Artificial Intelligence Review
https://doi.org/10.1007/s10462-021-10018-y

A review of deep learning used in the hyperspectral image analysis for agriculture

Chunying Wang[1] · Baohua Liu[1] · Lipeng Liu[1] · Yanjun Zhu[1] · Jialin Hou[1] · Ping Liu[1] · Xiang Li[2]

Abstract
Hyperspectral imaging is a non-destructive, nonpolluting, and fast technology, which can capture up to several hundred images of different wavelengths and offer relevant spectral signatures. Hyperspectral imaging technology has achieved breakthroughs in the acquisition of agricultural information and the detection of external or internal quality attributes of the agricultural product. Deep learning techniques have boosted the performance of hyperspectral image analysis. Compared with traditional machine learning, deep learning architectures exploit both spatial and spectral information of hyperspectral image analysis. To scrutinize thoroughly the current efforts, provide insights, and identify potential research directions on deep learning for hyperspectral image analysis in agriculture, this paper presents a systematic and comprehensive review. Firstly, its applications in agriculture are summarized, include ripeness and component prediction, different classification themes, and plant disease detection. Then, the recent achievements are reviewed in hyperspectral image analysis from the aspects of the deep learning models and the feature networks. Finally, the existing challenges of hyperspectral image analysis based on deep learning are summarized and the prospects of future works are put forward.

Keywords Agriculture · Classification · Detection · Deep learning · Hyperspectral imaging

Chunying Wang and Baohua Liu have contributed equally to this work.

✉ Ping Liu
liupingsdau@126.com

✉ Xiang Li
lixiang@sdau.edu.cn

[1] Shandong Provincial Key Laboratory of Horticultural Machinery and Equipment, Shandong Provincial Engineering Laboratory of Agricultural Equipment Intelligence, College of Mechanical and Electronic Engineering, Shandong Agricultural University, Taian 271018, China

[2] State Key Laboratory of Crop Biology, College of Life Sciences, Shandong Agricultural University, Taian 271018, China

Published online: 24 May 2021

1 Introduction

Precision agriculture is the science of improving crop yields and assisting management decisions using high technology sensor and analysis tools. Hyperspectral imaging is a technology to obtain spatial and spectral information simultaneously by using imaging and spectroscopy (Bioucas-Dias et al. 2013). Hyperspectral imaging techniques have been widely applied in precision agriculture (Awad 2019; Caballero et al. 2020; Ishida et al. 2018; Kuska et al. 2015). Hyperspectral data provides information on the temporal and spatial changes of farmland, which helps to quickly and accurately predict the growth of crops. For example, the remote sensing data can be used to inverse some important information such as nitrogen content, chlorophyll content, leaf area index, crop water content, etc., to predict crop growth and yield for precision agriculture (Awad 2019). Besides, many important aspects of agricultural production are effectively monitored using hyperspectral data, including the quantitative analysis of plant growth, the classification of plant seeds, the effective management of pests and diseases, the quality and safety analysis of agro-food products, and the effective assessment of soil texture.

However, hyperspectral imaging techniques have several challenges such as high dimensional data, the limited number of labeled samples, and the large spatial variability of adjacent spectra. The automatical learning of features from input hyperspectral image data makes deep learning more appealing for analyzing hyperspectral data (Petersson et al. 2016).

With the rapid development of deep learning, deep learning models have recently been successfully adopted in various areas such as computer vision (Dan et al. 2011; Athanasios et al. 2018; Voulodimos et al. 2018), speech recognition (Hinton et al. 2012), and other similarly complex data processing (Garcia-Garcia et al. 2017; Guo et al. 2016). Compared with traditional neural networks, deep learning architectures with more layers can potentially extract abstract and invariant features for better image analysis (Bengio et al. 2012). Meanwhile, the detection or classification tasks are also the need for precision agriculture. Consequently, deep learning has a greater potential for analyzing hyperspectral images including hyperspectral remote sensing data (Lin et al. 2014; Yang et al. 2016; Han et al. 2019) and laboratory data (Wang et al. 2019) in agriculture (Li et al. 2019). To scrutinize thoroughly the current efforts, provide insights, and identify potential research directions on deep learning for hyperspectral image analysis in agriculture, this paper mainly reviews the latest advances of hyperspectral image analysis in agriculture based on deep learning, focusing on the classification of varieties, the ripeness and component prediction, the remote sensing image classification, and the plant disease detection. The remainder of the review is organized as follows. Section 2 gives the common deep learning models used for hyperspectral image analysis. Section 3 reviews the recent applications and developments of deep learning for hyperspectral image analysis and gives the problems of these applications. Finally, Sect. 4 summarizes the challenges of hyperspectral image analysis and gives the prospects of the future.

2 Deep learning models used for hyperspectral image analysis

In this section, we focus on major types of deep learning models applied to hyperspectral image analysis, including convolutional neural networks (CNNs), fully convolutional networks (FCNs), tensor learning models (TLs), deep belief networks (DBNs), stacked auto-encoders (SAEs), recurrent neural networks (RNNs), semi-supervised learnings, generative adversarial networks (GANs), and active learning models (ALs).

2.1 Convolution neural network

In recent years, CNN models have been successfully applied to image recognition and segmentation. CNN models achieve more remarkable results than traditional machine learning (Simonyan and Zisserman 2015; Yang et al. 2017; Huang et al. 2017; Pound et al. 2017). CNN is proposed by Lecun and Bottou for image processing firstly (Lecun and Bottou 1998). This CNN model is featured by two key properties: spatially shared weights and spatial pooling. It adopted a local link model similar to the biological neural network, to achieve weight sharing, reduce the number of weights, and improve the speed of data processing. Generally, the typical CNN models are the various combinations of the convolutional layers, max-pooling layers, and fully-connected (FC) layers. LeNet5, a common two-dimension CNN model, is shown in Fig. 1a. The abstract features are learned by alternating convolutional layers and pooling layers, then extracted by the FC layer. Each convolutional layer convolves the input data with multiple local filters to generate translation-invariant local features. The subsequent pooling layers reduce the spatial dimension of the generated feature maps following several rules, such as average, max, and so on, which do not affect the depth (spectra) dimension. The high-level abstract features in the neural network are extracted via FC layers. The backpropagation algorithm is used to adjust or fine-tune the weights and biases. The trainable parameters are optimized by using the stochastic gradient descent algorithm to speed up the training procedure. LeNet5 is the first CNN model applied in hyperspectral image analysis (Makantasis et al. 2015a). Due to the lack of large-scale training data at that time and the computing power of the computer, LeNet5 is not ideal for complex problems. Since 2006, researchers have designed many methods to overcome the difficulty of training deep CNN. Among them, the most famous is AlexNet which is similar to LeNet5 but a bit deeper (Krizhevsky et al. 2012). AlexNet made a major breakthrough in image recognition tasks. After that, researchers have proposed deep convolutional neural networks (DCNNs) (Yoo 2015): ZFNet (Zeiler and Fergus 2014), VGG-Net (Simonyan and Zisserman 2015), GoogleNet (Szegedy et al. 2014) and ResNet (Russakovsky et al. 2015; He et al. 2016).

Deep neural networks can learn deep feature representation for hyperspectral image interpretation and achieve high classification accuracy in different datasets. The proposal of Deep Residual Network (ResNet) is a milestone event in the history of CNN. ResNet solves the problem that deep CNN models are difficult to train (He et al. 2016). ResNet has recently been successfully adopted in hyperspectral image analysis, such as hyperspectral image classification (Zhong et al. 2017a), hyperspectral image denoising (Yuan et al. 2018), increasing the spatial resolution of hyperspectral image (Wang et al. 2017) and unsupervised spectral–spatial feature learning of hyperspectral images (Mou et al. 2017a).

The fully Convolutional Network (FCN) is the first deep learning network architecture developed for pixel-level segmentation (Long et al. 2015). FCN changes image-label

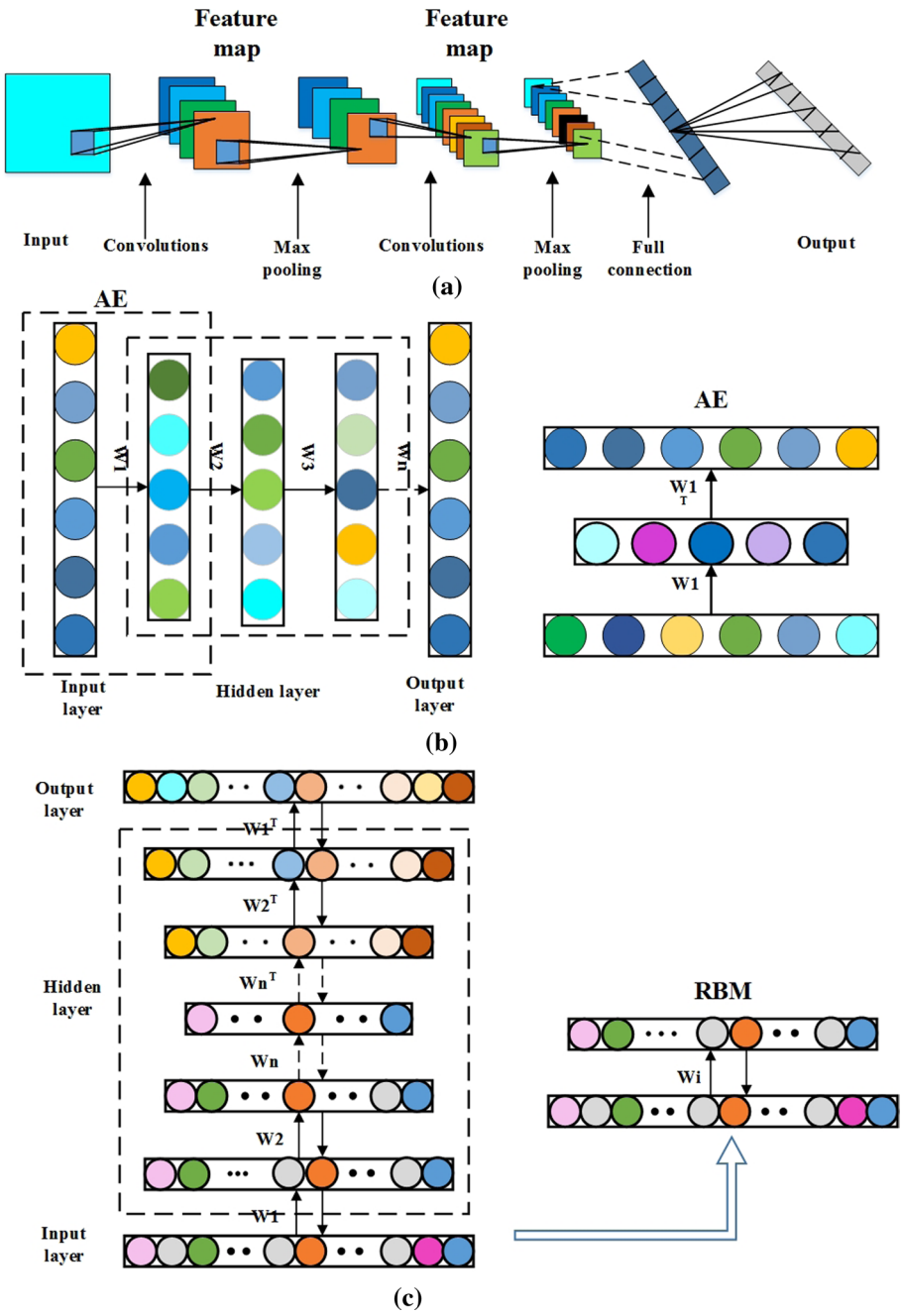

Fig. 1 Graphic depictions of CNN, SAE, and DBN. **a** CNN. **b** SAE, **c** DBN

mapping to pixel/voxel-label mapping for object detection and image segmentation. FCN takes the input of the arbitrary size and produces correspondingly-sized output with efficient inference and learning. In FCN architecture, the fully connected layers are replaced by 1×1 convolution layers on the top to generate prediction maps that contain classified information. Then, the prediction maps are upsampled to the same size of the input image through a transposed convolutional layer and then compared with the ground truth. The transposed convolution is also called fractionally strided convolution, which is performed in-network for end-to-end learning by backpropagation from the pixelwise loss. The traditional convolutional operation is a down-sampling process, while the transposed convolution is an up-sampling process. The prediction map and ground truth have the same size and resolution as the input image (Li et al. 2018a; Zhang et al. 2020a). FCN is also used to focuses on learning effective discriminant features for hyperspectral image classification (Jiao et al. 2017). Since the training samples in hyperspectral images are highly sparse, the training strategy in original fully convolutional networks is no longer feasible for hyperspectral images. A novel mask matrix is proposed to assist the back-propagation in the training stage of FCN (Xu et al. 2019). A novel hyperspectral change detection method is proposed for hyperspectral images, called the recurrent three-dimensional (3D) fully convolutional network (Re3FCN), which merged the advantages of a 3D fully convolutional network (FCN) and a convolutional long short-term memory (ConvLSTM) (Song et al. 2018). This study is the first to use a 3D FCN and a ConvLSTM for remote-sensing change detection. U-Net (Ronneberger et al. 2015), SegNet (Badrinarayanan et al. 2017), Deeplab series (Chen et al. 2015a, 2017a, b, 2018), and other FCN based networks have been proposed one after another and can realize the end-to-end segmentation of hyperspectral images providing a great breakthrough in semantic segmentation (Liu et al. 2020c).

The aforementioned approaches are almost vector or matrix-based methods, however, the intuitive representation of a hyperspectral image cube is a three-dimensional (3D) volumetric array, including one spectral dimension and two spatial dimensions. Therefore, it is more natural to treat the hyperspectral image as a 3D cube or tensor to preserve the high-order data structure. A series of 3D, contextual deep CNN, tensor-based methods have been successfully applied on HSI to conjunctively fuse the spatial features with spectral information. For instance, 3D Convolutional neural network (3D-CNN) (Li et al. 2017a), recurrent 2D Convolutional neural network (R-2D-CNN), and recurrent 3D Convolutional neural network (R-3D-CNN) are adopted to extract deep spectral–spatial features directly from raw hyperspectral image (Yang et al. 2018a; Song et al. 2018). Contextual deep CNN is based on Residual Network and FCN, which can optimally explore local contextual interactions by jointly exploiting local spatial–spectral relationships of neighboring individual pixel vectors (Lee and Kwon 2017; Zhang et al. 2018).

By preserving as many as possible the original spectral–spatial constraints, tensor representation helps to reduce the number of unknown parameters used in learning a linear dimensionality reduction model (Zhang et al. 2013; Zhong et al. 2015; Feng et al. 2017; Makantasis et al. 2018a; Huang et al. 2019a). Above Tensor-Based methods require any additional classifiers (e.g. SVM, CNN). But, a spectral–spatial classification framework based on low-rank tensor learning (lrTL) is proposed for classifying the high-dimensional hyperspectral images with the limited number of training samples (He et al. 2018). The unlabeled test samples by minimal error residuals in low-rank tensor learning algorithm, which doesn't require any additional classifiers. A novel tensorial approach, namely, generalized regression, is proposed for hyperspectral image classification (Liu et al. 2019). This model can be utilized to capture not only the intrinsic structure of data in a physical sense but also the generalized relationship of data in a logical sense.

2.2 Stacked auto-encoder model

The SAE model was used to get useful high-level features for remote sensing image classification in 2016 (Chen et al. 2016b), which introduces the concept of deep learning into hyperspectral image analysis firstly. The auto-encoders (AE) composed of encoder and decoder learns a new representation of the data by reconstructing the input data (Bengio et al. 2007). Each layer of SAE is associated with an AE trained separately. The abstract features are extracted by reconstructing the input data layer by layer (Fig. 1b). In the unsupervised pretraining stage, the learned features of one AE are used as the input data for training the next AE in a greedy way. The pretraining minimize the reconstruction error in each AE. After that, the parameters, such as weights and biases, in all AEs are used as the initial values for SAE optimized with respect to a supervised training criterion. Concretely, when the labeled data is used as the supervised signal, the parameters of each layer are fine-tuned using backpropagation of error, and the parameters of the structure are updated by the stochastic gradient descent algorithm. After that, researchers have proposed other improvement methods, which are denoising autoencoders (DAE) (Vincent et al. 2008) and stacked denoising autoencoders (SDAE) (Vincent et al. 2010).

2.3 Deep belief network

The deep belief network (DBN) is firstly proposed in 2006 by Hinton (Hinton and Salakhutdinov 2006; Hinton et al. 2006). DBN utilizes the Restricted Boltzmann Machine (RBM) as a learning module, similar to the way that SAE uses auto-encoders as their building blocks (Fig. 1c). Differently, the connection construction is symmetrical in the DBN. A DBN composes of multiple RBMs with connections between the layers but not between units within each layer, where the output of a layer is used as the input of the next layer. The RBM layers are firstly pre-trained using unlabeled samples in an unsupervised way to ensure that features are retained as much as possible. Then, the backpropagation algorithm fine-tunes the whole DBN network jointly with a small number of labeled samples. The extracted deep features are finally used for the detection and classification tasks. Importantly, the DBN networks overcome the main deficiencies of the backpropagation algorithm, such as local optimization and long training time, by the random initialization of weight parameters. Deep Boltzmann Machines (DBMs) are another type of deep model using RBM as their building block (Yang et al. 2019). The difference in the architecture of DBNs is that, in the latter, the top two layers form an undirected graphical model and the lower layers form a directed generative model, whereas in the DBM all the connections are undirected. Differing from 3D-CNN models which directly applying 3-dimensional kernel to hyperspectral image analysis, SAE, DBM, and DBN-based models firstly employ the dimensionality-reduction methods to reduce the spectral dimensionality to an acceptable level.

2.4 Recurrent neural networks

The recurrent neural network (RNN) was originally introduced in 1989 (Williams 1989). In contrast to a feed-forward neural network, RNN recognizes patterns in sequences of data and dynamic temporal characteristics by using a recurrent hidden state whose

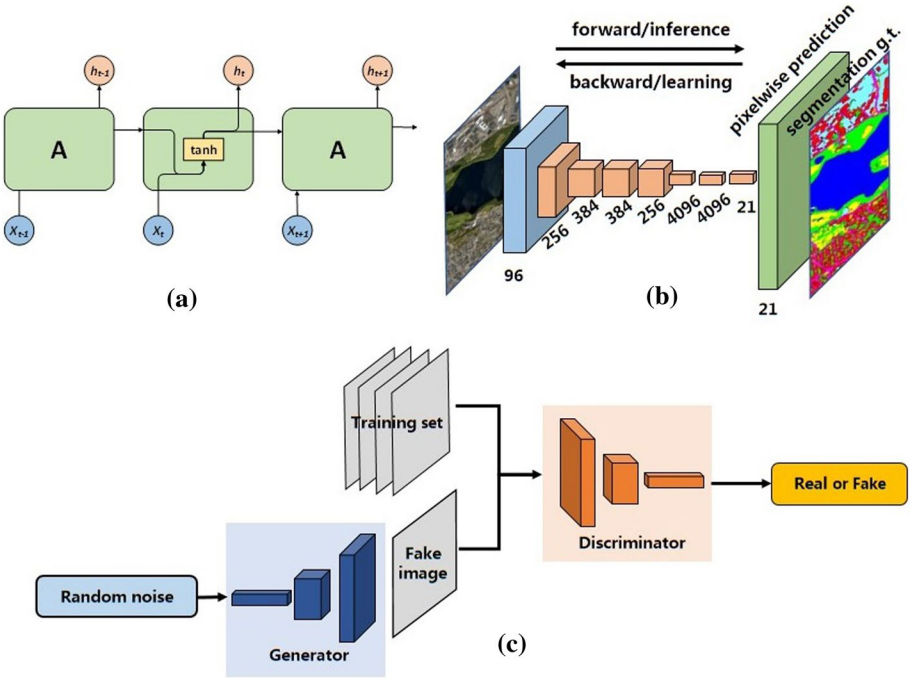

Fig. 2 Graphic depictions of RNN, GAN, and active learning. **a** RNN. **b** GAN. **c** AL

activation at each step depends on that of the previous steps (Fig. 2a). RNNs have been used for hyperspectral image classification firstly (Mou et al. 2017b), which can effectively analyze hyperspectral pixels as sequential data and then determine information categories via network reasoning. However, the traditional RNN model has the problem of gradient disappearance or gradient explosion. Long short-term memory networks (LSTMs) (Hochreiter and Schmidhuber 1997; Jiang et al. 2018) and gated recurrent units (GRUs) (Chung et al. 2014) solve this problem to a certain extent. LSTM variants include LSTM-F, a unidirectional LSTM model that operates in a many-to-one mode, and LSTM-S (unidirectional) and base LSTM (b-LSTM)-S (bidirectional) which can operate in a many-to-many model, generating an output class prediction output for every input timestep, and obviating the need for sliding-window segmentation. A bidirectional-convolutional long and short term memory network (Bi-CLSTM) is used to automatically learn the spectral–spatial features from hyperspectral images, which improves the classification performance by almost 1.5% as compared to 3D-CNN (Liu et al. 2017c).

2.5 Semi-supervised learning

Designing a hyperspectral image analysis system poses a few challenges: (1) Availability of only a limited number of labeled data. (2) Collection of labeled data is a highly time-consuming, expensive, and complicated process (Sawant and Prabukumar 2018). Semi-supervised learning is designed to alleviate the small-sample problem by utilizing both the

limited labeled samples and the wealth of unlabeled samples that can be easily obtained without significant cost (Russakovsky et al. 2015).

The semi-supervised methods can be roughly divided into four types: (1) generative models (Jin and Raich 2012) (2) Low density separation (Sun and Wang 2014; Adriana et al. 2015) (3) Graph-based methods (4) Wrapper-based methods (Jamshidpour et al. 2020). The self-training (Li and Jose 2013) and co-training (Zhang 2014) algorithms are commonly-used wrapper-based methods. The graph-based semi-supervised technique is gaining a lot of attention due to its ability to achieve a satisfactory hyperspectral image classification performance (Tian et al. 2015; Sawant and Prabukumar 2018; He et al. 2021; Jiang et al. 2021; Sawant and Prabukumar 2018; He et al. 2021). Semi-supervised HSI classification methods are also inspired by the generative adversarial networks (GANs) (He et al. 2017), CRNN (Wu and Saurabh 2017), SAE (Protopapadakis et al. 2021).

2.6 Generative adversarial networks

The Learning generative adversarial network (GAN) was firstly proposed in 2014, which utilizes adversarial training to generate the region of samples based on the required class label (Goodfellow et al. 2014) (Fig. 2b). Generative approaches try to learn the distribution parameters from data, and then they can generate new samples according to the learned models. Discriminate approaches attempt to model the dependence of labels on training data which can be used for predicting labels from training data. GANs (Zhan et al. 2018) and improved GANs such as Deep convolutional GAN (Chen et al. 2019), 1-D and 3-D GAN (Zhu et al. 2018), Capsule GAN (Xue 2020), Cascade conditional GAN (Liu et al. 2020b), MDGAN (Gao et al. 2019), 3DBF-GAN (He et al. 2017), have been used for hyperspectral image analysis. GANs make full use of the limited labeled samples as well as the sufficient unlabeled samples and achieve very promising results with a small number of labeled samples.

2.7 Active deep learning

Applying limited labeled samples to improve classification results is a challenge in hyperspectral images. GAN, semisupervised Learning, and active Learning are three promising techniques to address this challenge. Active learning (AL) is an iterative procedure of selecting the most informative examples from a subset of unlabeled samples (Fig. 2c). Selecting new training samples for active learning is based on the uncertainty and the intrinsic distribution and structure of the unlabeled samples (Cui et al. 2019; Liu et al. 2018d; Lei et al. 2021; Mu et al. 2020). The active learning method can train a deep network faster and with fewer training samples than traditional semisupervised learning methods.

Most of the research on active learning is combined with a special classifier. Examples include a kernel-based method, an active learning method combined with a support vector machine (SVM), logistic regression (LR), extreme learning machine (ELM) (Pradhan et al. 2019a, b), and Gaussian process regression.

Many spectral–spatial feature-based AL have been proposed for hyperspectral image classification (Bhardwaj et al. 2020a, b). CNN and Capsnet are also combining with AL for hyperspectral image classification (Cao et al. 2020; Paoletti et al. 2020). Most of these approaches are closely connected with a specific type or a specific structure of the classifier. Examples are random sampling (RS), maximum uncertainty sampling

(MUS), multiview (MV) (Li et al. 2020; Zhang et al. 2020b), query-by-committee (QBC), mutual information (MI)-based sampling approach (Liu et al. 2018d) as well as the breaking ties (BTs) selection criterion (Liu et al. 2018d).

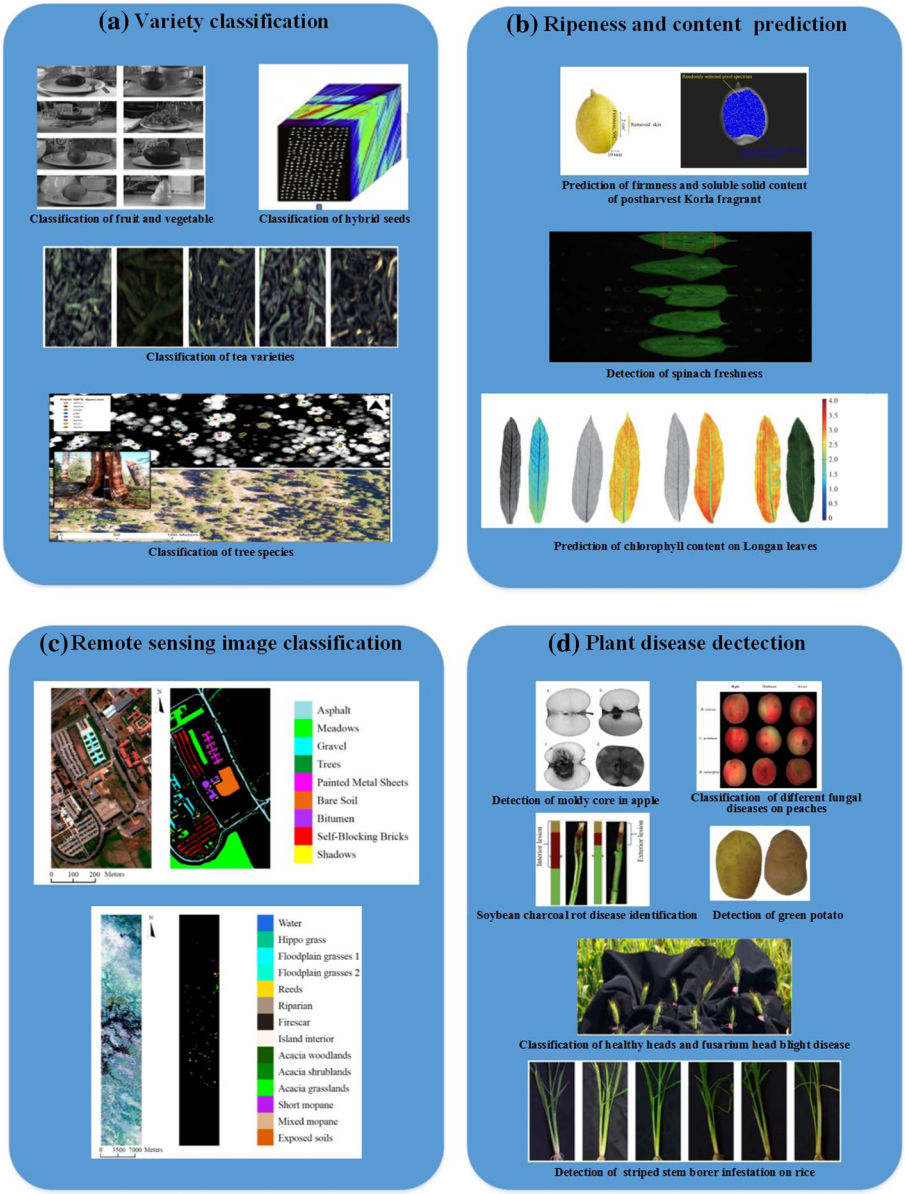

Fig. 3 Deep learning for hyperspectral image-based advanced application in agriculture

Table 1 Deep learning for hyperspectral image-based advanced application in agriculture

Category	Task	References
Different classification themes	Classification of fruit and vegetable	(Steinbrener et al. 2019)
	Classification of hybrid seeds	(Nie et al. 2019)
	Classification of tea varieties	(Sun et al. 2018a)
	Classification of tree species	(Fricker et al. 2019)
	Peanut hyperspectral image classification	(Liu et al. 2020a)
	Peanut planting area classification	(Huang et al. 2019b)
	Classification of film in seed cotton	(Ni et al. 2019)
Ripeness and content prediction	Prediction of firmness and soluble solid content of postharvest Korla fragrant	(Yu et al. 2018b)
	Detection of spinach freshness	(Xie et al. 2019)
	Prediction of chlorophyll content on Longan leaves	(Gan et al. 2018)
	Rapid detection of nitrogen concentration in oilseed rape (Brassica napus L.) leaf	(Yu et al. 2018a)
	Firmness detection for apples	(Rao et al. 2019)
	Potassium content prediction for citrus leaves	(Yue et al. 2019)
	Detection of apple multi-quality parameters	(Wang et al. 2020)
	Tomato maturity discrimination	(Jiang et al. 2021)
	Prediction of the total volatile basic nitrogen (TVB-N) content in peeled Pacific white shrimp	(Yu et al. 2019a)
	Prediction of the total viable count (TVC) content in peeled Pacific white shrimp	(Yu et al. 2019b)
	Freshness discriminating of shrimp	(Yu et al. 2018c)
	TVB-N measurement in pork	(Guo et al. 2017)
	Identification of lotus seed flour adulteration	(Hu et al. 2020)

Table 1 (continued)

Category	Task	References
Plant disease dectection	Detection of moldy core in apple	(Zhou et al. 2017)
	Classification of different fungal diseases on peaches	(Sun et al. 2018b)
	Soybean charcoal rot disease identification	(Nagasubramanian et al. 2017)
	Detection of green potato	(Li et al. 2016)
	Classification of healthy heads and fusarium head blight disease	(Jin et al. 2018)
	Detection of striped stem borer infestation on rice	(Fan et al. 2019)
	Detection of rice panicle blast	(Huang et al. 2017)
	Detection of Huanglongbing in citrus leaves	(Lu et al. 2019)
	Detection of peaches fungal diseases	(Jiang et al. 2021)
Remote sensing image classification	Classification of scene	others in references

3 Hyperspectral image analysis in agriculture using deep learning

A large amount of researches on hyperspectral images are proven that deep learning algorithms are very reliable, which give some inspiration to the specific applications in agriculture. The main applications are focusing on the classification of varieties, ripeness and component prediction, remote sensing image classification, and plant disease detection (Fig. 3; Table 1). The analysis of agricultural hyperspectral images using deep learning is discussed in this section.

3.1 Hyperspectral image analysis in agriculture using CNN models

According to the types of extracted features, these CNN-based networks for the hyperspectral image analysis are mainly divided into three manners, such as spectral feature-based networks, spatial feature-based networks, and joint spatial–spectral feature-based networks (Fig. 4). The classical CNN models applying to hyperspectral image analysis are summarized and listed in Table 2).

3.1.1 Spectral feature-based networks

In the hyperspectral images, each pixel in the spatial information contains a 1-D spectral vector composed of spectral data. The spectral feature-based network determines the attributes of all pixels only by using spectral information that is the most significant and directest feature. The CNN models autonomously extract discriminative features from the original spectral information (Slavkovikj et al. 2015). The learned features are used for classification tasks in the conjunction with classifiers. The spectral domain contains discriminant information in the hyperspectral images. Structures trained with merely spectral data have a good performance for hyperspectral image analysis. A 4-layer CNN is designed to classify hyperspectral images directly in spectral-domain for the first time (Fig. 5) (Hu et al. 2015). This CNN combines one convolutional layer, one max-pooling layer, and two fully connected layers. The input represents a pixel spectral vector, followed by the convolution layer and the max-pooling layer in turns to computing a set of 20 feature maps classified with a fully connected network. Three hyperspectral data, including Indian Pines, Salinas, and University of Pavia scenes, are employed to

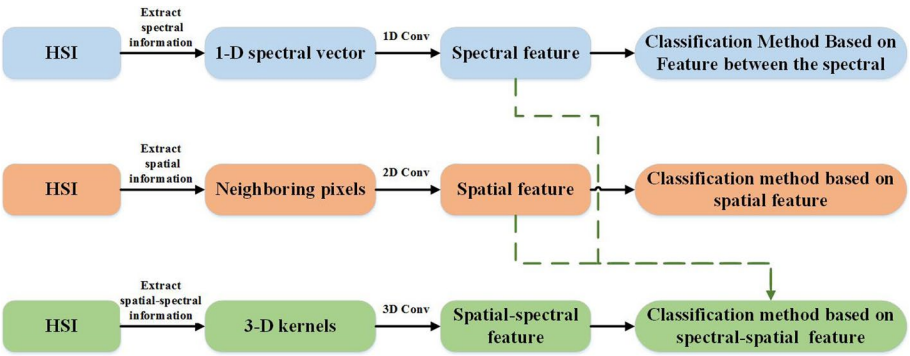

Fig. 4 Manners of the CNNs applying to hyperspectral images

Table 2 The classical CNN models applied for hyperspectral image analysis

Models	Key concept	Tasks	References	Source code
VGGNet	Small-sized (3 by 3) convolutional filters to increase the depth of CNNs (up to 19 layers)	Classification for hyperspectral imaging of Fusarium head blight disease symptom	(Jin et al. 2019)	http://www.robots.ox.ac.uk/~vgg/research/very_deep/
ResNet	Residual representation and skip connection scheme to enable the training of very deep CNNs (up to 1000 layers)	Classification for hyperspectral imaging of Fusarium head blight disease symptom	(Jin et al. 2019)	https://github.com/tensorflow/models/tree/master/official/vision/image_classification
GoogLeNet	A 22 layers deep network (allows for increasing the depth and width of the network while keeping the computational budget constant)	Peanut hyperspectral image classification	(Liu et al. 2020a)	https://github.com/PanJinquan/tensorflow_models_learning
SegNet	Consisted of an encoder network, a corresponding decoder network followed by a pixel-wise classification layer	Peanut planting area classification	(Huang et al. 2019b)	https://github.com/alexgkendall/caffe-segnet
AlexNet	A five-layer CNN architecture	Classification of scene	(Lee and Kwon 2017)	https://github.com/TensorFlow/models/blob/master/research/slim/nets/alexnet.py
DenseNet	An open source system that computes dense, multiscale features from the convolutional layers of a CNN based object classifier.	Classification of scene	(Yang et al. 2018b)	https://github.com/liuzhuang13/DenseNet
DeepLab	Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs.	Classification of scene	(Niu et al. 2019)	https://github.com/tensorflow/models/tree/master/research/deeplab
FCN	Fully Convolutional Network is the first deep learning network architecture developed for pixel-level segmentation	Classification of scene	(Jiao et al. 2017)	https://github.com/shekkizh/FCN.tensorflow
Mask RCNN	Mask R-CNN, extends Faster R-CNN by adding a branch for predicting an object mask in parallel with the existing branch for bounding box recognition	Classification of scene	(Kumar et al. 2020)	https://github.com/matterport/Mask_RCNN

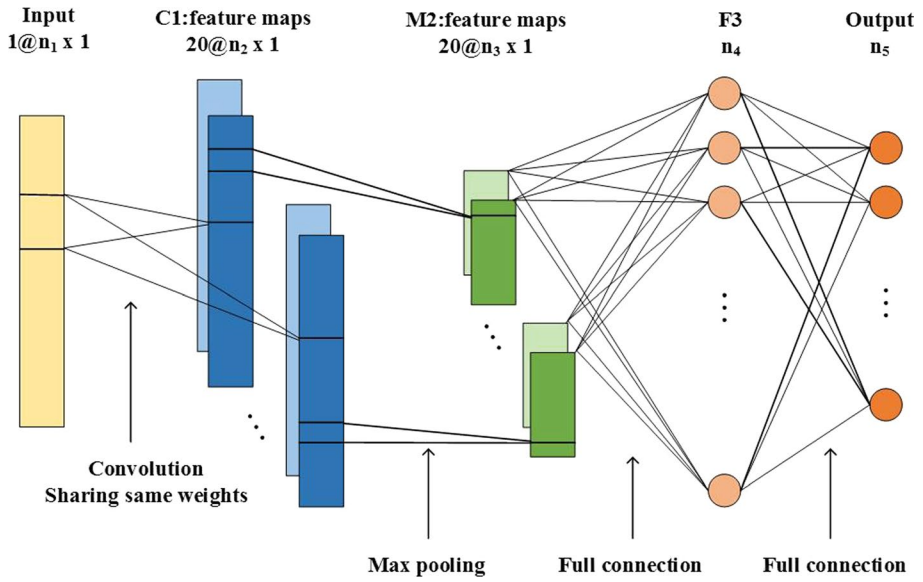

Fig. 5 A 5-layer CNN to classify hyperspectral images (Hu et al. 2015)

evaluate the effectiveness of the proposed CNN. For all the data, 200 labeled pixels per class are randomly selected for training and all other pixels in the ground truth map for testing. This CNN tested using three datasets achieves better classification performance with 90.16%, 92.6%, and 92.56%, respectively. The training time of this CNN on the Indian Pines dataset is 4300s and the testing time is 1.98s. A DCNN model which consists of two convolution modules and five FC layers is constructed to classify loofah seeds and hybrid okra seeds using spectral data (Nie et al. 2019). The average spectrum is convolved in two convolution modules to select features. Then, all the features are flattened into a 1-D structure to feed into the FC layers. Finally, the six classification results are obtained via the softmax layer. Moreover, the work demonstrated that CNNs combining with NIR hyperspectral imaging technology can be used to identify different hybrid seeds resulting from the cross-breeding of crops. Differing from the common 2D-CNN, the convolution and pooling operation of 1D-CNN are 1-D computations. It mainly considers the spectral domain while neglecting spatial information details, which is a shortage. The dataset contains six different varieties each of hybrid loofah and hybrid okra seeds. The hybrid okra seeds included six different varieties: 2014HK2, 2014HK4, 2014HK16, baiguo, cuizhi, and danzhi, and the hybrid loofah seeds also included six different varieties: zhesi35, 1.35 M, guangF, 25 F, guangM, and guangl. The overall classification accuracy of the DCNN is 98.94% and 98.24%, respectively. Another spectral feature-based framework for hyperspectral classification based on fully CNN is proposed (Li et al. 2018a). Through convolution, deconvolution, and pooling layers, the deep features of hyperspectral data are enhanced. After feature enhancement, the optimized extreme learning machine (ELM) is utilized for classification. 10% labeled data of Indian Pines and 5% labeled data of Pavia University are utilized as the training samples to train the proposed network. The overall classification accuracy of this method tested using the Indian Pines dataset is 96.7%. The overall classification accuracy of this method tested using the University of Pavia dataset is 85.11%.

3.1.2 Spatial feature-based networks

The hyperspectral image not only reflects the spectral information with hundreds of bands but also contains spatial characteristics. In the hyperspectral image analysis tasks, spatial information cannot be ignored, because adjacent pixels may belong to the same attribute (class) and there is other discriminant information related to the shape and size of different structures. As indicated in many spectral–spatial classification studies, the utilization of spatial features can significantly improve the detection or classification accuracy of hyperspectral images (Fauvel et al. 2013; Zhang et al. 2013). Hence, excavating spatial information is of great importance for hyperspectral image analysis. The spatial feature-based network considers the neighboring pixels of a certain pixel in the hyperspectral image to extract the spatial features. Subsequently, 2D-CNN models are adopted to extract the high-level spatial features, where the input data are the neighboring pixels patches. The learned spatial features are finally fed to classifiers such as the softmax layer, logical regression (LR) for detection or classification. However, due to the spectral dimension in a hyperspectral image, it is necessary to map the spectral information to an acceptable scale with a low information loss using dimensionality-reduction methods prior to the model training.

To reduce the spatial dimensionality and the computational burden, Makantasis and colleagues introduced randomized principal component analysis (PCA) of the spectral dimension to condense the whole image (Fig. 6). The first 10 to 30 principal components (PCs) are preserved. The CNN with three convolutional layers extracts spatial features from neighbors of 5×5 sizes in hyperspectral images, which are the input of the multi-layer perceptron classifier (Makantasis et al. 2015a). The Indian Pines and University of Pavia scenes datasets are employed to evaluate the effectiveness of the proposed CNN. Moreover, compared to the commonly-used CNN, this network has not max-pooling layer, which results in not take into account any translation and scale invariance. The split ratio ranges from 5 to 75% of the size of the whole dataset and corresponds to the size of the training set. The misclassification error of this method on the Indian Pines dataset is 2.5773 to 0.1861%. The misclassification error of this method on the University of Pavia dataset is 0.2281 to 0.0086%. When 80% of the tagged samples are randomly chosen as the training set, and 10% and 10% are randomly chosen for the validation and testing sets, respectively, overall classification accuracies of this network tested using the Pavia center, Pavia

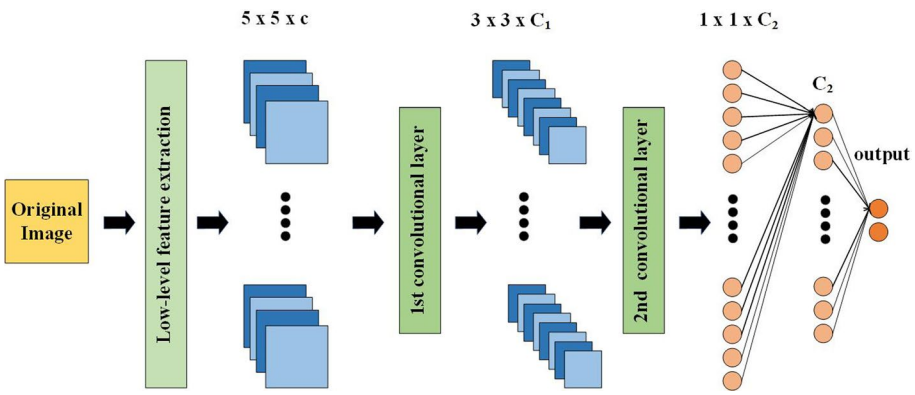

Fig. 6 A deep CNN model framework for classification (Makantasis et al. 2015a)

University, Salinas, and Indian Pines datasets are 99.91%, 99.62%, 99.53%, and 98.88%. Due to the fast convergence rate, this CNN network can be scaled for other large datasets. The spectral dimensionality is reduced by PCA, then the hyperspectral images are decomposed into patches of 27×27 sizes in pixel to train the eight-layer deep CNN model (Luo et al. 2017). University of Pavia scenes are employed to evaluate this CNN. This spatial size is larger than that in the CNN of Makantasis and colleagues and the overall classification accuracy of this method trained on 9% of dataset size is 97.57%.

In the practical applications based on the manually collected datasets, many CNN models have been used for those datasets. GoogLeNet-22, a deep CNN model, is a 22-layer network formed by repeatedly stacking basic Inception modules, which is proposed for detecting a panicle blast of rice (Huang et al. 2017). This rice panicle blast dataset is collected by authors using GaiaField-F-V10, consists of 4480 samples of panicle blast, 4704 samples without panicle blast. The GoogLeNet-22 learns the high-dimensionality features to extract and fuse the different scale local spatial features by using the multi-scale convolutional kernels of inception modules. Especially, feature extraction and classification learning are seamlessly integrated into a unified framework. Both of them are trained jointly under the supervision of label samples, which helps to improve the blast prediction performance. Experimental results show that the proposed GoogLeNet based method achieves a high classification accuracy of 92.0%. However, counterintuitively, the classification performance of deep learning models degrades as their depth increases. The deep residual networks (ResNets) have been used for hyperspectral image classification (Zhong et al. 2017a). In the Indian Pines dataset, 10%, 10%, and 80% of available annotated data are adopted for the learning, validation, and predicting, respectively. Similarly, in the University of Pavia dataset, 5%, 5%, and 90% are used for the same purposes. Overall classification accuracy obtained by ResNet-4 (40 convolutional kernels) tested using the Indian Pines dataset is 96%. Overall classification accuracy obtained by ResNet-4 (40 convolutional kernels) tested using the University of Pavia dataset is 99%. The two typical deep CNN networks, VGG and ResNet, are constructed for detecting Fusarium head blight disease through the proximal hyperspectral images collected in the fields (Jin et al. 2019). This Fusarium head blight disease dataset is collected by authors using Hy-Scan1211, contains 809200-pixel samples. The input of models is the reconstructed 2-D pixel spectral data with no loss of information. Comparing the different layer depths of VGG and ResNet, the VGG network with 22-layer (VGG-22) is the optimal detection network with the accuracy 72.4%, which extracts spectral features more effectively, results in obtaining obvious and accurate disease region. However, those CNN models are trained by small samples, which may cause the misclassification problem. To solve this problem, the CNN pre-trained for the RGB image dataset is utilized to classify fruit and vegetable hyperspectral images (Steinbrener et al. 2019). To create the fruit and vegetable classification custom dataset, a hyperspectral camera from Ximea (MQ022HG-IM-SM4X4-VIS) is used to record a total of 2700 images of 13 different classes of fruits and vegetables. The image size for each of the spectral bands is 512×272 pixels The pre-trained GooleNet is fine-tuned with the dataset of small hyperspectral images to obtain a classifier. Then, an additional layer is added after the input layer to classify the hyperspectral image with the RGB pre-trained network. Overall classification accuracy of this approach is 92.23%. What's more, this approach could easily be extended to other applications involving the hyperspectral image classification based on RGB pre-trained networks.

Except for the above proximal hyperspectral images, remote sensing images are also combined with deep learning for agricultural detection or classification. The SegNet network is developed for a remote sensing peanut planting area classification (Huang et al.

2019b). The dataset peanut planting remote sensing images is acquired from Sentinel-2. The number of the training images is 2316 and the number of the testing dataset is 1193. Firstly, the selected four-band spectral images are utilized for training the SegNet network including encoder and decoder. Each encoding layer corresponds to a decoding layer. Then, the first thirteen convolutional layers of the VGG-16 are the encoder to extract spatial features. The output features of the decoder are fed to the softmax layer that calculates probability in all categories of each pixel to obtain an optimal classification. The peanut planting area classification accuracy of this approach is 89.2%. The GoogleNet-22 network is also used for classifying the five different varieties of peanuts through hyperspectral images (Liu et al. 2020a). The peanut dataset contains 450 images acquired by using Finland Specim Imaging Spectroradiometer. The overall accuracy of the test set is 93.3%. Therefore, the CNN models have shown a great advantage in solving the hyperspectral images and remote sensing images. Similarly, a four-level deep learning architecture targets the land cover and crop species classification basing on remote sensing images, which is the ensemble of the spectral features learned by the 1D-CNN and the spatial features learned by the 2D-CNN (Nataliia et al. 2017). However, this architecture requires additional empirical processing to combine the learned features without a fully automatic feature representation learning procedure. The dataset is Landsat-8 and Sentinel-1A images. The study area is classified into eleven classes including major agricultural crops (water, forest, grassland, bare land, winter wheat, winter rapeseed, spring cereals, soybeans, maize, sunflowers, and sugar beet). The overall accuracy of the 2D-CNN is 94.6%.

3.1.3 Spatial–spectral feature-based networks

The rich spectral and spatial information are utilized in the hyperspectral image analysis. There are generally two types in extracting spatial–spectral features that further improve the network performance. The first is to extract spatial and spectral features respectively based on deep network models and various techniques, and then these spatial features are combined with spectral features to perform classification or detection. The second type is to extract spectral–spatial features simultaneously by 3D-CNNs. 3D-CNN models also use 3-D kernels to learn features correlation between the spatial and the spectral domain of the hyperspectral data cubes for classification or detection. Additionally, 3D-CNNs generated at different scales could be applied to extract deep spectral–spatial joint features (Li et al. 2017a; Chen et al. 2016a).

However, a huge amount of data contained in the hyperspectral images will increase the data analysis complexity and slow down the calculation efficiency. Therefore, it is necessary to reduce dimensionality in analyzing hyperspectral data (Zhang et al. 2016).

A hybrid classification method combined CNN with dimension reduction (DR) operated by PCAis proposed, which fully takes the spatial information and the spectral characteristics of hyperspectral images into account (Liu et al. 2017a). 50 training samples in class 1 to 9 and 500 training samples in class 10 to 13 are selected respectively. The overall accuracy of the DR-CNN-vs tested using the HYDICE dataset is 96.68 %. A novel deep learning framework for hyperspectral image classification using both spectral and spatial features is presented (Yue et al. 2015). The framework is a hybrid of principal component analysis, deep convolutional neural networks (DCNNs), and logistic regression (LR). The features generation algorithm is firstly present to generate the spectral and spatial features. After that, the DCNNs-LR classifier is trained to get useful high-level features and to fine-tune the whole model. The preprocessed features are the input of deep CNN, including

three 2-D spectral features and three spatial features. Spatial features are the first three PCs extracted in hyperspectral images, with 42 × 42 sizes in pixels. Particularly, the 1-D spectral features are structured into 2-D data. The experimental study is carried out with 3921 training samples and a total number of 42,776 testing samples of the Pavia dataset. The overall accuracy of the DCNNs-LR is 95.18 %. In the hyperspectral datasets with high-dimension, a balanced local discriminant embedding algorithm is also applied to reduce spectral dimension and extract spectral features in low dimensionality. For instance, a spectral–spatial feature-based classification (SSFC) framework is proposed and tested on Pavia center and University datasets (Zhao and Du 2016). 2D-CNN is utilized to find high-level spatial features after the dimensionality-reduction on the hyperspectral images, which is fusing the spectral and spatial features for classification. The average overall accuracy of SSFC is 97.83%. The overall accuracies of SSFC-SVM (50 samples from each class to form the training data set) tested using Pavia Center and Pavia University datasets are 99.29% and 96.03%. Another is that three wavelengths are screened by the grouping elite strategy genetic algorithm to reduce dimensionality in analyzing hyperspectral data (Xie et al. 2019). A 2D-CNN is constructed to extract the image features for identifying the fresh grade of spinach. The 2D-CNN consists of four convolutional layers and pooling layers, one FC layer, and two dropout layers. The dataset of spinach images is acquired using HSI-VNIR-00001. Training set and test set each containing 240 spinach samples. The spinach is divided into three grades of fresh, relatively fresh, and corruption according to the score calculated by considering 6 factors: fresh spinach days of storage, appearance, water content, chlorophyll a, chlorophyll b, and carotenoids. The average recognition accuracy of spinach fresh grade is 80.99%.

Many dual-channel CNN frameworks have been proposed for extracting spatial–spectral features. A similar strategy of CNN with two channels is used to learn jointly spectral–spatial features from hyperspectral image(Yang et al. 2016). Each channel CNN has

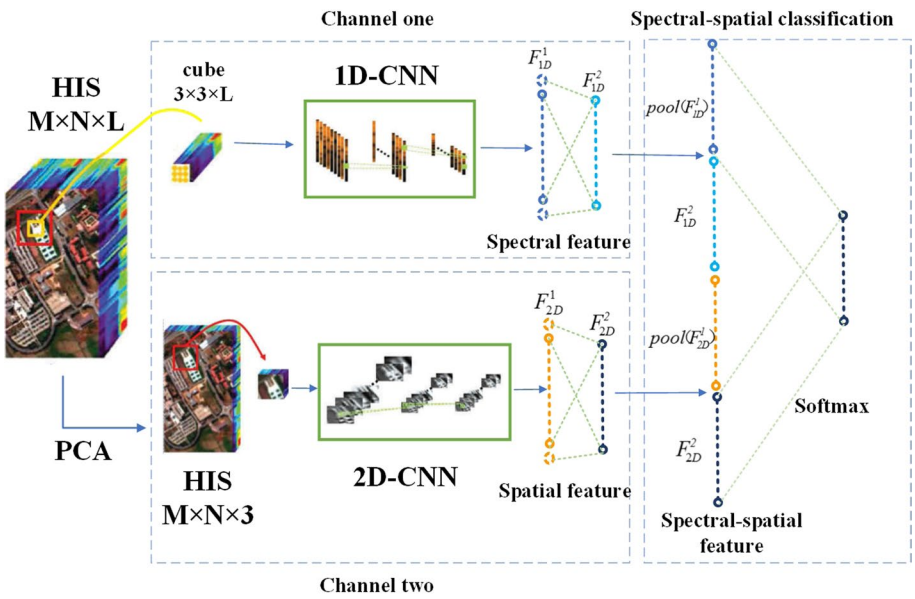

Fig. 7 A dual-channel CNN framework for classification (Zhang et al. 2017)

five convolutional layers. Salinas Valley dataset is used in transfer learning and the Indian pines dataset is used for the test. The overall classification accuracy of this dual-channel CNN is 95.58% when the number of training samples is 200. Another dual-channel CNN framework is proposed for classification (Zhang et al. 2017) (Fig.7). The spectral features and spatial features are extracted via 1D-CNN and 2D-CNN respectively after dimensionality-reduction processing. A Softmax classifier is used to combine the spectral and spatial features jointly and classify the scene eventually. In order to extract the high-level abstract spectral features, a $3 \times 3 \times$ L-sized cube is extracted from the eight neighborhoods of one pixel. The cubes are rearranged into nine pixel-vectors, feeding into the 1D-CNN. In channel two, a data cube with size $41 \times 41 \times 3$ is chosen from the neighborhood as the input of 2D CNN. Indian Pines and University of Pavia scenes are employed to evaluate the effectiveness of the proposed dual-channel CNN. 5% and 10% from each class of the ground truth data are randomly chosen as the training samples for Pavia University and Indian Pines, respectively, and take the remainders as testing samples. The overall classification accuracies of this dual-channel CNN tested using the Pavia University and Indian Pines datasets are 99.68% and 98.76%. The training time of this dual-channel CNN tested using the Pavia University and Indian Pines datasets are 3421.25s and 4860.57s, respectively. Considering the within-class variation and the similarity between classes in the spectral domain, the spectral features are extracted via constructing supervised within-class/between-class hypergraph (SWBH) (Kong et al. 2018). What's more, the workers present the sample expanded CNN model (SECNN), which is trained based on newly generating a large amount of labeled hyperspectral image samples by the random zero settings, to extract spatial features for forming spectral–spatial joint features. Indian Pines, Pavia University, and Salinas are used to evaluate the performance of the proposed SWBHCSECNN. 1717 is chosen to be the optimal window size for the three hyperspectral datasets. The number of training samples per class is set as 200. The average overall classification accuracies of the proposed SWBHCSECNN tested using these three datasets are 99.54%, 99.27%, and 99.34%. The average training time is 2701.94s, 2808.37s, and 5340.91s. A novel deep learning-based method for this task is proposed, by learning a non-linear end-to-end mapping between the noisy and clean HSIs with a combined spatial–spectral residual network (HSID-CNN) (Yuan et al. 2018). HSID-CNN achieves the highest MPSNR and MSSIM values and the lowest MSA values in all the noise levels. Overall classification accuracy obtained by HSID-CNN tested using the Indian Pines dataset is 85.65%. Overall classification accuracy obtained by HSID-CNN tested using the University of Pavia dataset is 86.99%. The average run-time for denoising of HSID-CNN is 3.5s. A spectral–spatial fully convolutional network (SSFCN) is proposed for HSI classification with an end-to-end, pixel-to-pixel architecture (Xu et al. 2019). The network starts with two branches, learning the hierarchical spectral and spatial features respectively. These two branches flow into a merging layer, Then, the joint features are followed by a convolutional layer and the softmax function to yield the classification map. The mask matrix is utilized during the network training stage. The number of training samples per class is set as 150. Overall classification accuracies obtained by SSFCN with CRF tested using the University of Pavia dataset, the Houston dataset, and the Salinas dataset are 98.11% 95.51%, and 98.4%, respectively. The run-time using those three datasets are 272.71s, 981.72, 323.55s, respectively.

It is more natural to treat the hyperspectral image as a 3D cube or tensor to preserve the high-order data structure. A series of 3D CNN, contextual deep CNN, tensor-based methods have been successfully applied on HSI to conjunctively fuse the spatial features with spectral information.

As the spectral and the spatial features are extracted simultaneously, the 3D-CNNs could take full advantage of the characteristics of the 3-D data and fully exploit the joint spatial/spectral correlations information. This method has the strength that 3D-CNN could model finer data information attributed to the 3-D convolution operations and directly focuses on classification tasks instead of two independent channels to extract spectral and spatial features (Liu et al. 2018a). A 3D convolutional neural network (3D-CNN) framework is proposed for accurate hyperspectral image classification (Li et al. 2017a). The proposed method views the HSI cube data altogether without relying on any preprocessing or post-processing, extracting the deep spectral–spatial-combined features effectively. Besides, it requires fewer parameters than other deep learning-based methods. Three hyperspectral data, including Indian Pines, Botswana, and University of Pavia scenes, are employed to evaluate the effectiveness of 3D CNN. 50% labeled samples are randomly chosen as training data, and the rest is used for testing. Overall classification accuracy obtained by this 3-D CNN-based FE model tested using the Indian Pines dataset is 99.07%. Overall classification accuracy obtained by this 3-D CNN-based FE model tested using the University of Pavia dataset is 99.39%. Overall classification accuracy obtained by this 3-D CNN-based FE model tested using the Kennedy Space Center dataset is 99.55%. A 3-D CNN-based FE model with combined regularization and virtual sample enhanced method is proposed to extract effective spectral–spatial features of hyperspectral imagery (Chen et al. 2016a). Overall classification accuracy obtained by this 3-D CNN-based FE model tested using the Indian Pines dataset is 98.53%. Overall classification accuracy obtained by this 3-D CNN-based FE model tested using the University of Pavia dataset is 99.66%. Overall classification accuracy obtained by this 3-D CNN-based FE model tested using the Kennedy Space Center dataset is 97.03%. Another 3D-CNN is constructed for detecting apple multi-quality parameters based on hyperspectral images, which reduces the training cost and the risk of overfitting (Wang et al. 2020). This 3D-CNN contains three 3-D convolutional layers, one 3-D pooling layer, and three separate FC layers to predict the Brix, firmness, and moisture simultaneously. The dataset of apple hyperspectral images for detecting apple multi-quality parameters is acquired using HIS-VNIR-0001. The classification model of apple's brix, firmness, and moisture established this algorithm achieved the prediction accuracy of 93.97%, 92.29%, and 93.36%, respectively. When the CNN model executes the divided patches of hyperspectral images, the border information cannot be defined around some pixels. Before the training of a 3D-CNN architecture with five layers, a specific preprocessing strategy is used for management of the borders of the image to properly account for the border information around some pixels, then the processed hyperspectral images are fed to the CNN to increase the classification accuracy (Paoletti et al. 2018). Overall classification accuracy obtained by this network tested using the Indian Pines dataset is 98.37%. Overall classification accuracy obtained by this network tested using the University of Pavia dataset is 97.8%. Another supervised 3D-CNN model is developed for the identification of healthy and charcoal rot samples on soybean plants (Nagasubramanian et al. 2017). This model consists of two convolutional layers interspersed with two max-pooling layers following by two FC layers. The size of kernels in convolutional layers are 3×3 in spatial dimension and 16 in spectral dimension. The relatively small 3D-CNN model and dropout mechanism are used to prevent overfitting. The plant disease identification dataset is acquired using Pika XC hyperspectral line imaging scanner. The dataset contains 111 hyperspectral stem images of size 500×1600×240. Among the 111 images, 64 represent healthy stems and 47 represent infected stems. Data patches of resolution 64×64×240 pixels are extracted from the stem images. The 64×64×240 image patches are applied as input to the 3D-CNN model. The training dataset consists

of 1090 images. Out of 1090 training images, 940 images represent the healthy stem and 150 images represent the infected stem. The validation dataset consists of 194 samples and the test dataset consists of 539 samples. The accuracy of identifying charcoal rot disease in soybeans using 3D-CNN is 95.73%. Due to the large calculation using full bands in training, the band selection method is needed to reduce the dimensionality of large hyperspectral images based on robust interpretability. A five-layer 3D-CNN is designed for crop species classification method through the multi-temporal remote sensing images (Fig. 8) (Ji et al. 2018). This 3D-CNN with fine-tuned parameters is trained for max out to make samples a score to each label by the VGG for features extraction. Moreover, an active learning strategy is introduced to improve the classification accuracy to a required level and increase the high-quality samples. Although the third dimension of samples is temporal data, this 3D-CNN is also suitable for spectral dimension. The crop species classification dataset is two GF2 (Gaofen 2) multi-temporal images and a GF1(Gaofen1) image. A novel change detection method (2DCNN-LSTM) for hyperspectral images (HSIs), including sample generation and a deep-learning network, called the recurrent three-dimensional (3D) fully convolutional network (Re3FCN), which merged the advantages of a 3D fully convolutional network (FCN) and a convolutional long short-term memory (ConvLSTM) (Song et al. 2018). This is the first to use a 3D FCN and a ConvLSTM for remote-sensing change detection. The strategy assisted in training fewer samples of representative feature expression. The Re3FCN is mainly comprised of spectral–spatial and temporal modules. Particularly, a spectral–spatial module with a 3D convolutional layer extracts the spectral–spatial features from the HSIs simultaneously, whilst a temporal module with ConvLSTM records and analyzes the multi-temporal HSI change information. The hyperspectral change detection dataset included two sites of hyperspectral EO-1 Of all the samples, 70% are used for training, and 30% are used for testing. Hyperion multi-temporal images obtained from Yancheng in the Jiangsu province of China. The overall accuracy of the multi-class change detection on the 2DCNN-LSTM is 96.2%. Another fully convolutional network-based framework, named deep multiscale spatial–spectral feature extraction algorithm is proposed, which focuses on learning effective discriminant features for hyperspectral image classification (Jiao et al. 2017). The well pretrained deep fully convolutional network based on VGG-very deep-16 is introduced to excavate the potential deep multiscale spatial structural information in the proposed hyperspectral imaging framework. Overall classification accuracies obtained by this network tested using the Indian Pines dataset, the University of Pavia dataset, and the Salinas dataset are larger than 97%. The training testing time of this network with SVM on the Indian Pines dataset is only 4.96s. Four new deep learning models, namely 2D Convolutional neural network (2D-CNN), 3D Convolutional neural network (3D-CNN), recurrent 2D Convolutional neural network (R-2D-CNN), and recurrent 3D Convolutional neural network (R-3D-CNN) are advocated for hyperspectral image classification (Yang et al. 2019). In the 2D-CNN model, three convolutional layers are utilized. To preserve the vital information of each pixel, the pooling layers are excluded from our 2D-CNN model. Finally, a fully connected layer, which takes the feature maps of the last 2D convolutional layer as inputs, is constructed to make the prediction. For the feature extraction phase, a 3D convolution operator instead of the 2D convolution operator is applied to the 3D-CNN model. The main difference with 2D-CNN is that the 3D-CNN model has one extra phase of reordering. The R-2D-CNN deep neural network comprises a recurrent CNN structure, where a basic 2D-CNN block is reused multiple times. More specifically, it uses the basic 2D-CNN block to extract the feature maps for the 1-st level instances at the beginning. These feature maps are then concatenated with the 2-nd level instances, which are fed to the same 2D-CNN block for extracting the next level feature

Table 3 Overall classification accuracies of different models using different datasets (Yang et al. 2019)

	Indian Pines scene (%)	Pavia University scene (%)	Botswana Scene (%)	Salinas scene (%)	Pavia Center scene (%)	Kennedy Space Center (%)
2D-CNN	97.08	95.46	97.60	98.96	96.02	97.76
3D-CNN	98.92	98.49	97.21	99.08	98.75	98.46
R-2D-CNN	99.19	99.19	98.54	99.47	99.88	99.22
R-3D-CNN	99.50	99.97	99.38	99.80	96.79	99.85

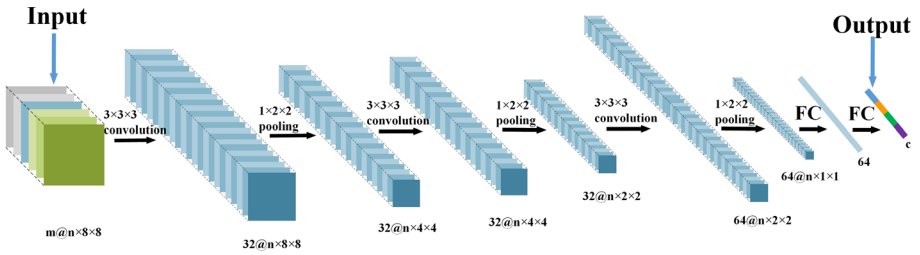

Fig. 8 A five-layer 3D-CNN based crop classification method (Ji et al. 2018)

maps. As for the R-2D-CNN model, the R-3D-CNN model is also underpinned by multi-level recurrent neural networks which shrink a patch gradually to form multi-level instances. Six hyperspectral data, including Indian Pines, Botswana, Salinas, Kennedy Space Center, Pavia Center, and University of Pavia scenes, are employed to evaluate the-effectiveness of those models. Overall classification accuracies of different models using different datasets are shown in Table 3.

A wider and deeper contextual deep CNN is developed to extract spatial–spectral features simultaneously using relatively small amounts of training samples (Lee and Kwon 2017). The contextual deep CNN optimally explores local contextual interactions by jointly exploiting local spatial–spectral relationships of neighboring individual pixel vectors. The joint exploitation of the spatial–spectral information is achieved by a multi-scale convolutional filter bank used as an initial component of the contextual deep CNN pipeline. The initial spatial and spectral feature maps obtained from the multi-scale filter bank are then combined together to form a joint spatial–spectral feature map. The joint feature map representing rich spectral and spatial properties of the hyperspectral image is then fed through a fully convolutional network that eventually predicts the corresponding label of each pixel vector. The number of training samples per class is set as 200. Overall classification accuracy obtained by this contextual deep CNN tested using the Indian Pines dataset is 93.61%. Overall classification accuracy obtained by this contextual deep CNN tested using the Salinas dataset is 95.07%. Overall classification accuracy obtained by this contextual deep CNN tested using the University of Pavia dataset is 95.97%. Overall classification accuracy obtained by pure contextual deep CNN (Lee and Heesung 2016) tested using the Indian Pines dataset is 92.06%. By this means the obtained model could generate more powerful feature expression with fewer samples. A novel simplified deep learning model, R-VCANet, is proposed, which achieves higher accuracy when the number of training samples is not abundant (Pan et al. 2017). In R-VCANet, the inherent properties of

hyperspectral image data, spatial information, and spectral characteristics are utilized to construct the network. Spectral and spatial information are combined via the rolling guidance filter (RGF), which could explore the contextual structure features and remove small details from hyperspectral images. a new network called Vertex Component Analysis Network (VCANet) is designed for deep features extraction from the smoothed hyperspectral images. 50% of all labeled pixels are selected as training samples. Overall classification accuracies obtained by R-VCANet tested using the Indian Pines, Pavia University and KSC datasets are 97.9%, 96.77%, and 97.9%, respectively. Different from commonly-used CNNs extracting features through a sliding window with a specific scale, a diverse region-based CNN (DR-CNN) learns contextual interactional features using the six diverse region inputs (Zhang et al. 2018). The joint representation containing rich spectral and spatial information is then fed to a fully connected network to predict the label of each pixel vector by a softmax layer. More importantly, a multi-scale summation module is used as the feature extractor to avoid overfitting in the designed network. The number of training samples per class is set as 200. Overall classification accuracy obtained by the DR-CNN tested using the Indian Pines dataset is 98.54%. Overall classification accuracy obtained by the DR-CNN tested using the Salinas dataset is 98.33%. Overall classification accuracy obtained by the DR-CNN tested using the University of Pavia dataset is 99.56%. Compared to pure contextual deep CNN and the contextual deep CNN, the performance of DR-CNN is much improved.

By preserving as many as possible the original spectral–spatial constraints, tensor representation helps to reduce the number of unknown parameters used in learning a linear dimensionality reduction model. A tensor extension of conventional supervised manifold-learning-based DR. A tensor organization scheme is defined for representing a pixel's spectral–spatial feature and develop tensor discriminative locality alignment (TDLA) for removing redundant information for subsequent classification (Zhang et al. 2013). Datasets are hyperspectral images of the mall in Washington, the airborne hyperspectral images of Pavia provided by the Data Fusion Technical Committee of the IEEE Geoscience and Remote Sensing Society, and hyperspectral images of a mixed forest/agricultural site at the Indian Pine test site gathered by the Airborne Visible/Infrared Imaging Spectrometer (A VIRIS) sensor. The classification accuracy improves significantly while using a small number of features. Overall classification accuracy using three datasets is 97.66%, 96.42%, and 89.54%, respectively. The linear tensor-based model exploits principles of logistic regression assuming the rank-1 canonical decomposition property among its weights. A modification of a feedforward neural network (rank-1 FNN) is proposed for hyperspectral data classification and analysis (Makantasis et al. 2018a). When the training set size is small, rank-1 FNN outperforms RBF-SVM, SAE, and CNN models. Overall classification accuracy of rank-1 FNN using the Indian Pines dataset (100 samples per class) is 91.63%. Overall classification accuracy of rank-1 FNN using the Pavia university dataset (100 samples per class) is 93.5%. The local tensor discriminant analysis (LTDA) is used for hyperspectral image classification (Zhong et al. 2015). 10% of samples from each class are randomly chosen as the training set Overall classification accuracy of LTDA using the Indian Pines dataset is 93.55%. Overall classification accuracy of LTDA using the Pavia university dataset is 98.03%. A restricted Boltzmann Machine based on a three-order tensor (Tensor3-RBM) is designed for multi-temporal hyperspectral remote sensing image change detection (Huang et al. 2019a). The traditional BP neural network on the top layer of DBN is replaced with the Support Tensor Machine (STM), and a DBN with multi-layer Tensor3-RBM and STM (TRS-DBN) is constructed. Dataset is two different-temporal AVIRIS HSRS images and two different-temporal EO-1 Hyperion HSRS images (L1G level). The average precision

of TRS-DBN is 94.85%. A novel tensorial approach (GTR) is proposed for hyperspectral image classification (Liu et al. 2019). This model can be utilized to capture not only the intrinsic structure of data in a physical sense but also the generalized relationship of data in a logical sense. 40 training samples per class are randomly chosen. Overall classification accuracy of GTR with MSK using the Indian Pines dataset is 97.92%. Overall classification accuracy of GTR with MSK using the Pavia university dataset is 98.58%. Overall classification accuracy of GTR with MSK using the University of Houston dataset is 77.89%. Above Tensor-Based methods require any additional classifiers (e.g. SVM, FNN, Rank-1 FNN). But, a spectral–spatial classification framework based on low-rank tensor learning (lrTL) is proposed for classifying the high-dimensional hyperspectral images with the limited number of training samples (He et al. 2018). The unlabeled test samples by minimal error residuals in low-rank tensor learning algorithm, which doesn't require any additional classifiers. 20 samples per class are selected as input training tensors of TDLA Overall classification accuracy of lrTL using the Indian Pines dataset is 72.66%. Overall classification accuracy of lrTL using the Pavia university dataset is 71.16%. Overall classification accuracy of lrTL using the Salinas dataset is 84.61%. What's more, a superpixel tensor sparse coding (STSC) based hyperspectral image classification (HIC) method is proposed (Feng et al. 2017), by exploring the high-order structure of the hyperspectral image and utilizing information along all dimensions to better understand data.

3.2 Responding to the main challenges of deep learning and especially CNN applied for hyperspectral image analysis

CNNs, the typical deep learning models, need massive training datasets to trigger their powers. These models also need a mass of labeled training samples when applying CNNs to the hyperspectral image analysis. Unfortunately, a limited number of labeled samples are usually given in hyperspectral image datasets, which results in breaking the balance between high-dimensionality characteristics of hyperspectral images and limited labeled training samples. However, the acquisition of huge training samples for specific tasks is time-consuming and labor-consuming work. On the one hand, increasing the labeled training samples on the limited data is significant. The data augmentation approaches are used to address the issue. A virtual sample enhanced method is proposed to create training samples from a given training sample to tackle the problem of limited labeled samples. Based on the fact that the hyperspectral characteristics of one class vary within a certain range, the same label of training samples is assigned to the virtual sample (Chen et al. 2016a). Moreover, the virtual sample method is also adopted to significantly increase new samples in the way which is simulated by multiplying the random factor by the training sample and adding random noise (Liu et al. 2018a, b). Besides, strategies are executed for data augmentation: adding Gaussian noise (Zhang et al. 2018), rotating and flipping (Zhang et al. 2017), discarding randomly band images, and translating randomly the brightness of average hyperspectral image augmentation (Huang et al. 2017).

The pixel-pair strategy is utilized to solve the scarcity of labeled samples of the proposed framework (Li et al. 2017b). Specifically, a new data combination is constructed via pairing with any two selected samples from the available labeled data and the data entry is relabeled. Then, the proposed method utilizes deep CNN to learn pixel-pair features within a fixed window, and the final label is determined via a majority voting strategy.

On the other hand, applying limited labeled samples to improve classification results is a challenge in hyperspectral images. Transfer learning, generative adversarial networks,

semi-supervised learning, and active learning are promising techniques to respond to the challenge.

3.2.1 Transfer learning models applied for responding to the challenges

Transfer learning models have been successfully and reliably applied for hyperspectral image analysis (Bharti et al. 2021; Xie et al. 2021; Zhao et al. 2021; Jiang et al. 2019). The bottom and mid layers of CNN could be transferred from models of other scenes using transfer learning, while a few top layers are trained using the limited training samples (Yang et al. 2016). The classification methods are applied to Indian Pines data. Salinas Valley data is used in transfer learning. The overall classification accuracy of CNN-transfer is higher than CNN when training samples below 180. Hyperspectral image superresolution is also a challenge. A new framework is proposed to enhance the resolution of hyperspectral images by exploiting the knowledge from natural images (Yuan et al. 2017). The relationship between low/high-resolution images is the same as that between low/high-resolution hyperspectral images. In the proposed framework, the mapping between low- and high-resolution images can be learned by the deep convolutional neural network and be transferred to hyperspectral images by borrowing the idea of transfer learning.

What's more, a supervised deep feature extraction method based on siamese CNN has been proposed to solve the lack of training samples (Liu et al. 2018c). Firstly, a CNN with five layers is designed to directly extract deep features from hyperspectral cubes, where the CNN can be intended as a nonlinear transformation function. Then, the siamese network composed of two CNNs is trained to learn features that show a low intraclass and high interclass variability. The important characteristic of the presented approach is that the S-CNN is supervised with a margin ranking loss function, which can extract more discriminative features for classification tasks.

3.2.2 Generative adversarial networks applied for responding to the challenges

What's more, the generative adversarial network (GAN) utilizes adversarial training to generate the region of samples based on the required class label. Therefore, GANs are proposed to solve the limited sample problem, avoiding the overfitting phenomena to a great extent, and some improved GANs are applied to the hyperspectral image applications. Table 4 shows GANs and improved GANs for generating the hyperspectral data.

3.2.3 Semi-supervised algorithms applied for responding to the challenges

Many semi-supervised algorithms have demonstrated that the adoption of unlabeled data is useful for improving classification performance (Tuia and Camps-Valls 2009; Munoz-Mar et al. 2010).

The graph-based semi-supervised technique is gaining a lot of attention due to its ability to achieve a satisfactory hyperspectral image classification performance (Tian et al. 2015; Sawant and Prabukumar 2018; He et al. 2021; Jiang et al. 2021; Sawant and Prabukumar 2018; He et al. 2021).

A novel graph-based semi-supervised algorithm combined with particle cooperation and competition is proposed to improve the model performance effectively by using unlabeled samples (He et al. 2021). First, to reduce the generation of label noise, we used an efficient constrained graph construction approach to calculate the affinity matrix, which is capable

Table 4 Improved GANs for generating the hyperspectral images

Methods	References	Brief description	Overall classification results	Times
Deep convolutional GAN	(Chen et al. 2019)	Spectral data are generated but the quality of bands is not evaluated	10% samples are randomly chosen for the training sets. Overall classification accuracy tested using the Indian dataset is 95.36%. Overall classification accuracy tested using the University of Pavia dataset is 97.35%	
1-D GAN and 3-D GAN	(Zhu et al. 2018)	Spectral data samples and spectral–spatial data samples are generated but with data loss	200 training samples are chosen. Overall classification accuracy of 1D-GAN with Logistic Regression tested using the Salinas dataset is 89.13%. Overall classification accuracy of 1D-GAN with Logistic Regression tested using the KSC dataset is 89.64%. Overall classification accuracy of 1D-GAN with RBF-SVM tested using the Indian Pines dataset is 68.64%. Overall classification accuracy of 3D-GAN tested using the Salinas dataset is 93.02%. Overall classification accuracy of 3D-GAN tested using the KSC dataset is 96.89%. Overall classification accuracy of 3D-GAN tested using the Indian Pines dataset is 89.09%	Runtime of 3D-GAN using the Salinas dataset is 55.10s. The runtime of 3D-GAN using the KSC dataset is 48.57s. The runtime of 3D-GAN using the Indian Pines dataset is 72.46s
Cascade conditional GAN	(Liu et al. 2020b)	Two steps consist of spatial information generation and spatial–spectral information generation for hyperspectral images complete spatial–spectral sample generation	5% samples are randomly chosen for the training sets. Overall classification accuracy using C2GAN+ResNet on the Salinas dataset is 99.85%. Overall classification accuracy using C2GAN+CNN on the Pavia University dataset is 99.74%	

Table 4 (continued)

Methods	References	Brief description	Overall classification results	Times
Capsule GAN	(Xue 2020)	Network using the 3D-CNN and 3D CapsNet as the generator and discriminator, respectively generates the spectral and spatial data	Overall classification accuracy tested using the Pavia University dataset is 99.23%. Overall classification accuracy tested using the Indian Pines hyperspectral dataset is 99.12%	
HSGAN	(Zhan et al. 2018)	Generating hyperspectral samples that are similar to the real data	Overall classification accuracy with 10.0% TTR tested using the Indian Pines dataset is 83.53%	

of constructing a highly correlated similarity relationship between the graph and the samples. Then, we introduced a particle competition and cooperation mechanism into label propagation, which could detect and re-label misclassified samples dynamically, thus stopping the propagation of wrong labels and allowing the overall model to obtain better classification performance by using predicted labeled samples. Finally, we applied the proposed model to hyperspectral image classification. Overall classification accuracies obtained by CLPPCC tested using the Indian Pines image dataset, the Pavia University scene dataset, and the Salinas dataset are 95.71% 94.54%, and 98.81%, respectively. The average run-time using those three datasets is 49.82s, 958.87s, 1432.51s, respectively. An HSI classification method is proposed based on a semi-supervised BLS (SBLS) (Yi et al. 2018). Firstly, to make full use of abundant spectral and spatial information of hyperspectral imagery, hierarchical guidance filtering is performing on the original HSI to get its spectral–spatial representation. Then, the class-probability structure is incorporated into the broad learning model to obtain a semi-supervised broad learning version, so that limited labeled samples and many unlabeled samples can be utilized simultaneously. Finally, the connecting weights of broad structure can be easily computed through the ridge regression approximation. Overall classification accuracies obtained by SBLS tested using the Indian Pines image dataset, the Salinas dataset, and the Botswana dataset are 92.47% 96.14%, and 99.32%, respectively. The average run-time using those three datasets is 420.02s, 240.26s, 70.97s, respectively.

A semi-supervised method based on hyperspectral imaging technology is proposed for tomato maturity discrimination by using a small number of labeled samples (Jiang et al. 2021). A semi-supervised algorithm based on Laplacian score and spectral information divergence (SIDLS), which used class probability information to construct graphs, is designed to select a representative waveband subset. The sparse representation model based on class probability information (CSR) is established to construct a connection graph, and the label propagation algorithm is used to discriminate tomato maturity. The Tomato maturity discrimination dataset contains 192 tomatoes (48 tomatoes each maturity stage) acquired by using GaiaSorter. Overall classification accuracies obtained by CSR distinguishing different color distribution of tomato slices such as green, turning-color, hard, and red are 100.00%, 92.11%, 89.19%, and 94.87% respectively.

Semi-supervised HSI classification methods are also inspired by the generative adversarial networks (GANs) (He et al. 2017), CNN (Liu et al. 2017b), CRNN (Wu and Saurabh 2017), SAE (Protopapadakis et al. 2021).

A semisupervised framework is designed for HSI data based on a 1-D GAN (HSGAN) (Zhan et al. 2018). This framework enables the automatic extraction of spectral features for HSI classification. When HSGAN is trained using unlabeled hyperspectral data, the generator can generate hyperspectral samples that are similar to the real data while the discriminator contains the features which can be used to classify hyperspectral data with only a small number of labeled samples. In the case of the training to total sample ratio is comes to 10%, overall classification accuracy obtained by HSGAN tested using the Indian Pines dataset is 83.53%. The training time is 161s when unable to total sample ratio is 10%. A semi-supervised HSI classification method is proposed inspired by the generative adversarial networks (GANs) and the three-dimensional bilateral filter (3DBF) (He et al. 2017). Overall classification accuracies obtained by 3DBF-GAN tested using the Indian Pines dataset, the University of Pavia dataset, and the Salinas dataset are 75.62% 77.94%, and 87.63%, respectively, which lower than that of HSGAN tested using the Indian Pines dataset. Semi-supervised learning that can jointly exploit labeled and unlabeled samples to deal with the problem of dimensionality. A novel semi-supervised convolutional neural

network (SS-CNN) is proposed for the classification of hyperspectral images (Liu et al. 2017b). The proposed network can automatically learn features from complex hyperspectral image data structures. Furthermore, skip connection parameters are added between the encoder layer and decoder layer in order to make the network suitable for semi-supervised learning. 200 training samples are chosen. Overall classification accuracy obtained by SS-CNN tested using the University of Pavia scenes dataset is 98.32%. The average training time and testing time are 1220.2s and 1.6s, respectively. A deep few-shot learning method is proposed to address the small sample problem of hyperspectral image classification (Liu et al. 2018b). SpectralCspatial features are extracted to reduce the labeling uncertainty via a deep residual 3-D convolutional neural network. It is worth noting that the proposed method in this way can outperform conventional semi-supervised methods with only a few labeled samples. A Stack Auto-encoder (SAE)-Driven and Semi-Supervised (SSL)-Based Deep Neural Network (DNN) is proposed to extract buildings from relatively low-cost satellite near-infrared images (Protopapadakis et al. 2021). Semi-supervised deep learning for hyperspectral image classification and a constrained Dirichlet process mixture model (C-DPMM) is proposed for hyperspectral image classification tasks (Wu and Saurabh 2017). Deep convolutional recurrent neural networks (CRNN) are used for hyperspectral image classification by treating each hyperspectral pixel as a spectral sequence. C-DPMM is injected to utilize spatial information in the hyperspectral images. 10 labeled pixels per class are chosen for learning. Overall classification accuracies obtained by PL-SSDL (C-DPMM) tested using the Pavia University dataset, the Houston University dataset, and the Wetland dataset are 88.43% 82.61%, and 97.33%, respectively. The average training time using those three datasets is 42.5 min, 53.2 min, 83.5 min, respectively.

The self-training (Li and Jose 2013) and co-training (Zhang 2014) algorithms are commonly-used wrapper-based methods. Stacked convolutional autoencoder self-taught learning (SCAR) is proposed to learn feature extracting frameworks from unlabeled hyperspectral imagery (Kemker and Kanan 2017). When chose 5 % per class training set, overall classification accuracies obtained by SCAE-Hyperion tested using the Indian Pines dataset, the Salinas Valley dataset, and the Pavia University are 96.61%, 99.0%, and 99.5%, respectively.

3.2.4 Active deep learning models applied for responding to the challenges

Active learning methods have been widely studied for remote sensing applications. Active learning has obtained great success in supervised remotely sensed hyperspectral image classification since it is able to select highly informative training samples. An active learning algorithm based on a weighted incremental dictionary learning (WI-DL) is proposed (Liu et al. 2017d). Both uncertainty and distribution are utilized in selecting new training samples in the proposed active learning algorithm. The WI-DL achieves higher accuracy with fewer training samples by actively selecting training samples. The performance of WI-DL is better than those of other algorithms, and the classification accuracy of WI-DL improves faster than other algorithms as more samples are added. Overall classification accuracies of WI-DL tested using the Pavia university dataset, the Pavia center dataset, and the Botswana dataset are 97.2%, 92.4%, and 91.6%, respectively.

Most of the research on active learning is combined with a special classifier or a special remote sensing application. Examples include a kernel-based method, an active learning method combined with a support vector machine (SVM), logistic regression (LR), and Gaussian process regression. Extreme Learning Machine (ELM) classifier

often provides good generalization ability at a much faster learning speed in comparison to traditional classification approaches such as SVM, LR, BPNN. An active learning approach based on ELM that effectively decreases the computational time while maintaining the classification accuracy is proposed (Pradhan et al. 2019b). Final accuracies obtained by ELM-RS and ELM-MV after competition of 300 iterations tested using the Kennedy Space Center dataset are 87.57% and 90.08%. Final accuracies obtained by ELM-RS and ELM-MV after competition of 300 iterations tested using the Botswana dataset are 87.57% and 90.08%. The average computation time of ELM-RS for Botswana and Botswana datasets are 7.01s, 4.81s. An approach (KELM-AL) using kernel-based ELM (KELM) with AL techniques is proposed (Pradhan et al. 2019a). The proposed KELM-AL achieved classification accuracy up to 91.15% in the Kennedy Space Center dataset while 95.02% in the case of the Botswana dataset with computation times of 149.78 s and 104.98 s, respectively. The classification accuracy and computation time are better than that of ELM-RS or ELM-MV. A spatial prior generalized fuzziness extreme learning machine autoencoder (GFELM-AE) based active learning is also proposed, which contextualizes the manifold regularization to the objective of ELM-AE (Ahmad et al. 2020a).

Although active learning has been applied to many applications in remote sensing, most of these approaches are closely connected with a specific type or a specific structure of the classifier. Examples are random sampling (RS), maximum uncertainty sampling (MUS), multiview (MV) (Li et al. 2020) and query-by-committee (QBC). mutual information (MI)-based sampling approach as well as the breaking ties (BTs) selection criterion (Liu et al. 2018d). Hyperspectral data inherently owns high dimensionality, which makes it suitable for multi-view learning algorithms. Multi-view learning methods aim to construct separate but also sufficient views that can accurately learn a classifier with each one. A novel subpixel-pixel-superpixel-based multiview AL (MAL) (SPS-MAL) method is proposed for hyperspectral image classification (Li et al. 2020). The multiple views are generated via extracting the subpixel-level, pixel-level, and superpixel-level information. The multiple views can reflect various characteristics of hyperspectral images, i.e., spectral mixture, spectral discrimination, and spectral–spatial structure. In addition, a coarse-to-fine MAL algorithm is introduced to effectively select the most representative samples with the most uncertainty. Overall classification accuracies obtained by SPS-MAL tested using the Indian Pines dataset, the Pavia university dataset, and the Salinas dataset are 98.66%, 98.1%, and 98.34%, respectively. Although many view generation methods have been proposed so far, few of them consider the sufficiency of each view. Another novel multi-view multi-learner (MVML) active learning method is introduced, in which the different views are generated by a genetic algorithm (GA) (Jamshidpour et al. 2020). The GA-based view generation method attempts to construct diverse, sufficient, and independent views by considering both inter and intra-view confidences. A spectral–spatial graph-based semi-supervised learning (SSL) method is also implemented as the classifier, which improved the performance of the classification task in comparison with supervised learning. Overall classification accuracies obtained by MVML tested using the Indian Pines dataset, the Pavia university dataset, and the Salinas dataset are 97.14%, 98. 59%, and 99.45%, respectively. The wrapper GA-based view generation method constructs sufficient views at the first iteration of AL for the classification purpose. The two distinct views generated by the GA-wrapper achieved 78.47–78.94%, 84.09–83.68%, and 85.94–86.21% of the overall accuracy for Indian pines, Pavia University, and Salinas data sets, respectively. Although the GA-wrapper method constructs the more efficient views, the GA-Comb method could build a set of views that are efficient and simultaneously diverse.

Active learning generally favors the selection of samples following discriminative distributions, which are located in low-density areas. However, hyperspectral data are often highly class-mixed. In this case, the potential of active learning to select effective training samples is more limited. The concept of feature-driven AL is introduced with Gabor filtering and morphological profiles for instantiation (Liu et al. 2018d). 80 initial samples and 350 actively selected samples of the Indian Pines dataset are used for training. 45 initial samples and 350 actively selected samples of the Pavia university dataset are used for training. 65 initial samples and 350 actively selected samples of the Kennedy Space Center dataset are used for training. Overall classification accuracies obtained by Gabor-MI tested using the Indian Pines dataset, the Pavia university dataset, and the Kennedy Space Center dataset are 98.66%, 98.1%, and 98.34%, respectively. Overall classification accuracies obtained by Gabor-MI tested using the Indian Pines dataset, the Pavia university dataset, and the Kennedy Space Center dataset are 98.66%, 98.1%, and 98.34%, respectively. Overall classification accuracies obtained by Gabor-BT tested using the Indian Pines dataset, the Pavia university dataset, and the Kennedy Space Center dataset are 99.5%, 99.84%, and 99.53%, respectively. Another feature-driven AL is proposed based on dimensionality reduction, spectral–spatial information, and classification with active learning. Overall classification accuracies obtained by this method tested using the Kennedy Space Center dataset, the Pavia university dataset, and the Indian Pines dataset are 99.71%, 99.66%, and 99.44%, respectively. The computing time of this method tested using the Kennedy Space Center dataset, the Pavia university dataset, and the Indian Pines dataset are 6.95s, 17.15s, and 915.5s, respectively. A spectral–spatial AL model for the classification of HSI having limited training samples is proposed. In the proposed model first the spectral and spatial information of the HSI is integrated by constructing an extended superpixel profile (ESPP) (Bhardwaj et al. 2020a). First v samples are selected using uncertainty criterion and the final h samples are selected out of v samples using diversity criterion. The batch size h is kept 20 in the experiments. The value of v is kept as $3 \times h$ in all the experiments. Overall classification accuracies obtained by ESPP MS-ABD, ESPP MCLU-ECDB, and ESPP GA-Multic tested using the Kennedy Space Center dataset are 99.9%, 99.83%, and 99.89%, respectively. Overall classification accuracies obtained by ESPP MS-ABD, ESPP MCLU-ECDB, and ESPP GA-Multic tested using the Kennedy Space Center dataset are 98.3%, 98.61%, and 98.66%, respectively. A model that incorporates spatial information exploiting attribute profiles is also proposed by Bhardwaj and collages for improving the performance of active learning techniques (Bhardwaj et al. 2020b). Attribute profiles successfully integrate spectral and spatial information which helps in more accurate class discrimination in hyperspectral imaging. A multi-class non-randomized AL method is also proposed based on Multinomial Logistic Regression (MLR)-LORSAL classifier in conjunction with fuzziness as sample selection method whilst exploiting both spatial and spectral information of hyperspectral data (Ahmad et al. 2020b). The size of the training set is fixed to the maximum of 50 samples randomly selected from all the classes while giving equal representation to each class. Overall classification accuracies obtained by this method with fuzziness tested using the Indian Pines dataset, the Salinas dataset, the Pavia university dataset, and are 81%, 90%, and 87%, respectively. This work also validates the sufficient number of samples required to train a classifier; i.e., 500–1000 samples are more than enough to train a classifier to produce an acceptable accuracy for HSI classification tasks.

The high dimensionality of hyperspectral data, the presence of substantial noise, and the overlap of classes all contribute to the difficulty of automatically clustering and segmenting hyperspectral images. To solve this problem, an unsupervised learning technique called spectral–spatial diffusion learning (DLSS) combines a geometric estimation of class modes

with diffusion-inspired labeling that incorporates both spectral and spatial information (Murphy and Maggioni 2018). DLSS is robust to choices of parameters and has low computational complexity. Another diffusion-based method for active learning of hyperspectral images (LAND) is proposed (Tasissa et al. 2021). A deep variational autoencoder (VAE) extracts smoothed, denoised features from a high-dimensional hyperspectral image, which are then used to make labeling queries based on graph diffusion processes. With just10 labeled points, the overall accuracy of V AE-LAND is 96.97%, a 12.5% improvement to the competitive LAND algorithm. The proposed method could reduce the labeling cost of hyperspectral image classification, only 480 labeled samples are used to train the proposed method (Active DenseNet121). CapsNets are now widely regarded as state-of-the-art within the deep learning field. A new AL-based approach (AL-CapsNet) is presented for HSI data classification that integrates the spectral and the spatial information contained in the HSI data and enhances the performance of CapsNets when very limited training samples are available (Paoletti et al. 2020). The overall accuracies of AL-CapsNet tested using the Indian Pines and Kennedy Space Center datasets are 98.85% and 99.78%.

3.3 Hyperspectral image analysis in agriculture using SAE models

The first attempt that applying the SAE model to hyperspectral remote-sensing data for classification is executed. The SAE model is constructed to extract joint spectral–spatial features hierarchically in a layer-wise training fashion (Chen et al. 2014). The concept of deep learning is introduced into hyperspectral data classification for the first time. Spectral features and spatial features are extracted from the 1-D pixel vector and the neighborhood region of a certain pixel via SAE respectively. These features are fed as input to a final layer classifier like LR. Like other spatial feature-extraction methods that reduced the input dimension, this framework introduces PCA to condense the whole image. The proposed joint spectral–spatial deep neural network opens a new window for future research, showcasing the deep learning-based methods' huge potential for accurate hyperspectral data classification. The tagged parts of the image are split into three sets, i.e., training, validation, and testing data, with a split ratio of 6: 2: 2. Overall classification accuracy obtained by SAE-LR tested using the KSC dataset is 98.76%. Overall classification accuracy obtained by SAE-LR tested using the Pavia dataset is 98.52%. Differing from the frequently-used PCA dimensionality-reduction method that cannot process complicated data, the locality preserving projection method reduces dimensionality efficiently without losing any information and provides the extra functionality of yielding non-linear extension, as it is defined everywhere (Simranjit and Singara 2018). A DCNN is constructed in this work with Autoencoders which is then passed to the logistic regression for classification. 50% of the sample are randomly selected for training. Overall classification accuracy obtained by DCNN tested using the Indian Pines dataset is 84.44%. Overall classification accuracy obtained by DCNN tested using the Salinas dataset is 87.2%. The autoencoders are trained with 1000 epoch and the softmax layer is trained with 5000 epochHowever, the classification accuracy is lower than that of Chen's work. So, PCA is often used to condense the whole image.

In addition, the non-linear deep AEs are also trained by BP to reduce the dimensionality of high-dimensional data, which is an integral part of the hyperspectral image processing (Hinton et al. 2006).

Some researchers regard the SAEs as a part of the framework to extract effective features (Chen and Li 2017). The proposed framework serves as an engine for merging the

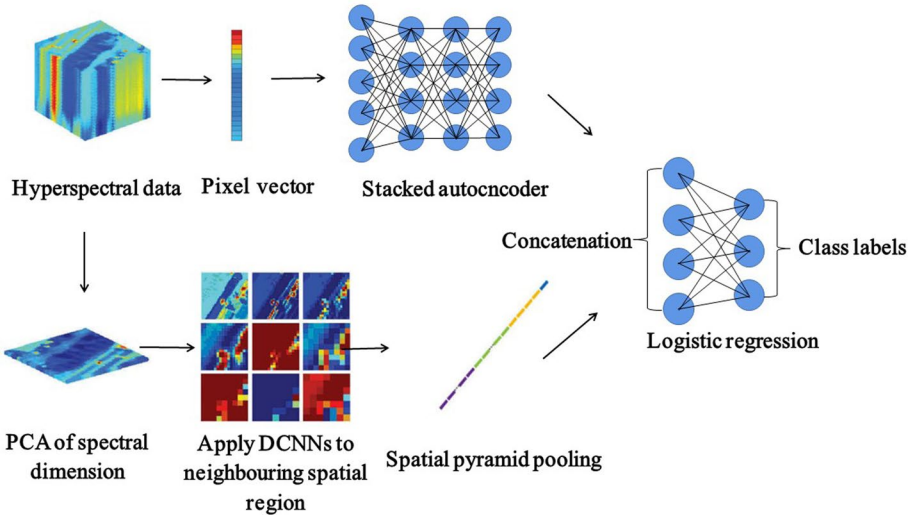

Fig. 9 The framework for classification merging the spectral and spatial features via SAEs and deep CNNs (Yue et al. 2016)

spatial and spectral features via suitable deep learning architecture: SAEs and DCNNs followed by a logistic regression classifier (Yue et al. 2016) (Fig.9). The SAE is used for getting the high-level spectral features and deep CNN is used to learn spatial feature samples automatically. Then the combined features are fed into the LR classifier to predict the attribute of each pixel. Meanwhile, workers solve the problem of large-scale variance of the hyperspectral image using the spatial pyramid pooling method. The spatial feature maps of the top convolutional layers are pooled into a fixed-length feature to aggregate the multi-scale feature maps, which improves the classification performance. 10% of the randomly chosen samples are used to determine the spectral–spatial feature adjustment parameter. For the Pavia data set, the best classification results are achieved with $\alpha = 0.7$ by the randomly chosen samples. Overall classification accuracy obtained by SSDL tested using the Pavia University data dataset is 95.27%. Also, the segmented SAE is proposed to reduce complexity but improve the efficiency of feature representation and the accuracy of hyperspectral image classification (Jaime et al. 2016). The original spectral domains of samples are decomposed into smaller segments of the spectrum and processed by different smaller SAEs separately. Although the segmented SAE consists of several SAEs, they are simpler than the one employed in the common case among smaller SAE which is a two-layer configuration with 40 units. The training set is selected randomly with an equal sample rate of 5% in each class Overall classification accuracy obtained by S-SAE tested using the Indian Pines dataset is 80.66%. Overall classification accuracy obtained by S-SAE tested using the Pavia CA dataset is 97.42%. The time required in extracting features from an original pixel is reduced by 60% and about 44% for the Indian Pines and Pavia CA data sets, respectively. The variable weighted SAE (VW-SAE-PSO-EL)is developed to achieve spectral dimensionality reduction, which could retain useful information hidden in the noise (Ni et al. 2019). The non-linear high-level features are extracted by the variable weighted SAE for classifying the film covering of seed cotton. The film sorting in the seed cotton dataset

contains 107 hyperspatial images (21 for training), with 384 × 210 × 288 sizes in pixels. Overall classification accuracy obtained by VW-SAE-PSO-EL is 95.5%.

Besides the CNN models have the better performance on the fewer hyperspectral data, SAEs (Ozdemir et al. 2014) and DBNs (Chen et al. 2014) are also very promising deep learning for hyperspectral classification and detection with limited training samples and augment the training data by various ways and means. Since SAE can yield promising results in hyperspectral image classification and detection, some improved SAE models are presented for better performance. Stacked sparse AE (SSAE) is an improved SAE based on sparsity constraint which is added into the hidden layers (Tao et al. 2015). The SSAE could adaptively learn discriminative feature representations from unlabeled data. Sparse spectral features and multiscale spatial features are learned from unlabeled samples based on the SSAE with two hidden layers respectively. The whole procedure could simplify the numbers of training data and learn suitable feature representations from end to end without supervision. Overall classification accuracy obtained by the proposed method tested using the Pavia Center dataset is 99.62%. Overall classification accuracy obtained by the proposed method tested using the Pavia University dataset is 98.63%. When using the small training set, the overall classification accuracies obtained by the proposed method tested using the Pavia learned from the Pavia University to classify the Pavia Center, the classification performance with 0.9699 in kappa is obtained. Also, the sparse spectral features are directly extracted by utilizing a contextual deep learning network (CDL-MLR) which is considered to be suitable for classification using relatively few training samples (Ma et al. 2015). When the number of training samples is more than 20 % of each class in the ground truth, the overall classification accuracy is trending towards stability. Overall classification accuracy obtained by CDL-MLR tested using the Indian Pines image dataset is 98.23%. Overall classification accuracy obtained by CDL-MLR tested using the Salinas image dataset is 98.26%. Overall classification accuracy obtained by CDL-MLR tested using 9% of the University of Pavia dataset is 99.86%. The weight decay is imposed on the sparse stacked AE to avoid over-fitting (Mughees et al. 2017). The joint spectral–spatial features are rotation-invariant features extracted by sparse stacked AE. The neighborhood information in images is sorted according to the difference between the first pc information of all pixels and the central pixel in the PCA image block. Then the central pixel and neighborhood information are stitched together like a 1-D input vector. 9% training samples are used. Overall classification accuracy obtained by SAE-HMRF tested using the Indian Pines dataset is 90.08%. Overall classification accuracy obtained by SAE-HMRF tested using the Pavia University dataset is 98.72%. The overall accuracies of the above improved SAE models are greater than 98% tested using the Pavia University and Center dataset.

At present, SAEs have wide applications in agriculture. The SAE combined with a fully-connected neural network (FNN) is applied to quantify nitrogen (N) concentration in oilseed rape leaf (Yu et al. 2018a). The deep spectral features are extracted by SAE trained on a large number of pixel spectra in an unsupervised manner. Then spectral features are fed to the FNN layer following the last encoding layer to fine-tune the whole model using the mean spectra with measured N concentrations in a supervised manner. Finally, SAE-FNN quantitatively predicts N concentration using the mean spectra. It is worth noting that the pixel spectra data is selected to create a big data set for deep learning, which overcomes the problem of insufficient samples. The nitrogen concentration detection dataset consists of 192 oilseed rape (Brassica napus L.) leaf samples. SAE-FNN model based upon 10 deep spectral features gave reasonable accuracy for the prediction of N concentration in oilseed rape leaf with 0.903 R^2. At the same time, they design a model based on this SAE to predict firmness and soluble solids content (SSC)

of postharvest Korla fragrant pear (Pyrus Bretschneider Rehd) ($R^2 = 0.931$) (Yu et al. 2018b), and total viable count (TVC) in peeled Pacific white shrimp ($R^2 = 0.927$) (Yu et al. 2019a). This research group also develops a DL-based model for detecting TVB-N content of Pacific white shrimp (Yu et al. 2019b). The nonlinear spectral features are likewise obtained by training SAE. The least-squares support vector machine (LS-SVM) is established based on the calculated spectral features for quantitatively predicting the corresponding TVB-N. The SAE-LS-SVM model obtained a result with R^2 value of 0.921. Similarly, the optimum wavelengths are extracted based on the SAE for detecting the cadmium residue in lettuce leaves (Zhou et al. 2020). Then the partial least squares support vector machine regression (LS-SVR) is used for Cd concentration prediction in the lettuce leaf. The Cd concentration prediction dataset is 160 leaf samples in each category under 7 different concentrations of Cd stress. The LS-SVR model obtained a result with R^2 value of 0.9487. In addition, the detection task is implemented by utilizing an SAE combined with an extreme learning machine (ELM) method to detect the firmness of apples (Rao et al. 2019). The dataset of firmness detection for apples consists of 126 Yantai red apples of similar size. SAE-ELM(19) model gave reasonable accuracy for the prediction of firmness detection for apples with 0.7703 R^2.

These works using different networks based on SAE for predicting demonstrate that the SAE could extract discriminative features and has a high potential in the multivariate analysis. For classification tasks, the feature extraction procedure via SAE in their work is the same as above, and the LR classifier is added to classify the freshness grade of shrimp (Yu et al. 2018c). The result shows that the SAE-LR algorithm could achieve satisfactory classification accuracy. The dataset of nondestructive freshness discriminating of shrimp consists of 256 hyperspectral images obtained for shrimps. The best SAEs-LR model had R^2 value of 0.858. All works manifest SAE could obtain informative features and has a better application in detection or classification tasks. The improved SAE models are also constructed for the classification of hyperspectral images. A stacked denoising AE (SDAE) combined RF classifier is built for detecting Huanglongbing of citrus (Lu et al. 2019). The SDAE model contained three-layer DAEs to extract deep spectral features which are inputted into the last layer RF. The RF classifier could take the place of Softmax classification and reverse the fine-tuning ability of BP. The DAE extracts robust features by artificially adding the noise to input hyperspectral images. The RF classifier has high classifier accuracy and stronger stability. The Huanglongbing detection dataset consists of the near-infrared spectrum data of citrus leaves provided by a company. The classification accuracy of the SDAE-RF model is 99.05%. The low-rank SAE is designed to learn spectral features for classifying different tea varieties by denoising of low-rank matrix recovery and dimensionality reduction of SAE (Sun et al. 2018a). The low-rank matrix recovery method is carried out to separate noise matrix from hyperspectral images in an optimal manner. Then SAE is trained for extracting low-rank robust features. The tea variety identification dataset consists of five kinds of tea samples containing 618 bands of hyperspectral images. The classification accuracy of the SDAE-RF model is 99.37%.

Considering that only spectral information is used in many works, the spectral features with spatial features are combined to feed the sparse AE for predicting the distribution of chlorophyll content in longan leaves (Gan et al. 2018). Firstly, the relationships between the spectral response and chlorophyll contents of the Longan leaves in three growth periods are measured based on Pearson correlation coefficient, obtaining effective spectral features, and spatial features are calculated under three feature bands. Then, the spectroscopy and texture features are imported to the spare auto-encoder

(SAE) model in deep learning to predict the chlorophyll content of Longan leaves. The distribution of chlorophyll content is predicted using the SAE model based on the mapping information. The dataset of chlorophyll content of Longan leaves measurement consists of 20 fruiting longan leaf samples (Dimocarpus longan Lour). SAE model gave reasonable accuracy for the prediction of chlorophyll content of Longan leaves with 0.6212 R^2. A stacked sparse autoencoder-deep learning network (SSAE-DLN) is proposed to predict the potassium content of citrus leaves using the hyperspectral information in four different phenological periods (Yue et al. 2019). The SSAE could extract the features in different tasks to learn adaptively nonlinear features from hyperspectral data in four phenological periods. The DLN is used to migrate and fuse different probabilities spectral features, which improves the accuracy, stability, generalization, and transferability of the model. Deep learning could describe the nutriture change (chlorophyll content, potassium content) in different growth periods. The potassium content of the citrus leave detection dataset consists of 463 citrus leaf samples. SSAE-DLNs model based on first-order differential spectral features achieves the reasonable accuracy of the inversion of potassium content for citrus leaves with 0.8771 R^2.

The early striped stem borer infestation on rice is detected using a cost-sensitive SSAE-based model whose cost function is modified for becoming sensitive towards the early infested samples (Fan et al. 2019). Specifically, this study modifies the indicator function in the cost function of SSAE to impose greater punishment on misclassification for detecting early infestation, and then SSAE could learn high-level and more sophisticated features in a hierarchical way. The improved SSAE not only executes disease detection but also focuses on detecting early infestation. The dataset of striped stem borer infestation on rice detection consists of 114 rice plants (Y Liangyou689, nonglutinous rice) grown in Zhejiang University. The detection accuracy of the cost-sensitive SSAE model is 99.98%. The application of cost-sensitive SSAE has significance in controlling diseases and should extend to more kinds of pest infestation. What's more, a sparse AE-based model is proposed for detecting TVB-N content in pork (Guo et al. 2017). Differing from the feature extraction of sparse AE in the many works, SAE explores the internal structure of the selected wavelength, getting more useful information from few wavelengths. The dataset of TVB-N measurement in pork consists of 186 chilled pork longissimus muscle samples. Root-mean-square errors of prediction

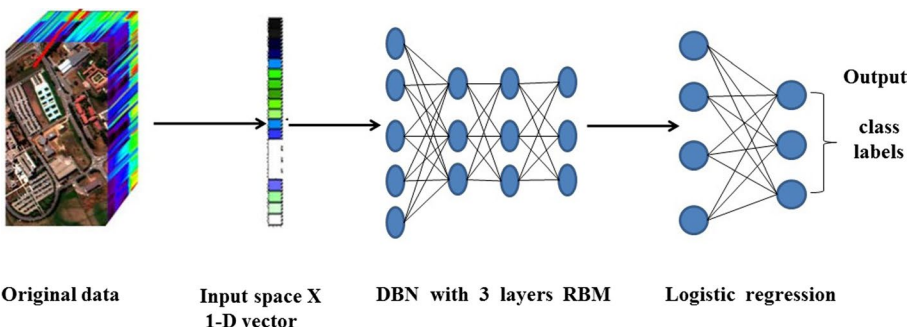

Fig. 10 The five-layers DBN structures for hyperspectral image pixel-wise classification (Chen et al., 2015a, b). Here, a DBN is employed for unsupervised feature learning and an logistic regression above the DBN is add to constitute a DBNLR framework

(RMSEP) and correlation coefficient of prediction (RP) obtained by SPA-SAE-PCA selected wavelength using PLS are 3.48 and 1.80, respectively.

3.4 Hyperspectral image analysis in agriculture using DBN models

The hyperspectral image features are extracted using deep learning model DBN within the common datasets for the first time (Fig. 10) (Chen et al. 2015b). The robust spectral, spatial, and spatial–spectral features are extracted by the five-layers DBN respectively. Specifically, the features are learned in an unsupervised way by pre-training the DBN layer-by-layer. The raw data (spectral, spatial, and spectral–spatial data) from the hyperspectral image cube is input. The output of the last hidden layer is regarded as the most representative features and a logical regression structure is added to classify the target object on top of the hidden layers. The complete framework could be deemed as a joint neural network. So when finalizing the pre-training, the classifier fine-tuned features are learned from the pre-trained DBN. It is worth noting that the spatial data are flattened to a 1-D vector in the same way with the previous SAE structure. For evaluating the classification accuracy, labeled samples are randomly divided into the training set and test set with a ratio of 1:1. The training of DBNs is complex, but they outperform other methods and are super-fast on the testing time. Overall classification accuracy obtained by DBN-LR tested using the Indian pines dataset is 95.95%. Overall classification accuracy obtained by DBN-LR tested using the Pavia dataset is 99.05%. The time obtained by DBN-LR tested using the Indian pines dataset is 0.0198s, well below that obtained by SVM. The same method with the five-layers DBN is utilized on the two real-world datasets and demonstrates that DBN has a better result in terms of classification accuracy (Fig. 11) (Sun et al. 2017). A hyperspectral image not only contains the general 2-dimensional (2-D) spatial data but also reflects the spectral information with hundreds of bands. To improve the classification efficiency, PCA is used to reduce the dimension of hyperspectral images. If the number of principal components (PCs) of a hyperspectral image is n and m × m neighbor pixels are selected, a pixel block can be represented as m × m × n members. The m × m × n tensor could be unfolded into a 1-D vector with m^2n elements being the spatial information of each class in the hyperspectral image. By combining this vector with the spatial vector, new 1-D data containing both spectral and spatial information could be available. Each pixel of a hyperspectral image can be represented by such a 1-D vector which will be the input of the DBN

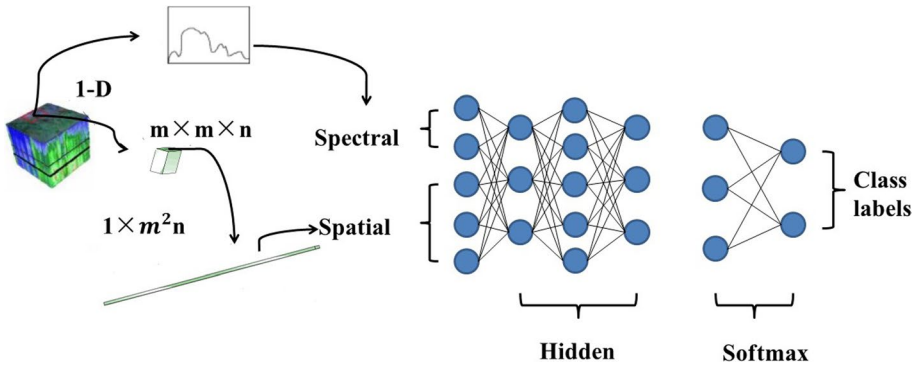

Fig. 11 The same method with Chen et al. (2015a, b) on the two real-world datasets

network. Then, each layer of the DBN can be trained without supervision. By retaining the information as much as possible, the output of each layer could be taken as the input of the next layer. Finally, a softmax is selected as the classifier in the last layer. Overall classification accuracies obtained by PCA-DBN tested using the URBAN and HYDICE datasets are 95.25% and 97.16%. Nevertheless, the pre-training and fine-tuning method in the learned DBNs would make many hidden units behave very similarly, or perform as never responding or always responding latent factor. A diversified DBN is proposed to solve the above problems (Zhong et al. 2017b). The diversified DBN introduces a diversity promoting prior over the latent factors to regularize the unsupervised pre-training and supervised fine-tuning processes, leading to the trained DBNs with more diverse latent factors. Therefore, DBNs could capture a large proportion of information to increase classification performance. The problem that "dead" (never responding) or "potential over-tolerant" (always responding) latent factors(neurons) by promoting the hyperspectral classification framework is also addressed based on an optimal DBN. 200 labeled pixels per class are randomly selected for training. Overall classification accuracy obtained by D-DBN-PF tested using the Indian pines dataset is 91.03%. Overall classification accuracy obtained by D-DBN-PF tested using the Pavia dataset is 93.11%. An advanced algorithm enhances the texture of hyperspectral images and then a two-hidden-layer DBN is developed to extract abstract features for determining the label of each testing pixel (Li et al. 2018b). 300 labeled pixels per class are randomly selected for training. This method is outperformed. Overall classification accuracy obtained by the proposed method tested using the Indian pines dataset is 97.56%. Overall classification accuracy obtained by the proposed method tested using the Pavia dataset is 96.96%. Overall classification accuracy obtained by the proposed method tested using the Salinas dataset is 96.22%.

A mass of training samples is also needed for training DBN. Considering that few labeled samples, it is also necessary to conduct sample augment before training the networks. The method of averaging the adjacent pixels by adding them together is performed to increase new samples in two manners, which not only increases training samples but also exploited the correlation of spatial data (Li and Huang 2016). One is adjacent pixels from horizontal, vertical, left-oblique, right-oblique directions of each pixel respectively acquiring fourfold samples. Another is one sample from an adjacent 3×3 size block of pixels. Differing from the common classifiers, a BP neural network algorithm which is a supervised classifier with the gradient descent algorithm is added on the top of DBNs to fine-tune the network (Lu et al. 2018). Importantly, this DBN could overcome the shortcoming of BP networks which are easy to fall into local optimum, thus achieving better classification, because of its unsupervised layer-by-layer pre-training with the contrastive divergence algorithm. A model including the Conditional Random Field (CRF) and DBN is exploited (Zhong et al. 2017). CRFs potentials are defined over the deep features learned by a DBN. The DBN-CRF model takes advantage of the strength of DBNs in learning a good representation and the ability of CRFs to model spatial information in both the observations and labels. The CRF training usually involves repeated inference, so that an end-to-end method is developed to train the joint DBN and CRF model, which meaning the parameters in DBN and CRF can be jointly trained. The overall accuracy and average accuracy of the full DBN-CRF over the Indian Pines dataset are 92.15% and 94.22%, which are much better than 88.34% and 91.53% of the DBN-CRF-U. Over the University of Pavia dataset, the full DBN-CRF obtained 94.02% overall accuracy and 94.62% average accuracy, which are also higher than 91.24% and 92.43% obtained by the DBN-CRF-U. High dimensionality and lack of labeled samples are the difficulties in feature extraction for hyperspectral

image (HSI) processing. A deep-learning-based feature extraction method is proposed (Yang et al. 2019). The result data contain the joint spectral and spatial information of the objects. The local receptive field and weight sharing are introduced into a deep Boltzmann machine(DBM, another type of deep model using RBM) to establish a novel feature extractor, called local-global DBM (LGDBM). The LGDBM has two advantages: (1) it can learn both the local and global features of the high-dimensional input data and (2) it has much fewer parameters than the DBM. Therefore, only a few labeled samples are needed for training, and the local and global spectral–spatial features are extracted intrinsically.

A few DBNs are practically applied for disease detection, which makes sense that identify the type of disease to control the disease of products. The methods developed on the hyperspectral remote image datasets could apply and improve the hyperspectral images collected in the lab. The DBN-based deep neural network is employed for detecting nondestructively various degrees of the moldy core in apple (Zhou et al. 2017). This DBN consists of the input layer, three RBM networks, and one BP layer. The BP classifier with four outputs on the last layer of DB is trained in a supervised manner and fine-tunes the whole network. Finally, the four outputs represent the disease degree of moldy core in apples, dividing health, mild, moderate, and severe degrees according to the lesion area. The moldy core dataset contains 225 apple samples including a training set of 150 samples and a test set of 75 samples. The classification accuracy of the DBN model is 88.00%. The DBN-based deep neural network could effectively overcome the deficiencies of the BP network (Li and Huang 2016). Besides, three infection levels (slight, moderate, and severe) on peach are classified by DBN (Sun et al. 2018b). The DBN is trained using both spectral and spatial information in hyperspectral images, and all 420 spatial bands are employed. 54 spatial features are extracted from one sample without deep learning methods, including 6 color features, 16 from GLCM, and 32 from histogram counts. Then, the spectral and spatial features are combined to train DBN in an unsupervised, layer-by-layer manner to obtain the classification result. The peaches fungal diseases dataset contains 300 peaches (30 samples for the control group, 270 samples for the three treated groups, 90 samples for each treated group). After the peaches were air-dried at room temperature, the treatment groups are artificially inoculated with the spore suspensions of R. stolonifera, B. cinerea, and C. acutatum, respectively, using a syringe with a steel needle. The DBN model based on integrated information (494 features) showed the highest classification results for the three diseases, with accuracies of 82.5%, 92.5%, and 100% for slightly-decayed, moderately-decayed, and severely-decayed samples, respectively.

Many DBNs are applying for agriculture outside of disease detection. A DBN is built for identifying the adulteration of lotus seed flour, which includes three RBMs and one BP layer (Hu et al. 2020). The lotus seed flour dataset contains 600 samples with different ratios of other crop flours adulteration. Based on the DBN model, lotus seed flour adultered with the other crops could be effectively identified and the average recognition rate for very few adulteration levels is about 96% (training numbers $\geq$ 600). Another DBN based on spectral and spatial features is established to detect the slight green potatoes placed randomly (Li et al. 2016). Firstly, the processing of dimensionality reduction in spatial information and spectral data is executed using maximum variance unfolding and local tangent space alignment methods respectively. Then, the spectral and spatial information is used to train DBN for obtaining the informative spectral–spatial features and getting better detection results. However, the work not specifically introduces how to fuse spectral and spatial information. The green potato detection dataset consists of

225 potato samples (green potatoes and normal potatoes). According to these tasks presented above, we can see the DBN models gradually mature in the agricultural hyperspectral images classification.

3.5 The application of RNN models in agricultural hyperspectral image analysis

Hyperspectral pixels intrinsically have a sequence-based data structure. In contrast to a feed-forward neural network, an RNN can recognize patterns in sequences of data and dynamic temporal characteristics by using a recurrent hidden state whose activation at each step depends on that of the previous steps. RNN frameworks have been used for hyperspectral image classification (Mou et al. 2017b), which can effectively analyze hyperspectral pixels as sequential data and then determine information categories via network reasoning. A modified GRU with PRetanh (RNN-GRU-PRetanh) is developed to effectively analyze hyperspectral data. The total training number of the Pavia university dataset, the Houston dataset, and the Indian Pines dataset are 3921, 2832, and 695, respectively. Overall classification accuracies of RNN-GRU-PRetanh using the Pavia university dataset, the Houston dataset, and the Indian Pines dataset are 88.85%, 89.85%, and 88.63%, respectively. The training times of the three datasets are 77.4s, 88.8s, and 19.9s, respectively. The testing efficiency is 8396.76 pixels/s.

A bidirectional-convolutional long and short term memory (Bi-CLSTM) network is used to automatically learn the spectral–spatial features from hyperspectral images, which improves the classification performance by almost 1.5 % as compared to 3D-CNN (Liu et al. 2017c). 10% pixels from each class is used as the training set Overall classification accuracies of Bi-CLSTM using the Indian Pines dataset, the Pavia university dataset and the Kennedy Space Center dataset are 96.78%, 99.1%, and 98.29%, respectively. The training times of the three datasets are 535.5s, 432s, and 112.5s, respectively. The testing times of the three datasets are 12.62s, 12.95s, and 2.65s, respectively. What's more, a deep convolutional recurrent neural network that integrates convolutional layers with recurrent layers is used for accurately identifying the disease region in the early stage (Jin et al. 2018). Specifically, the convolutional layers extract abstract features of spectra data. The contextual information of features is learned from the recurrent layer to improve the generalization of the classification model. Moreover, the input layer of the network is the 2-D grayscale pixel images reshaped based on spectral information. The last softmax layer classifies the healthy or Fusarium head blight diseased wheat. The hybrid structure could overcome the complicated environmental conditions and irregular disease patterns to improve the model with respect to robustness and generalizability. The total size of the training and validation datasets is 227,484. The total size of the testing dataset is 581,716. The training time of 2D-CNN-BidGRU is 12.4h and the testing efficiency is 3658 pixels/s. The accuracy of 2D-CNN-BidGRU is 74.3%

4 Discussion

This paper reviews and shows hyperspectral image analysis in agriculture using deep learning. Hyperspectral images containing spectral and spatial information have great prospects in agriculture such as variety classification, ripeness and component prediction, remote sensing image classification, and plant disease detection. With the development of hyperspectral imaging technology, the application of hyperspectral images in agriculture is wider

and more mature. From the deep learning models on classification or detection of agriculture using hyperspectral images, deep learning has become a widely used powerful tool for hyperspectral image analysis. Deep learning and its applications in agriculture using hyperspectral images turn to be reliable and have better accuracy. What's more, both spectral and spatial information of hyperspectral images is used to get the full advantages of deep learning. The deep learning models mentioned above could extract the abstract features for classification or detection, which have advantages over traditional methods. However, these aspects have not yet been popular in agriculture. A large amount of research focused on the classification of remote sensing images using the Indian Pines dataset, the Pavia university dataset, the Kennedy Space Center dataset, and so on.

There is no doubt that the use of deep learning and hyperspectral images is an irresistible general trend for the future of agriculture. There are some bottlenecks in the agricultural applications of hyperspectral image analysis based on deep learning. It is thus necessary to consider the below points in future agricultural works.

1. The rich spectral and spatial information is utilized in hyperspectral image analysis. A series of 3D CNNs and contextual deep CNNs have been successfully applied on HSI to conjunctively fuse the spatial features with spectral information. It is crucial that how to make full use of features according to the target of hyperspectral image analysis. Ensuring overall accuracy while improving computation efficiency is very important. By preserving as many as possible the original spectral–spatial constraints, tensor representation helps to reduce the number of unknown parameters used in learning a linear dimensionality reduction model. Tensor-based learning will be a promising technique for hyperspectral image analysis.
2. High dimensional data and the limited number of labeled samples are the most bottlenecks in hyperspectral image analysis. Transfer learning, generative adversarial networks, semi-supervised learning, and active learning are promising techniques to respond to the challenge of limited labeled training samples. Transfer learning introduces the useful information learned from the source data to the target data, which can significantly decrease the demand for training samples. Transfer learning can effectively solve the degradation of network performance when the available training hyperspectral images are limited in agriculture. GANs utilize adversarial training to generate the hyperspectral data based on the required class label. Semi-supervised learning can jointly exploit labeled and unlabeled samples to deal with the problem of dimensionality. Active learning models achieve higher accuracy with fewer training samples by actively selecting training samples. Hyperspectral data inherently owns high dimensionality, which increases the computational complexity of hyperspectral image analysis. Transfer learning and active learning can also respond to the challenges of high-dimensionality and superresolution characteristics of hyperspectral images.
3. The methods of reducing the interferences are important for modeling. The acquisition processes of hyperspectral images are affected by the collecting condition, especially the lighting of the scene. The deep learning models trained with a mass of samples could reduce the effect but not separate the noise from informative data. The denoising methods should be explored in agricultural applications.

5 Conclusions

Deep learning applying to the hyperspectral images in agriculture is a newly emerging direction. In this paper, we summarize its applications in agriculture include variety classification, ripeness and component prediction, remote sensing image classification, and plant disease detection. Then, we review the recent achievements in deep learning-based hyperspectral image analysis from the aspects of the deep learning models and the feature networks. The critical challenges of deep learning applied for hyperspectral image analysis are analyzed, especially CNN. These methods for addressing the challenges are discussed in detail. Finally, we discuss some bottlenecks in the agricultural applications of hyperspectral image analysis based on deep learning that could be the basis for future work in this area.

Acknowledgements This work was supported by NSFC (Nos. 31871543, 31700644), Natural Science Foundation of Shandong (No. ZR2020KF002), and the project of Shandong provincial key laboratory of horticultural machinery and equipment (No. YYJX201905). The authors are grateful to all study participants. The authors declared that they have no conflicts of interest in this work. We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Declaration

Conflict of interest The authors declare that they have no conflict of interest.

References

Adriana R, Carlo G, Gustau C (2015) Unsupervised deep feature extraction for remote sensing image classification. Remote Sens 54(3):1349–1362

Ahmad M, Shabbir S, Oliva D, Mazzara M, Distefano S (2020a) Spatial-prior generalized fuzziness extreme learning machine autoencoder-based active learning for hyperspectral image classification. Optik 206:163712

Ahmad M, Mazzara M, Raza RA, Distefano S, Sohaib A (2020b) Multiclass non-randomized spectral–spatial active learning for hyperspectral image classification. Appl Sci 10(14):4739

Athanasios V, Nikolaos D, Anastasios D, Eftychios P (2018) Deep learning for computer vision: a brief review. Comput Intell Neurosci 2018:1–13

Awad MM (2019) An innovative intelligent system based on remote sensing and mathematical models for improving crop yield estimation. Inf Process Agric 6(3):316–325

Badrinarayanan V, Kendall A, Cipolla R (2017) Segnet: a deep convolutional encoder–decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell 39(12):2481–2495

Bengio Y, Lamblin P, Popovici D, Larochelle H (2007) Greedy layer-wise training of deep networks. Adv Neural Inf Process Syst 19:153

Bengio Y, Courville A, Vincent P (2012) Representation learning: a review and new perspectives. IEEE Trans Pattern Anal Mach Intell 35(8):1798–1828

Bhardwaj K, Das A, Patra S (2020a) Spectral–spatial active learning with superpixel profile for classification of hyperspectral images. In: 2020 6th international conference on signal processing and communication (ICSC), pp 149–155. https://doi.org/10.1109/ICSC48311.2020.9182764

Bhardwaj K, Das A, Patra S (2020b) Spectral–spatial active learning with attribute profile for hyperspectral image classification. In: International conference on intelligent computing and smart communication 2019. Springer, Singapore, pp 1219–1229. https://doi.org/10.1007/978-981-15-0633-8_119

Bharti R, Saini D, Malik R (2021) A novel approach for hyper spectral images using transfer learning. IOP Conf Ser Mater Sci Eng 1022(1):012120

Bioucas-Dias JM, Plaza A, Camps-Valls G, Scheunders P, Nasrabadi N, Chanussot J (2013) Hyperspectral remote sensing data analysis and future challenges. IEEE Geosci Remote Sens Mag 1(2):6–36

Cao XY, Yao J, Xu ZB, Meng DY (2020) Hyperspectral image classification with convolutional neural network and active learning. IEEE Trans Geosci Remote Sens 58(7):4604–4616. https://doi.org/10.1109/TGRS.2020.2964627

Caballero D, Calvini R, Amigo JL (2020) Hyperspectral imaging in crop fields: precision agriculture. Data Handl Sci Technol 32:453–473

Chen YS, Lin ZH, Zhao X, Wang G, Gu YF (2014) Deep learning-based classification of hyperspectral data. IEEE J Sel Top Appl Earth Obs Remote Sens 7(6):2094–2107

Chen L, Papandreou G, Kokkinos I, Murphy K, Yuille AL (2015a) Semantic image segmentation with deep convolutional nets and fully connected CRFs. Paper presented at the international conference on learning representations 40(4):834–848

Chen YS, Zhao X, Jia XP (2015b) Spectral–spatial classification of hyperspectral data based on deep belief network. IEEE J Sel Top Appl Earth Obs Remote Sens 8(6):2381–2392

Chen Y, Jiang H, Li C, Jia X, Ghamisi P (2016a) Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. IEEE Trans Geosci Remote Sens 54(10):6232–6251

Chen X, Ma L, Yang XQ (2016b) Stacked denoise autoencoder based feature extraction and classification for hyperspectral images. Journal of Sensors 3632943. http://dx.doi.org/10.1155/2016/3632943

Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL (2017a) Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans Pattern Anal Mach Intell 40(4):834–848

Chen ZY, Li WH (2017) Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network. IEEE Trans Instrum Meas 99:1–10

Chen L, Papandreou G, Schroff F, Adam H (2017b) Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587

Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H (2018) Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European conference on computer vision (ECCV). Springer, Cham, pp 833–851. https://doi.org/10.1007/978-3-030-01234-2_49

Chen FJ, Li JM, Yang DY (2019) Hyperspectral image classification based on generative adversarial networks. Comput Eng Appl 55(22):172–179

Chung J, Gulcehre C, Cho K H et al (2014) Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555

Cui Y, Ji XW, Xu K, Wang LG (2019) A double-strategy-check active learning algorithm for hyperspectral image classification. Photogramm Eng Remote Sens 85(11):841–851

Dan C, Meier U, Masci J, Gambardella LM, Schmidhuber J (2011) Flexible, high performance convolutional neural networks for image classification. In: Proceedings of the 22nd international joint conference on artificial intelligence, pp 1237–1242. https://doi.org/10.5591/978-1-57735-516-8/IJCAI11-210

Fan YY, Zhang C, Liu ZY, Qiu ZJ, He Y (2019) Cost-sensitive stacked sparse auto-encoder models to detect striped stem borer infestation on rice based on hyperspectral imaging. Knowl-Based Syst 168(2019):49–58

Fauvel M, Tarabalka Y, Benediktsson JA, Chanussot J, Tilton JC (2013) Advances in spectral–spatial classification of hyperspectral images. Proc IEEE 101(3):652–675

Feng Z, Wang M, Yang S (2017) Superpixel tensor sparse coding for structural hyperspectral image classification. IEEE J Sel Top Appl Earth Obs Remote Sens 4:1–8

Fricker GA, Ventura JD, Wolf JA, North MP (2019) A convolutional neural network classifier identifies tree species in mixed-conifer forest from hyperspectral imagery. Remote Sens 11(19):2326–2347

Gan HM, Yue XJ, Hong TS, Ling KJ, Wang LH, Cen ZZ (2018) A hyperspectral inversion model for predicting chlorophyll content of Longan leaves based on deep learning. J South China Agric Univ 39(3):102–110

Gao H, Yao D, Wang M et al (2019) A hyperspectral image classification method based on multi-discriminator generative adversarial networks. Sensors 19(15):3269

Garcia-Garcia A, Orts-Escolano S, Oprea SO, Villena-Martinez V, Garcia-Rodriguez J (2017) A review on deep learning techniques applied to semantic segmentation. arXiv preprint arXiv:1704.06857

Guo TF, Huang M, Zhu QB, Guo Y (2017) Hyperspectral image-based spare autoencoder network for TVB-N measurement in pork. In: 2017 ASABE annual international meeting, 1700450. https://doi.org/10.13031/aim.201700450

Guo YM, Liu Y, Ard O, Lao SY (2016) Deep learning for visual understanding: a review. Neurocomputing 187(Apr.26):27–48

Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. In: Proceedings of the international conference on neural information processing systems, Montreal, QC, Canada, pp 2672–2680

Han LQ, Zhang YN, Qin QM (2019) Endmember extraction of farmland hyperspectral image using deep learning autoencoder and shuffled frog leaping algorithm. Trans Chin Soc Agric Eng 35(6):167–173
He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp 770–778. https://doi.org/10.1109/CVPR.2016.90
He Z, Hu J, Wang YW (2018) Low-rank tensor learning for classification of hyperspectral image with limited labeled samples. Signal Process 145(2018):12–25
He ZP, Xia KW, Li TJ, Zu BK, Yin ZX, Zhang JN (2021) A constrained graph-based semi-supervised algorithm combined with particle cooperation and competition for hyperspectral image classification. Remote Sens 13(2):193
He Z, Liu H, Wang YW, Hu J (2017) Generative adversarial networks-based semi-supervised learning for hyperspectral image classification. Remote Sens 9(10):1042
Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. Science 313(5786):504–507
Hinton GE, Osindero S, Teh YW (2006) A fast learning algorithm for deep belief nets. Neural Comput 18(7):1527–1554
Hinton G, Deng L, Yu D, Dahl GE, Mohamed AR, Jaitly N, Senior A, Vanhoucke V, Nguyen P, Sainath T (2012) Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. IEEE Signal Process Mag 29(6):82–97
Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780
Hu W, Huang YY, Wei L, Zhang F, Li HC (2015) Deep convolutional neural networks for hyperspectral image classification. J Sens 2015:1–12
Hu RW, Yu Y, Ni ML, Yu J, Zhao JW, Zhu C, Li ZM (2020) Identification of lotus seed flour adulteration based on near-infrared spectroscopy combined with deep belief network. Food Sci 41(06):298–303
Huang SP, Sun C, Qi L, Ma X, Wang WJ (2017) Rice panicle blast identification method based on deep convolution neural network. Trans Chin Soc Agric Eng 33(20):169–176
Huang Y, Tang LB, Li Z, Long T (2019b) Research on peanut planting area classification technology using remote sensing image based deep learning. J Signal Process 35(4):617–622
Huang FH, Yu Y, Feng TH (2019) Hyperspectral remote sensing image change detection based on tensor and deep learning. J Vis Commun Image Represent 58(JAN.):233–244
Ishida T, Kurihara J, Viray FA, Namuco SB, Marciano JJ (2018) A novel approach for vegetation classification using UAV-based hyperspectral imaging. Comput Electron Agric 144:80–85
Jaime Z, Ren JC, Zheng JB, Zhao HM, Qing CM, Yang ZJ, Stephen M (2016) Novel segmented stacked auto-encoder for effective dimensionality reduction and feature extraction in hyperspectral imaging. Neurocomputing 185(12):1–10
Jamshidpour N, Safari A, Homayouni S (2020) A GA-based multi-view, multi-learner active learning framework for hyperspectral image classification. Remote Sens 12(2):297
Ji SP, Zhang C, Xu AJ, Shi Y, Duan YL (2018) 3D convolutional neural networks for crop classification with multi-temporal remote sensing images. Remote Sens 10(2):75–92
Jiao LC, Liang MM, Chen H, Yang SY, Liu HY, Cao XH (2017) Deep fully convolutional network-based spatial distribution prediction for hyperspectral image classification. IEEE Trans Geosci Remote Sens 55(10):5585–5599
Jiang ZC, Pan WD, Shen H (2018) LSTM based adaptive filtering for reduced prediction errors of hyperspectral images. In: 2018 6th IEEE international conference on wireless for space and extreme environments (WISEE), pp 158–162. https://doi.org/10.1109/WiSEE.2018.8637354
Jiang XF, Zhang Y, Li Y, Li SY, Zhang Y (2019) Hyperspectral image classification with transfer learning and Markov random fields. IEEE Geosci Remote Sens Lett 17(3):544–548. https://doi.org/10.1109/LGRS.2019.2923647
Jiang YP, Chen SF, Bian B, Li YH, Sun Y, Wang XC (2021) Discrimination of tomato maturity using hyperspectral imaging combined with graph-based semi-supervised method considering class probability information. Food Anal Methods 2:1–16
Jin G, Raich R (2012) On surrogate supervision multiview learning. In: IEEE international workshop on machine learning for signal processing (MLSP), pp 1–6. https://doi.org/10.1109/MLSP.2012.6349759
Jin X, Jie L, Wang S, Qi H, Li S (2018) Classifying wheat hyperspectral pixels of healthy heads and fusarium head blight disease using a deep neural network in the wild field. Remote Sens 10(3):395–415
Jin X, Lu J, Fu YZ, Wang S, Xu GJ, Li SW (2019) A classification method for hyperspectral imaging of Fusarium head blight disease symptom based on deep convolutional neural network. Acta Agriculturae Zhejiangensis 31(2):315–325
Kemker R, Kanan C (2017) Self-taught feature learning for hyperspectral image classification. Remote Sens 55(5):2693–2705

Kong Y, Wang XS, Cheng YH (2018) Spectral–spatial feature extraction for HSI classification based on supervised hypergraph and sample expanded. IEEE J Sel Top Appl Earth Obs Remote Sens 11(11):4128–4140

Krizhevsky A, Sutskever I, Hinton G (2012) ImageNet classification with deep convolutional neural networks. In: Proceedings of the 25th international conference on neural information processing systems, pp 1097–1105

Kumar S, Torres C, Ulutan O et al (2020) Deep remote sensing methods for methane detection in overhead hyperspectral imagery. In: Winter conference on applications of computer vision (WACV), pp 1765–1774. https://doi.org/10.1109/WACV45572.2020.9093600

Kuska M, Wahabzada M, Leucker M, Dehne HW, Kersting K, Oerke EC, Steiner U, Mahlein AK (2015) Hyperspectral phenotyping on the microscopic scale: towards automated characterization of plant pathogen interactions. Plant Methods 11(1):28–41

Lecun Y, Bottou L (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324

Lee H, Heesung K (2016) Contextual deep CNN based hyperspectral classification. In: 2016 IEEE international geoscience and remote sensing symposium (IGARSS)

Lee H, Kwon H (2017) Going deeper with contextual CNN for hyperspectral image classification. IEEE Trans Image Process 26(10):4843–4855

Lei Z, Zeng Y, Liu P, Su XH (2021) Active deep learning for hyperspectral image classification with uncertainty learning. IEEE Geosci Remote Sens Lett 99:1–5

Li Y, Zhang HK, Shen Q (2017a) Spectral–spatial classification of hyperspectral imagery with 3-D convolutional neural network. Remote Sens 1(9):67–87

Li W, Wu GD, Zhang F, Du Q (2017b) Hyperspectral image classification using deep pixel-pair features. IEEE Trans Geosci Remote Sens 52(2):844–853

Li JJ, Zhao X, Li YS, Du Q, Xi BB, Hu J (2018a) Classification of hyperspectral imagery using a new fully convolutional neural network. IEEE Geosci Remote Sens Lett 15(2):292–296

Li JJ, Xi BB, Li YS, Du Q, Wang KY (2018b) Hyperspectral classification based on texture feature enhancement and deep belief networks. Remote Sens 10(3):396–416

Li S, Song W, Fang L, Chen Y, Benediktsson JA (2019) Deep learning for hyperspectral image classification: an overview. IEEE Trans Geosci Remote Sens 57(9):6690–6709

Li Y, Lu T, Li ST (2020) Subpixel-pixel-superpixel-based multiview active learning for hyperspectral images classification. IEEE Trans Geosci Remote Sens 99:1–13

Li XG, Huang XQ (2016a) Deep neural networks based on hyperspectral image classification. Electron Meas Technol 39(7):81–86

Li XY, Ku J, Yan YY, Xu ML, Xu SM, Jin R (2016b) Detection method of green potato based on hyperspectral imaging. Trans Chin Soc Agric Mach 47(3):228–233

Li J, Jose M (2013) Semi-supervised hyperspectral image classification using soft sparse multinomial logistic regression. IEEE Lett Geosci Remote Sens 10:318–322

Lin ZH, Chen Y, Zhao X, Wang G (2014) Spectral–spatial classification of hyperspectral image using autoencoders. In: 2013 9th international conference on information, communications and signal processing, pp 1–5. https://doi.org/10.1109/ICICS.2013.6782778

Liu B, Yu XC, Zhang PQ, Tan X, Yu AZ, Xue ZX (2017b) A semi-supervised convolutional neural network for hyperspectral image classification. Remote Sens Lett 8(9):839–848

Liu XF, Sun QQ, Liu B, Huang B (2017a) Hyperspectral image classification based on convolutional neural network and dimension reduction. In: 2017 Chinese Automation Congress (CAC), pp 1686–1690. https://doi.org/10.1109/CAC.2017.8243039

Liu QS, Zhou F, Hang RL, Yuan XT (2017c) Bidirectional-convolutional lstm based spectral-spatial feature learning for hyperspectral image classification. Remote Sens 9(12):1330

Liu P, Zhang H, Eom KB (2017d) Active deep learning for classification of hyperspectral images. IEEE J Sel Top Appl Earth Obs Remote Sens 10(2):712–724

Liu XF, Sun QQ, Meng Y, Wang CC, Fu M (2018a) Feature extraction and classification of hyperspectral image based on 3D-convolution neural network. In: 2018 IEEE 7th data driven control and learning systems conference, pp 918–922. https://doi.org/10.1109/DDCLS.2018.8515930

Liu B, Yu XC, Yu AZ, Zhang PQ, Wan G, Wang RR (2018b) Deep few-shot learning for hyperspectral image classification. IEEE Trans Geosci Remote Sens 99:1–15

Liu B, Yu XC, Zhang PQ, Yu AZ, Fu QY, Wei XP (2018c) Supervised deep feature extraction for hyperspectral image classification. IEEE Trans Geosci Remote Sens 56(4):1909–1921

Liu CY, He L, Li ZT, Li J (2018d) Feature-driven active learning for hyperspectral image classification. IEEE Trans Geosci Remote Sens 56(1):341–354

Liu JJ, Wu ZB, Xiao L, Sun J, Yan H (2019) Generalized tensor regression for hyperspectral image classification. IEEE Trans Geosci Remote Sens 58(2):1244–1258
Liu CL, Lin L, Yu CC, Wu JZ (2020a) Research on peanut hyperspectral image classification method based on deep learning. Comput Simul 37(03):189–192
Liu X, Qiao Y, Xiong Y, Cai Z, Liu P (2020b) Cascade conditional generative adversarial nets for spatial–spectral hyperspectral sample generation. Sci China Inf Sci 63(4):1–16
Liu Z, Jiang J, Qiao X et al (2020) Using convolution neural network and hyperspectral image to identify moldy peanut kernels. LWT Food Sci Technol 132:109815
Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. IEEE Trans Pattern Anal Mach Intell 39(4):640–651
Lu W, Guo YM, Dai DJ, Zhang CY, Wang XY (2018) Rice germination rate detection based on fluorescent spectrometry and deep belief network. Spectrosc Spectr Anal 38(4):1303–1312
Lu HX, Wei MM, Yang HH, Liu ZB, Hu JQ (2019) Detecting huanglongbing by stacked denoising auto-encoders combined random forest. Laser Infrared 49(4):460–466
Luo JH, Li MQ, Zhang ZZ, Li J (2017) Hyperspectral remote sensing images classification using a deep convolutional neural network model. J Xihua Univ 36(4):13–20
Ma XR, Geng J, Wang HY (2015) Hyperspectral image classification via contextual deep learning. Eurasip J Image Video Process 1:20
Makantasis K, Karantzalos K, Doulamis A, Loupos K (2015a) Deep learning-based man-made object detection from hyperspectral data. In: International symposium on visual computing, pp 717–727. https://doi.org/10.1007/978-3-319-27857-5_64
Makantasis K, Doulamis ND, Nikitakis A, Doulamis AD (2018a) Tensor-based classification models for hyperspectral data analysis. IEEE Trans Geosci Remote Sens 56(12):6884–6898
Mou L, Ghamisi P, Zhu XX (2017a) Unsupervised spectral–spatial feature learning via deep residual conv–deconv network for hyperspectral image classification. IEEE Trans Geosci Remote Sens 56(1):391–406
Mou LC, Ghamisi P, Zhu XX (2017b) Deep recurrent neural networks for hyperspectral image classification. IEEE Trans Geosci Remote Sens 55(7):3639–3655
Mu CH, Liu J, Liu Y, Liu YJ (2020) Hyperspectral image classification based on active learning and spectral–spatial feature fusion using spatial coordinates (October 2019). IEEE Access 8:11
Mughees A, Tao L (2017) Hyperspectral image classification based on deep auto-encoder and hidden Markov random field. In: 13th international conference on natural computation, fuzzy systems and knowledge discovery (ICNC-FSKD), pp 59–65. https://doi.org/10.1109/FSKD.2017.8393336
Munoz-Mari J, Bovolo F, Gmez-Chova L, Bruzzone L, Camp-Valls G (2010) Semi-supervised one-class support vector machines for classification of remote sensing data. IEEE Trans Geosci Remote Sens 48(8):3188–3197
Murphy JM, Maggioni M (2018) Unsupervised clustering and active learning of hyperspectral images with nonlinear diffusion. IEEE Trans Geosci Remote Sens 57(3):1829–1845
Nagasubramanian K, Jones S, Singh AK, Singh A, Ganapathysubramanian B, Sarkar S (2017) Explaining hyperspectral imaging based plant disease identification: 3D CNN and saliency maps. In: 31st conference on neural information processing systems (NIPS, 2017) 4–9 December 2017, Long Beach, CA, USA
Nataliia K, Mykola L, Sergii S, Andrii S (2017) Deep learning classification of land cover and crop types using remote sensing data. IEEE Geosci Remote Sens Lett 14(5):778–782
Ni C, Li ZY, Zhang X, Zhao L, Zhu TT, Jiang XS (2019) Film sorting algorithm in seed cotton based on near-infrared hyperspectral image and deep learning. Trans CSAE 50(12):170–179
Nie P, Zhang J, Feng X, Yu C, He Y (2019) Classification of hybrid seeds using near-infrared hyperspectral imaging technology combined with deep learning. Sens Actuators B Chem 296:126630–126641
Niu Z, Liu W, Zhao J et al (2019) DeepLab-based spatial feature extraction for hyperspectral image classification. IEEE Geosci Remote Sens Lett 16(2):251–255
Ozdemir AO, Gedik BE, Cetin YY (2014) Hyperspectral classification using stacked autoencoders with deep learning. In: Proceedings of the 2014 6th workshop on hyperspectral image and signal processing: evolution in remote sensing (WHISPERS), pp 1–4. https://doi.org/10.1109/WHISPERS.2014.8077532
Pan B, Shi ZW, Xu X (2017) R-VCANet: a new deep-learning-based hyperspectral image classification method. IEEE J Sel Top Appl Earth Obs Remote Sens 10(5):1975–1986
Paoletti ME, Haut JM, Plaza J, Plaza A (2018) A new deep convolutional neural network for fast hyperspectral image classification. ISPRS J Photogramm Remote Sens 145(2018):120–147

Paoletti ME, Haut JM, Plaza J, Plaza A (2020) Training Capsnets via active learning for hyperspectral image classification. In: IGARSS 2020-2020 IEEE international geoscience and remote sensing symposium, pp 40–43. https://doi.org/10.1109/IGARSS39084.2020.9324302

Petersson H, Gustafsson D, Bergstrom D (2016) Hyperspectral image analysis using deep learning: a review. In: International conference on image processing theory, tools and applications, pp 1–6. https://doi.org/10.1109/IPTA.2016.7820963

Pound MP, Atkinson JA, Townsend AJ, Wilson MH, Griffiths M, Jackson AS, Bulat A, Tzimiropoulos G, Wells DM, Murchie EH, Pridmore TP, French AP (2017) Deep machine learning provides state-of-the-art performance in image-based plant phenotyping. Gigascience 6(10):1–10

Pradhan MK, Minz S, Shrivastava VK (2019a) A kernel-based extreme learning machine framework for classification of hyperspectral images using active learning. J Indian Soc Remote Sens 47(3):1693

Pradhan MK, Minz S, Shrivastava VK (2019b) Fast active learning for hyperspectral image classification using extreme learning machine. IET Image Process 13(4):549–555

Protopapadakis E, Doulamis A, Doulamis N, Maltezos E (2021) Stacked autoencoders driven by semi-supervised learning for building extraction from near infrared remote sensing imagery. Remote Sens 13(3):371

Rao LB, Pang T, Ji RS, Chen XY, Zhang J (2019) Firmness detection for apples based on hyperspectral imaging technology combined with stack autoencoder-extreme learning machine method. Laser Optoelectron Progr 56(11):113001-1-113001–7

Ronneberger O, Fischer P, Brox T (2015) U-Net: convolutional networks for biomedical image segmentation. Paper presented at the medical image computing and computer assisted intervention 759:195–202

Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang ZH, Karpathy A, Khosla A, Bernstein M, Berg AC, Li FF (2015) ImageNet large scale visual recognition challenge. Int J Comput Vis 115(3):211–252

Sawant SS, Prabukumar M (2018) A review on graph-based semi-supervised learning methods for hyperspectral image classification. Egypt J Remote Sens Space Sci 23(2):243–248

Simonyan K, Zisserman A (2015) Very deep convolutional networks for arge-scale image recognition. In: International conference on learning representations, 7–9 (May 2015) San Diego, CA, pp 1–14

Simranjit S, Singara SK (2018) Efficient classification of the hyperspectral images using deep learning. Multimed Tools and Applications 77:27061–27074

Song A, Choi J, Han Y, Kim Y (2018) Change detection in hyperspectral images using recurrent 3D fully convolutional networks. Remote Sens 10(11):1827

Slavkovikj V, Verstockt S, De Neve W, Van Hoecke S, Van De Walle R (2015) Hyperspectral image classification with convolutional neural networks. In: Proceedings of the 23rd annual ACM conference on multimedia conference, 13–15 October 2015, Brisbane, Australia, pp 1159–1162

Steinbrener J, Posch K, Leitner R (2019) Hyperspectral fruit and vegetable classification using convolutional neural networks. Comput Electron Agric 162:364–372

Sun Z, Wang C (2014) Semi-supervised classification for hyperspectral imagery with transductive multiple-kernel learning. IEEE Lett Geosci Remote Sens 11:1991–1995

Sun QQ, Liu XF, Fu M (2017) Classification of hyperspectral image based on principal component analysis and deep learning. In: 2017 7th IEEE international conference on electronics information and emergency communication (ICEIEC), pp 356–359. https://doi.org/10.1109/ICEIEC.2017.8076581

Sun J, Jin HT, Wu XH, Lu H, Shen JF, Dai CX (2018a) Tea variety identification based on low-rank stacked auto-encoder and hyperspectral image. Trans CSAE 49(8):316–323

Sun Y, Wei KL, Liu Q, Pan LQ, Tu K (2018b) Classification and discrimination of different fungal diseases of three infection levels on peaches using hyperspectral reflectance imaging analysis. Sensors 18(4):1295–1308

Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich, A (2014) Going deeper with convolutions. In: 2015 IEEE conference on computer vision and pattern recognition (CVPR), pp 1–9. https://doi.org/10.1109/CVPR.2015.7298594

Tao C, Pan HB, Li YS, Zou ZR (2015) Unsupervised spectral–spatial feature learning with stacked sparse autoencoder for hyperspectral imagery classification. IEEE Geosci Remote Sens Lett 12(12):2438–2442

Tasissa A, Nguyen D, Murphy J (2021) Deep diffusion processes for active learning of hyperspectral images. arXiv preprint arXiv:2101.03197

Tian YP, Tao C, Zou ZR, Yang ZX, He XF (2015) Semi-supervised graph-based hyperspectral image classification with active learning. Acta Geodaetica et Cartographica Sinica 44(8):919–926

Tuia D, Camps-Valls G (2009) Semi-supervised remote sensing image classification with cluster kernels. IEEE Geosci Remote Sens Lett 6(2):224–228

Vincent P, Larochelle H, Bengio Y, Manzagol PA (2008) Extracting and composing robust features with denoising autoencoders. In: Proceedings of the 25th international conference on machine learning, pp 1096–1103. https://doi.org/10.1145/1390156.1390294

Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol PA (2010) Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. J Mach Learn Res 11(12):3371–3408

Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E (2018) Deep learning for computer vision: a brief review. Comput Intell Neurosci 2018:1–13

Wang C, Liu Y, Bai X, Tang WZ, Lei P, Zhou J (2017) Deep residual convolutional neural network for hyperspectral image super-resolution. In: International conference on image and graphics, pp 370–380. https://doi.org/10.1007/978-3-319-71598-8_33

Wang DY, Vinson R, Holmes M, Seibel G, Bechar A, Nof S, Tao Y (2019) Early detection of tomato spotted wilt virus by hyperspectral imaging and outlier removal auxiliary classifier generative adversarial nets (OR-AC-GAN). Sci Rep 9(1):4377–4390

Wang HY, Li XF, Li YB, Sun YX, Xu HL (2020) Non-destructive detection of apple multi-quality parameters based on hyperspectral imaging technology. J Nanjing Agric Univ 43(1):178–185

Williams Z (1989) A learning algorithm for continually running fully recurrent neural networks. Neural Comput 1(2):270–280. https://doi.org/10.1162/neco.1989.1.2.270

Wu H, Saurabh P (2017) Semi-supervised deep learning using pseudo labels for hyperspectral image classification. IEEE Trans Image Process 27(3):1259–1270

Xie ZZ, Xu HL, Huang QG, Wang P (2019) Spinach freshness detection based on hyperspectral image and deep learning method. Trans Chin Soc Agric Eng 35(13):277–284

Xie F, Gao Q, Jin C, Zhao F (2021) Hyperspectral image classification based on superpixel pooling convolutional neural network with transfer learning. Remote Sens 13(5):930

Xu YH, Bo D, Zhang LP (2019) Beyond the patchwise classification: spectral–spatial fully convolutional networks for hyperspectral image classification. IEEE Trans Big Data 6(3):492–506

Xue ZX (2020) A general generative adversarial capsule network for hyperspectral image spectral–spatial classification. Remote Sens Lett 11(1):19–28

Yang JX, Zhao YQ, Chan CW, Chen Y (2016) Hyperspectral image classification using two-channel deep convolutional neural network. In: 2016 IEEE international geoscience and remote sensing symposium (IGARSS), pp 5079–5082. https://doi.org/10.1109/IGARSS.2016.7730324

Yang GG, Bao YD, Liu ZY (2017) Localization and recognition of pests in tea plantation based on image saliency analysis and convolutional neural network. Trans Chin Soc Agric Eng 33(6):156–162

Yang XF, Ye YM, Li XT, Lau R, Zhang XF, Huang XH (2018a) Hyperspectral image classification with deep learning models. IEEE Trans Geosci Remote Sens 56(9):5408–5423

Yang G, Gewali UB, Ientilucci E et al (2018b) Dual-channel densenet for hyperspectral image classification. In: 2018 IEEE international geoscience and remote sensing symposium, pp 2595–2598. https://doi.org/10.1109/IGARSS.2018.8517520

Yang JG, Guo YH, Wang XL (2019) Feature extraction of hyperspectral images based on deep Boltzmann machine. IEEE Geosci Remote Sens Lett 17(6):1077–1081. https://doi.org/10.1109/LGRS.2019.2937601

Yi K, Wang X, Cheng Y, Chen C (2018) Hyperspectral imagery classification based on semi-supervised broad learning system. Remote Sens 10(5):685

Yoo Hyeon-Joong (2015) Deep convolution neural networks in computer vision. IEIE Trans Smart Process Comput 4(1):35–43

Yu XJ, Lu HD, Liu QY (2018a) Deep-learning-based regression model and hyperspectral imaging for rapid detection of nitrogen concentration in oilseed rape (*Brassica napus* L.) leaf. Chemometr Intell Lab Syst 172(2018):188–193

Yu XJ, Lu HD, Wu D (2018b) Development of deep learning method for predicting firmness and soluble solid content of postharvest Korla fragrant pear using Vis/NIR hyperspectral reflectance imaging. Postharvest Biol Technol 141(2018):39–49

Yu XJ, Tang L, Wu XF, Lu HD (2018c) Nondestructive freshness discriminating of shrimp using visible/near-Infrared hyperspectral Imaging technique and deep learning algorithm. Food Anal Methods 11:768–780

Yu XJ, Yu X, Wen ST, Yang JQ, Wang JP (2019a) Using deep learning and hyperspectral imaging to predict total viable count (TVC) in peeled Pacific white shrimp. J Food Meas Charact 2(2019):2082–2094

Yu XJ, Wang JP, Wen ST, Yang JQ, Zhang FF (2019b) A deep learning based feature extraction method on hyperspectral images for nondestructive prediction of TVB-N (total volatile basic nitrogen (TVB-N) content in Pacific white shrimp (*Litopenaeus vannamei*). Biosyst Eng 178(2019):244–255

Yuan Y, Zheng X, Lu X (2017) Hyperspectral image superresolution by transfer learning. IEEE J Sel Top Appl Earth Obs Remote Sens 5:1–12

Yuan QQ, Zhang Q, Li J, Shen HF, Zhang LP (2018) Hyperspectral image denoising employing a spatial-spectral deep residual convolutional neural network. IEEE Trans Geosci Remote Sens 57(2):1205–1218

Yue J, Zhao WZ, Mao SJ, Liu H (2015) Spectral–spatial classification of hyperspectral images using deep convolutional neural networks. Remote Sens Lett 6(6):468–477
Yue J, Mao SJ, Mei L (2016) A deep learning framework for hyperspectral image classification using spatial pyramid pooling. Remote Sens Lett 7(9):875–884
Yue XJ, Ling KJ, Wang LH, Cen ZZ, Lu Y, Liu YX (2019) Inversion of potassium content for citrus leaves based on hyperspectral and deep transfer learning. Trans CSAE 50(3):186–195
Zeiler M, Fergus R (2014) Visualizing and understanding convolutional neural networks. ECCV 2014, pp 818–833
Zhan Y, Hu D, Wang Y, Yu X (2018) Semisupervised hyperspectral image classification based on generative adversarial networks. IEEE Geosci Remote Sens Lett 15(2):212–216
Zhang L, Zhang L, Tao D, Huang X (2013) Tensor discriminative locality alignment for hyperspectral image spectral–spatial feature extraction. IEEE Trans Geosci Remote Sens 51(1):242–256
Zhang X (2014) Modified co-training with spectral and spatial views for semi-supervised hyperspectral image classification. Appl Earth Obs Remote Sens IEEE 7:2044–2055
Zhang C, Guo C, Liu F, Kong W, He Y, Lou B (2016) Hyperspectral imaging analysis for ripeness evaluation of strawberry with support vector machine. J Food Eng 179(2016):11–18
Zhang HK, Li Y, Zhang YZ, Shen Q (2017) Spectral-spatial classification of hyperspectral imagery using a dual-channel convolutional neural network. Remote Sens Lett 8(5):438–447
Zhang MM, Li W, Du Q (2018) Diverse region-based CNN for hyperspectral image classification. IEEE Trans Image Process 27(6):2623–2634
Zhang M, Jiang Y, Li C et al (2020a) Fully convolutional networks for blueberry bruising and calyx segmentation using hyperspectral transmittance imaging. Biosyst Eng 192:159
Zhang Z, Pasolli E, Crawford MM (2020b) An adaptive multiview active learning approach for spectral–spatial classification of hyperspectral images. IEEE Trans Geosci Remote Sens 58(4):2557–2570
Zhao WZ, Du SH (2016) Spectral–spatial feature extraction for hyperspectral image classification: a dimension reduction and deep learning approach. IEEE Trans Geosci Remote Sens 54(8):4544–4554
Zhao S, Qiu Z, He Y (2021) Transfer learning strategy for plastic pollution detection in soil: calibration transfer from high-throughput HSI system to NIR sensor. Chemosphere 7:129908
Zhong Z, Fan B, Duan J et al (2015) Discriminant tensor spectral–spatial feature extraction for hyperspectral image classification. IEEE Geosci Remote Sens Lett 12(5):1028–1032
Zhong ZL, Li J, Ma LF, Jiang H, Zhao H (2017a) Deep residual networks for hyperspectral image classification. In: IEEE international geoscience and remote sensing symposium (IGARSS), pp 1824–1827. https://doi.org/10.1109/IGARSS.2017.8127330
Zhong P, Gong ZQ, Li ST, Schonlieb CB (2017b) Learning to diversify deep belief networks for hyperspectral image classification. IEEE Trans Geosci Remote Sens 55(6):3516–3530
Zhong P, Gong ZQ (2017) A hybrid DBN and CRF model for spectral–spatial classification of hyperspectral images. Stat Optim Inf Comput 5(2):75
Zhou ZY, He DJ, Zhang HH, Lei Y, Su D, Chen K (2017) Non-destructive detection of moldy core in apple fruit based on deep belief network. Food Sci 38(14):297–303
Zhou X, Sun J, Tian Y, Chen QS, Wu XH, Hang YY (2020) A deep learning based regression method on hyperspectral data for rapid prediction of cadmium residue in lettuce leaves. Chemometr Intell Lab Syst 200(15):103996
Zhu L, Chen Y, Ghamisi P, Benediktsson JA (2018) Generative adversarial networks for hyperspectral image classification. IEEE Trans Geosci Remote Sens 56(9):5046–5063