Multimedia Tools and Applications
https://doi.org/10.1007/s11042-023-16456-w

Two-stage multi-dimensional convolutional stacked autoencoder network model for hyperspectral images classification

Yang Bai[1,2] · Xiyan Sun[1,2,3] · Yuanfa Ji[1,2] · Wentao Fu[1,3] · Jinli Zhang[1]

Received: 21 December 2022 / Revised: 11 May 2023 / Accepted: 6 August 2023

Abstract
Deep learning models have been widely used in hyperspectral images classification. However, the classification results are not satisfactory when the number of training samples is small. Focused on above-mentioned problem, a novel Two-stage Multi-dimensional Convolutional Stacked Autoencoder (TMC-SAE) model is proposed for hyperspectral images classification. The proposed model is composed of two sub-models SAE-1 and SAE-2. The SAE-1 is a 1D autoencoder with asymmetric structre based on full connection layers and 1D convolution layers to reduce spectral dimensionality. The SAE-2 is a hybrid autoencoder composed of 2D and 3D convolution operations to extract spectral-spatial features from the reduced dimensionality data by SAE-1. The SAE-1 is trained with raw data by unsupervised learning and the encoder of SAE-1 is employed to reduce spectral dimensionality of raw data. The data after dimension reduction is used to train the SAE-2 by unsupervised learning. The fine-tuning of SAE-2 encoder and the training of classifier are implemented simultaneously with small number of samples by supervised learning. Comparative experiments are performed on three widely used hyperspectral remote sensing data. The extensive comparative experiments demonstrate that the proposed architecture can effectively extract deep features and maintain high classification accuracy with small number of training samples.

Keywords Hyperspectral image classification · Deep learning · Stacked autoencoder · Multi-dimensional convolutional neural networks

✉ Xiyan Sun
sxy@guet.edu.cn

[1] School of Information and Communication, Guilin University of Electronic Technology, Guilin, China

[2] Guangxi Key Laboratory of Precision Navigation Technology and Application, Guilin University of Electronic Technology, Guilin, China

[3] National & Local Joint Engineering Research Center of Satellite Navigation and Location Service, Guilin University of Electronic Technology, Guilin, China

Published online: 16 August 2023

1 Introduction

Hyperspectral images (HSIs), which are comprised of hundreds of spectral bands and provide rich spectral and spatial information, are widely used in agriculture [24], environmental monitoring [32], mineral exploration [27], military and security [31], astronomy [13], medicine [25], chemistry [34], urban planning [38], etc. For these applications, the HSIs classification, which is to specify specific class for each pixel, is an important basic task. Because the effectiveness of all applications is directly affected by the classification accuracy. Unfortunately, the unbalance between the high dimensionality of spectral bands and the limited number of labeled samples make it very difficult to improve classification accuracy. On the one hand, the explosion of dimensionality not only provides abundant spectral information, but also contains enormous redundant information and noise, which makes the classification accuracy not increase but also decrease. This phenomenon is known as curse of dimensionality. On the other hand, the high cost of labeling samples results in a small number of labeld samples for training model. Therefore, how to extract deep discriminative features from a small number of training samples become a key step of HSI classification tasks [21, 35].

The traditional feature extraction(FE) methods consists of band selection(BS) and dimensionality reduction(DR). The purpose of BS is to select a subset from all spectral bands, which contains not only smaller dimensions, but also enough features representing the raw data for classification [11, 30, 37]. The purpose of DR is to find a lower dimensional representation of raw high dimensional data according to some mapping algrorithms, such as principal component analysis(PCA) [9, 18, 22, 41, 45], linear discriminant analysis(LDA) [7, 16, 19, 28, 43], morphological attribute profiles(MAPs) [2, 8, 10, 23, 40], etc. In the various BS algorithms, only the features of the subset bands are used for classification, that is, the features of other bands are discarded, so it will cause the waste of valuable feature information. The DR algorithms are mainly based on handcrafted features, so only shallow features can be obtained. Due to the inability to obtain deep features, it is difficult for traditional classification methods to further improve the classification accuracy.

In recent years, deep learning(DL) has shown amazing ability in deep feature extraction and achieved great successs in machine vision. So researchers are inspired to introduce the DL models into HSI classification. The DL models for HSI classification mainly include deep belief network(DBN), convolutional neural network(CNN) and autoencoder(AE), etc. Chen [5] proposed a novel deep model architecture for HSI classification, which combined the PCA for dimensionality reduction, the DBN model for spectral feature extraction and logistic regression as a classifier. Ghassemi [12] proposed a HSI classification framework in which the DBN was applied to extract spectral-spatial features. Because the DBN is a one-dimensional(1D) model, it is necessary to expand the two-dimensional(2D) spatial data into 1D vectors before extracting spatial features. The above-mentioned flatten processing of spatial features will cause the loss of spatial features and limits the improvement of classification accuracy.

CNN models, which is the most widely used DL model for HSIs classification, mainly contains three categories: 1D-CNN, 2D-CNN and 3D-CNN [1]. Hu [39] proposed a 1D-CNN model, which consisted of a convolutional layer, a max pooling layer and a full connection layer, for HSI classification with spectral features only. Li [20] proposed a pixel-pair 1D-CNN method combining the spectral and spatial information as the input of model to imrpove the classification accuracy. Yue [44] presented a framework, which consisted of PCA for dimensionality reduction, a deep 2D-CNN for

spectral-spatial feature extraction and a logistic regression classifier. Yu [42] introduced a deconvolution layer into a deep 2D-CNN model to enhance the extracted features from raw data. Haque [14] proposed a multi-scale 2D-CNN model named PCA-MS-CNN for HSIs classification. Li [21] proposed a lighter 3D-CNN framwork, which consisted of 3D convolution layers and full connection layers. Roy [29] proposed a hybrid CNN consisting of a spectral-spatial 3D-CNN followed by a spatial 2D-CNN. Zhang [46] proposed a Attention-Dense-HybridSN network based on 3D-CNN and 2D-CNN. In the network, a 3D-Dense block was used for extracting spectral-spatial features, and the channel and spatial attention were introduced to refine the extracted features. Because the 1D-CNN and 2D-CNN models cannot extract the spectral-spatial joint features of HSI data, the HSI classification methods based on 1D-CNN and 2D-CNN will lead to the loss of effective information. The 3D convolution kernel structrually matches the 3D cube data, so it can be used to extract spatial-spectral joint features. In addition, all above-mentioned CNNs are supervised learning models and the satisfactory classification accuracy(CA) can be obtained with sufficient labeled training samples, but the CA will decline rapidly when the training samples is few.

In recent years, AE as a unsupervised learning model has gained much attention. Chen [4] proposed three 1D-SAE models which were used for HSIs classification with spectral information, spatial information and spectral-spatial features respecively. Palma [17] proposed a hybrid unsupervised model based on 1D stacked AE(SAE) by introducding CNN in the training process of encoder and decoder. Mei [26] proposed a 3D convolutional autoncoder(3D-CAE), which consisted of a encoder with 3D convolutional operations only to maximally explore spatial-spectral information and a decoder to reconstruct the raw data. Sun [33] proposed a multi-scale 3D-CAE model composed of 3D convolutional layers and deconvolutional layers. The AE is composed of an encoder which can learns a representation for input data without labeled samples and a decoder which is used to resconstruct the input data.

Targeting the problem that the classification accuracy of models declines significantly with the decrease of the number of training samples, a novel deep learning framwork named Two-stage Multi-dimensional Convolutional Stacked Autoencoder(TMC-SAE) for HSI classification is proposed in this paper. The main contributions of this paper are summarized as follows.

(1) The TMC-SAE model was proposed for classification of hyperspectral remote sensing images. The highest classification accuracy was achieved with small number of training samples compared to other state-of-the-art models.
(2) The TMC-SAE consists of two independent stacked autoencoders SAE-1 and SAE-2. They are trainded independently by unsupervised learning. This architecture not only makes that the depth of SAE-1 and SAE-2 is not too large, but also ensures that TMC-SAE can extract depth features from HSIs.
(3) The SAE-1 is designed to be a 1D asymmetric SAE for spectral dimentionality reduction. The encoder of SAE-1 with 5 layers contains more trainable parameters than the decoder with 3 layers. This makes the feature extraction ability of the encoder obtain more attention during training.
(4) The SAE-2 is designed to be a hybrid network with 3D convolution and 2D convolution operations. The deep spatial-spectral-joint features extracted by SAE-2 make sure that the classification accuracy remains high when the number of training samples is small.

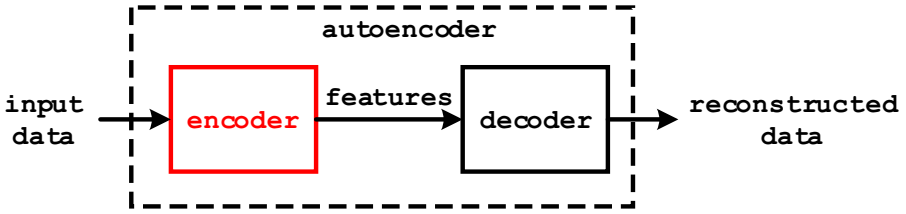

Fig. 1 The architecture of autoencoder

The remaining part of this paper is organized as follows. The related theoretical basis is described in Section 2. The framework details of TMC-SAE is presented in Sections 3 and 4. The experimental results over three benchmark hyperspectral datasets are shown in Section 5. Finally, conclusion are drawn in Section 6.

2 Related works

2.1 Stacked autoencoder

Figure 1 shows the general architecture of the autoencoder(AE), which consists of an encoder and a decoder. The function of the encoder is to extract the features of the input data and reduce the dimensionality of the data. The purpose of the decoder is to reconstructes the original data from the features extracted by the encoder.

During training, the encoder maps the input $X \in R^h$ to low dimensional representations $Y \in R^i$ through some algorithm and the decoder recovers $\widetilde{X} \in R^h$ from $Y \in R^i$ through inverse transformation. The purpose of training is to minimize the error between X and $\widetilde{X}$. This stage can be formulated mathematically as

$$\begin{aligned} Y &= f(W_e X + b_e) \\ \widetilde{X} &= g(W_d Y + b_d) \\ &\arg\ \min[loss(X, \widetilde{X})] \end{aligned} \tag{1}$$

where W_e, b_e and $f(\cdot)$ denote the weights, bias and activation function of encoder respectively, W_d, b_d and $g(\cdot)$ denote the weights, bias and activation function of decoder respectively.

During testing, only the encoder is adopted for feature extraction, and the features extracted by encoder are fed into the classifier for classification as shown in Fig. 2. The decoder is only used to obtain reconstructed data during training phase. The reconstructed data is closer to the input data, it is considered that the features are more representative.

Fig. 2 Testing process of the autoencoder

An AE which encoder and decoder contain more than one layer neural network is called a stacked autoencoder(SAE). In general, the number of operation layers in encoder and decoder are equal and the operations of decoder and encoder are inverse. In other words, the encoder and decoder are structurally symmetrical. The symmetrical structure makes SAE easy to be constructed. However, it is difficult to increase the depth of SAE because when the encoder is added one layer, the decoder must be added one layer, which makes the number of SAE layers be increased by 2. In order to improve the depth of encoder, an asymmetric structure of SAE is proposed, where the number of layers in decoder is smaller than that in encoder. This makes that there are more layers and trainable parameters in encoder to extract deep features for classification.

2.2 2D and 3D convolution

The 2D convolution and 3D convolution, which principle is shown in Fig. 3, are basic operations for extracting features in convolutional neural networks.

In the 2D convolution operation, input data is convolved with 2D kernels. The output data $y_{i,j}^{x,y}$ at spatial position (x, y) in the jth feature map of the ith layer is denoted as

$$y_{i,j}^{x,y} = f\left(\sum_{m}\sum_{p=0}^{W_1-1}\sum_{q=0}^{W_2-1} w_{i,j,m}^{p,q} v_{(i-1),m}^{(x+p)(y+q)} + b_{i,j}\right) \tag{2}$$

where m is the index of the feature maps in the $(i-1)$th layer, $w_{i,j,m}^{p,q}$ is the weight of position (p, q) connected to the mth feature map, W_1 and W_2 are the width and height of the kernel, $b_{i,j}$ is the bias for the jth feature map in the ith layer and $f(\cdot)$ is the activation function. Through 2D convolution operations, deep spatial features of input data can be extracted into output data.

In the 3D convolution operation, input data is convolved with 3D kernels. The output data $y_{i,j}^{x,y,z}$ at position position (x, y, z) of the jth feature map in the ith layer is given by

$$y_{i,j}^{x,y,z} = f\left(\sum_{m}\sum_{p=0}^{W_1-1}\sum_{q=0}^{W_2-1}\sum_{r=0}^{W_3-1} w_{i,j,m}^{p,q,r} x_{(i-1),m}^{(x+p)(y+q)(z+r)} + b_{i,j}\right) \tag{3}$$

where $w_{i,j,m}^{x,y,z}$ is the weight of position (p, q, r) connected to the mth feature map in the ith layer, W_3 is the size of kernel along toward spectral dimension, and other parameters are the same as the Eq. (2). The structure of 3D kernel is consistent with that of HSI data cube, so 3D convolution operations can extract spatial and spectral features simultaneously.

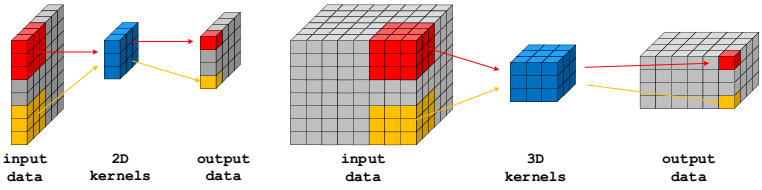

Fig. 3 **a** 2D convlution operation; **b** 3D convolution operation

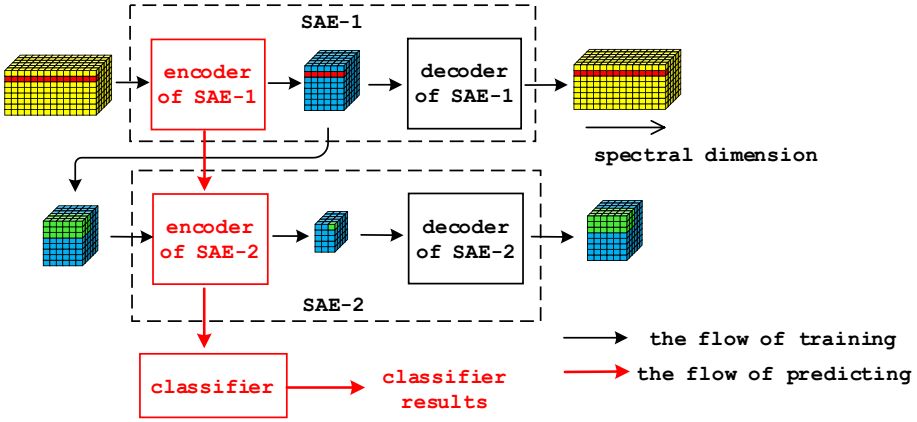

Fig. 4 Framework of TMC-SAE

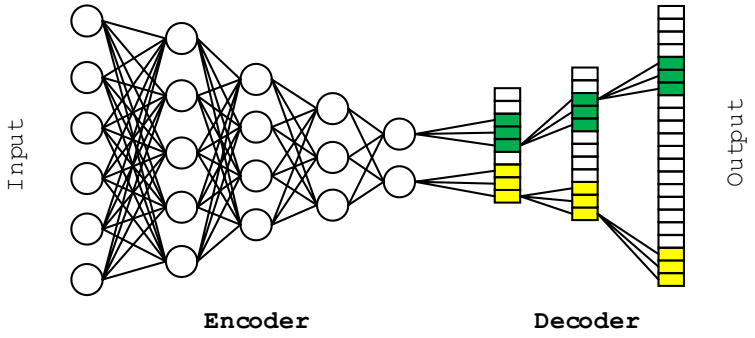

Fig. 5 Structure of SAE-1

3 Proposed TMC-SAE

3.1 Framework of the proposed TMC-SAE

In this paper, the TMC-SAE is proposed for HSI classification. As shown in Fig. 4, the TMC-SAE is composed of two stacked autoencoders(SAE) SAE-1 and SAE-2 respectively and a classifier. Both SAE-1 and SAE-2 contain a encoder and a decoder. The function of encoders and decoders are to extract features and reconstruct input data respectively. The decoders are designed only for training the encoders and not for classification. The network for classification is composed of the SAE-1 encoder, SAE-2 encoder, and classifier. The structures and training details of SAE-1, SAE-2 and classifier will be described in below.

The SAE-1 is a 1D SAE with asymmetric structure as shown in Fig. 5, in which the encoder and decoder are based on full connection(FC) layers and 1D convolutional layers respectively. The purpose of this asymmetric structre is to make the encoder contains more trainable parameters than decoder to improve its ability of feature extraction.

The encoder of SAE-1 consists of five FC layers which contain k1, k2, k3, k4, k5 neurons respectively and each FC layer is followed by a batch normalization(BN) layer,

activation layer with ReLU activation function and dropout layer(rate = 0.5). The decoder of SAE-1 is composed of three 1D deconvolution(DC) layers and each DC layer is followed by a BN layer and activation layer.

It is assumed that the raw HSI data is represented by $\boldsymbol{X} \in \mathbb{R}^{M \times N \times B}$, where M and N are the height and width of the image and B is the number of spectral bands. After the dimension reduction of spectral by encoder, the pixel data vector $x \in \mathbb{R}^{B}$ is mapped to the feature vecotr h with k5 dimensionality. The trained encoder will be used to reduce the dimension of raw HSI data and the output of encoder with size of $M \times N \times k5$ will be taken as the input of ASE-2. The encoder of SAE-1 reduces the number of spectral bands from B to k5 while maintaining the same spatial dimensions.

A hybrid network SAE-2 is proposed to further extract spectral-spatial features from the data after dimension reduction by encoder of SAE-1. The framework of SAE-2 is shown in Fig. 6. It consists of a encoder, which stacks three 3D convolution layers and three 2D convolution layers to extract spatial-spectral features simultaneously, and a companion decoder, which is composed of three 3D deconvolution layers and three 2D deconvolution layers to reconstruct the input data from the features extracted by the encoder.

The SAE-1 encoder output $X \in \mathbb{R}^{M \times N \times k_5}$ is divided into the 3-D neighboring patches $\boldsymbol{P} \in \mathbb{R}^{S \times S \times k_5}$, which is taken as the input of SAE-2. Each patch $P_{x,y} \in \boldsymbol{P}$ centered at the spatial location (x, y) pixel is generated by covering the $S \times S$ window and all spectral bands. The function of reshape layer is to combine the spectral dimension and channel dimension of the feature maps to make it suitable for next 2D convolution layer. There is none trainable parameter in the reshape layer. The backpropagate method is used to train the SAE-2 with a MSE loss function. In both the encoder and decoder, the ReLU activation function is adopted for every convolution and deconvolution layer to improve network fitting ability.

After the ASE-2 is trained, the encoder of ASE-2 is used independently to provide extracted spatial-spectral features for classifier. The classifier consists of a flatten layer, which expands the extracted features by ASE-2 encoder to 1D vectors, and three FC layers. The first two FC layers with ReLU activation function are designed to extract features further and followed by a dropout layer to prevent overfitting. The last FC layer

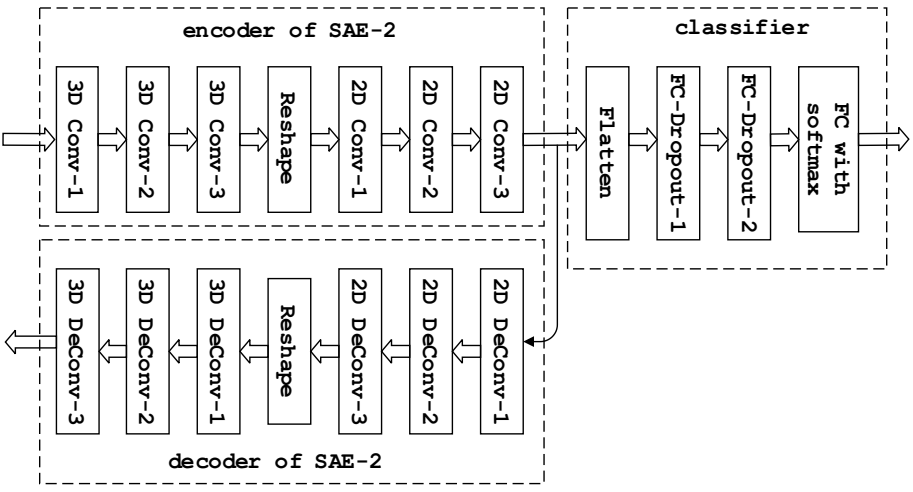

Fig. 6 Structure of SAE-2 and classifier

with the same number of neurons as the number of classes of pixels uses softmax activation function to implement the classifier.

3.2 Details of training

The training of TMC-SAE is a three-phase process: (1) the training of SAE-1 based on unsupervised learning. In this step, the encoder of SAE-1 automatically extracts features form raw spectral data and the decoder reconstructs the raw data from the output of encoder. The training dataset is composed of all pixel vectors. The trained encoder of SAE-1 realizes dimension reduction from the raw HSI data $\boldsymbol{X} \in \mathbb{R}^{M \times N \times B}$ to $\boldsymbol{Y} \in \mathbb{R}^{M \times N \times k_5}$ only in spectral dimension. (2) the training of SAE-2 based on unsupervised learning. This process is as same as step (1) except that the training data is the extracted features of trained SAE-1 encoder. In this phase, the 3D neighboring pathces dataset $\boldsymbol{Z} \in \mathbb{R}^{P \times P \times k_5}$, which contains the information of all labeled pixels and is generated from $\boldsymbol{Y} \in \mathbb{R}^{M \times N \times k_5}$, is taken as the training dataset. The parameters P represents the patch window size of the training sample. (3) the training of classifier and fine-tuning of SAE-2 based on supervised learning with small labeled smaples. In this phase, the dataset $\boldsymbol{Z} \in \mathbb{R}^{P \times P \times k_5}$ is divided into training and testing groups, respectively. The classifier training and SAE-2 encoder fine-tuning are performed simultaneously based on the training group. After the above process, the classification performance of TMC-SAE is verifed based on the testing group. It can be seen from the above details that the features of all pixels can be used for the ASE-1 and ASE-2 training. This allows the encoders of ASE-1 and ASE-2 make maximum use of the information in the dataset instead of relying on only a small number of labeled samples. Thanks to the deep features extraction ability of SAE-1 and SAE-2 encoders, the high classification accuracy can still be obtained based on a small samples training group. The detailed flowchart of TMC-SAE training and testing is shown in Fig. 7.

4 Details of experimental

4.1 Data description

In this paper, three benchmark hyperspectral datasets with different environmental settings are adopted to validate our proposed network. The first dataset was gathered by the Airborne Visible Infrared Imaging Spectrometer(AVIRIS) instrument over a mixed vegetation site in northwestern Indiana (Indian Pines, IP). It contains 145×145 pixels with 220 spectral channels covering the range from 0.4 to 2.5 μm. The second dataset was acquired over Kennedy Space Center(KSC), Florida. It consists of 512×614 pixels with 176 spectral bands. There are 13 different land-cover classes in the raw dataset. The third dataset was gathered over SalinasValley(SV), California. It contains 512×217 pixels and 224 bands in the range of 0.4–2.5 μm. There are 204 bands in the corrected data after 20 water absorption bands are removed. The land-cover classes and the labeled pixel numbers of each class for all datasets are listed in Table 1. The ground truth images of all datasets are shown in Fig. 8. All experiments are conducted on a computer with Intel(R) Core i7- CPU, Nvidia Geforce GTX 3090 GPU and 64 Gb RAM.

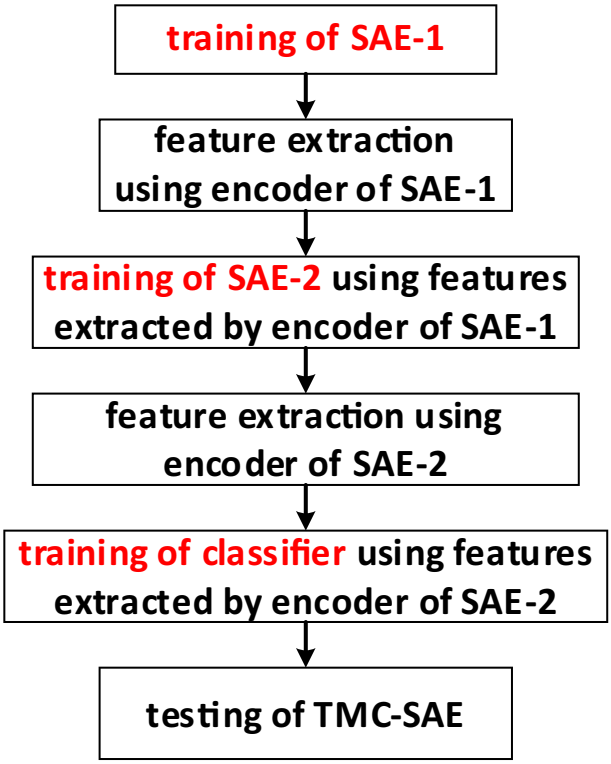

Fig. 7 Flowchart of training and testing for TMC-SAE

Table 1 The Class labels and number of training and testing samples

No	IP		KSC		SV	
	Class	Number	Class	Number	Class	Number
1	Alfalfa	46	Scrub	761	Brocoli green weeds 1	2009
2	Corn-notil	1428	Willow swamp	243	Brocoli green weeds 2	3726
3	Corn-min	830	CP hammock	256	Fallow	1976
4	Corn	237	CP/Oak	252	Fallow rough plow	1394
5	Grass-pasture	483	Slash pine	161	Fallow smooth	2678
6	Grass-trees	730	Oak/Broadleaf	229	Stubble	3959
7	Grass-mowed	28	Hardwood swamp	105	Celery	3579
8	Hay-windrowed	478	Graminoid marsh	431	Grapes untrained	11,271
9	Oats	20	Spartina marsh	520	Soil vinyard develop	6203
10	Soybean-notill	972	Cattail marsh	404	Corn senesced green weeds	3278
11	Soybean-mintill	2455	Salt marsh	419	Lettuce romaine 4wk	1068
12	Soybean	593	Mud flats	503	Lettuce romaine 5wk	1927
13	Wheat	205	Water	927	Lettuce romaine 6wk	916
14	Woods	1265			Lettuce romaine 7wk	1070
15	Buildings-Grass	386			Vinyard untrained	7268
16	Stone-Steel	93			Vinyard vertical trellis	1807
Total		10,249		5211		54,129

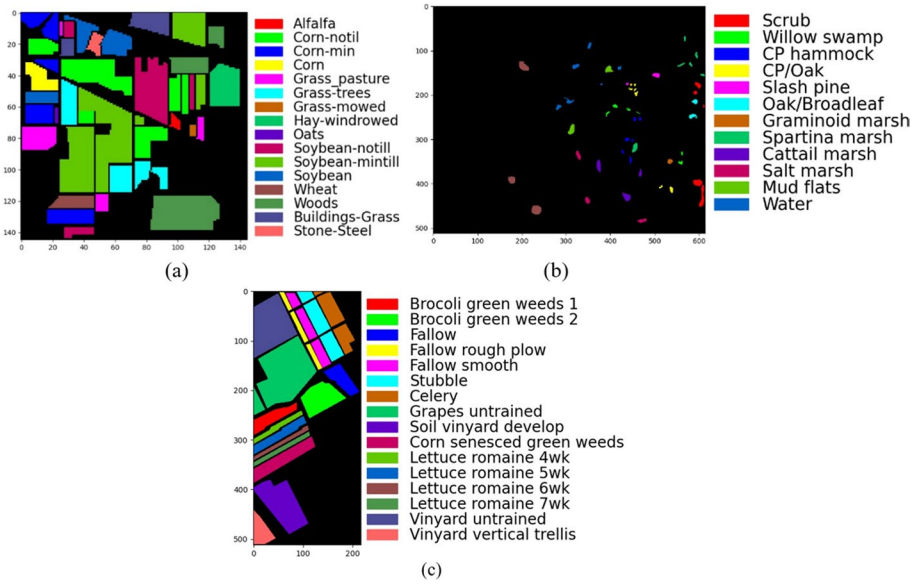

Fig. 8 Ground truth image. **a** IP dataset. **b** KSC dataset. **c** SV dataset

4.2 Network construction

Because the numbers of bands in three datasets are different, the numbers of neuron(k1 ~ k5) in the FC layers of SAE-1 encoder are different. In general, the spectral band compression ratio of the SAE-1 encoder is about 1/8. The network structure of SAE-1 is given in Table 2. It can be seen from Table 2 that the number of trainable parameters in encoder is much larger than those in the decoder. This asymmetric structure imporves the feature extraction ability of encoder.

The parameters of all layers in SAE-2 are the same for all datasets. The structure of SAE-2 and classifier is given in Table 3. In the SAE-2, the kernel sizes and strides of all layers are based on 3 and 1, respectively. The purpose of this design is to reduce the trainable parameters and the loss of spatial-spectral information during training process. The activation function employed in network is ReLU except for the last layer of the classifier.

Table 2 Network structures of SAE-1

Layer		For IP	For KSC
Encoder	Dense-1	k1 = 160	k1 = 100
	Dense-2	k2 = 120	k2 = 80
	Dense-3	k3 = 90	k3 = 60
	Dense-4	k4 = 50	k4 = 40
	Dense-5	k5 = 25	k5 = 13
Decoder	1D DeConv-1	kernel = 16@3,strides = 2	
	1D DeConv-2	kernel = 64@3,strides = 2	
	1D DeConv-3	kernel = 1@3,strides = 2	

Table 3 Network structures of SAE-2 and classifier

Layer		kernel	strides
Encoder	3D Conv-1	64@(3,3,3)	(1,1,1)
	3D Conv-2	32@(3,3,3)	(1,1,1)
	3D Conv-3	16@(3,3,3)	(1,1,1)
	2D Conv-1	256@(3,3)	(1,1)
	2D Conv-2	128@(3,3)	(1,1)
	2D Conv-3	64@(3,3)	(1,1)
Decoder	2D DeConv-1	128@(3,3)	(1,1)
	2D DeConv-2	256@(3,3)	(1,1)
	2D DeConv-3	n*@(3,3)	(1,1)
	3D DeConv-1	16@(3,3,3)	(1,1,1)
	3D DeConv-2	8@(3,3,3)	(1,1,1)
	3D DeConv-3	1@(3,3,3)	(1,1,1)
Classifier	FC-1	units = 256,rate of dropout = 0.4	
	FC-2	units = 256,rate of dropout = 0.4	
	FC-3	units = the number of classes	

*: The value of n is equal to the number of channels of the reshaped output of 3D Conv-3 layer

The learning rates of the ASE-1 and ASE-2 training are both 0.001, but the learning rate is 0.0001 when the classifier is trained and the encoder of ASE-2 is fine-tuned.

5 Experimental results and analysis

5.1 Analysis of parameters

In the architecture of TMC-SAE, the depth of SAE-1 is an important parameter for the classification performance. A series of experiments were conducted to evaluate the impact of SAE-1 depth on classification results. In the experiment, the depth of SAE-1 encoder was set eight different values from 1 to 8 and the overall accuracy(OA) was used to evaluate the classification performance of TMC-SAE with different depth on three datasets repectively. The experimental results are shown in Fig. 9. It can be seen that the OA first increases and then decreases as the depth of SAE-1 increases. This indicates that deeper SAE-1 can extract representative and deep features but will encounter the overfitting. Based on the experimental results, the depth of SAE-1 encoder was determined to be 5.

The encoder of SAE-2 consists of 2D convolution layers and 3D convolution layers. The purpose of 3D convolution operations is to extract spatial-spectral joint features from data that have been dimensionally reduced by SAE-1. The function of 2D convolution operations is to extract deeper features for classification task. In order to evaluate the effectiveness of 3D convolution and 2D convolution operations, the incomplete SAE without 3D convolution branch and that without 2D branch were used for classification experiments separately. The experimental results shown in Fig. 10 indicate that the SAE without 2D or 3D operations slightly reduce classification accuracy.

The loss and classification accuracy convergence curves of training group are portrayed in Fig. 11. It can be seen that both curves of all datasets converge at about 200 epochs.

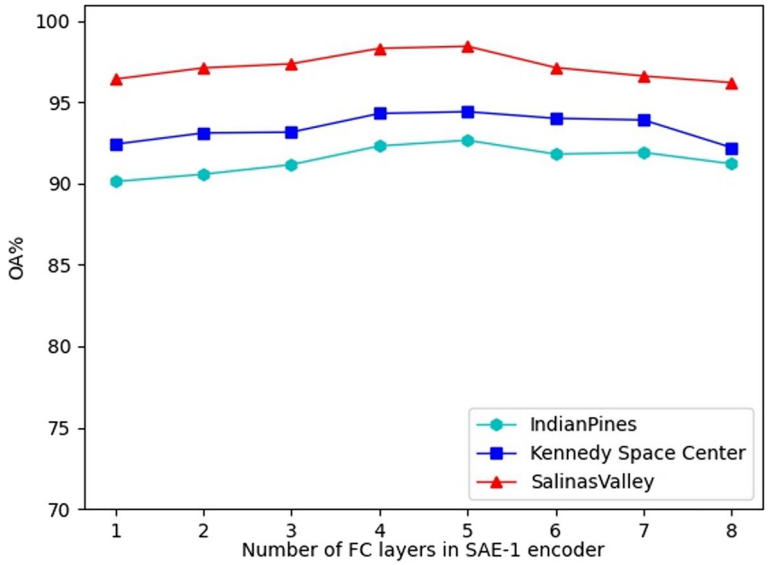

Fig. 9 Impact of SAE-1 encoder depth on overall accuracy

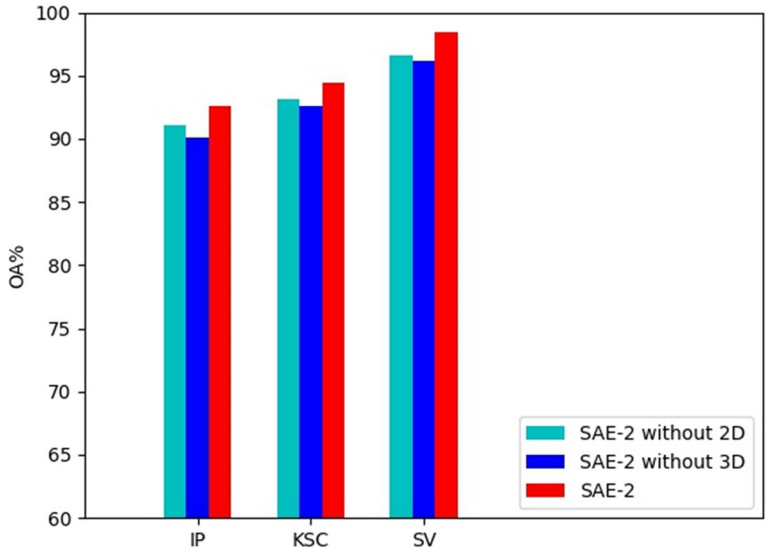

Fig. 10 Imapct of 2D or 3D operations in SAE-2 on overall accuracy

5.2 Visualization and analysis of ASE-1

In order to gain detailed understanding of the SAE-1, visualization about spectral information is provided in this section. The spectral curves are used to visualize the features before and after extraction by SAE-1. The raw spectral curves of graminoid marsh(class

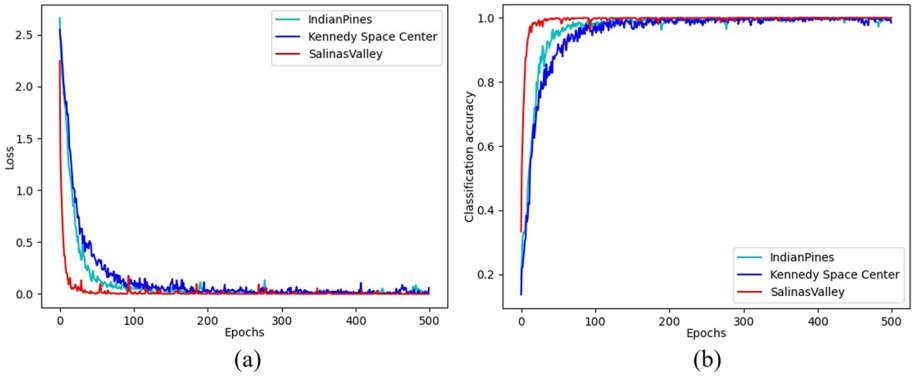

Fig. 11 The training losses and classification accuracy curves

8) and spartina marsh(class 9) in KSC are shown in Fig. 12a and b. Obviously, the two curves are very similar and fifficult to distinguish. The extracted feature curves by SAE-1 are shown in Fig. 12c and d. These two features, which dimensions are reduced from 175 to 20, become more discriminable and abstract.

5.3 Comparison of classification results

In this experiment, the overall accuracy(OA), average accuracy(AA), and Kappa coefficient(Kappa) are introduced to evaluate the classification results. In addition, the results of the proposed TMC-SAE are compared with six state-of-the-art HSI classification models, which cover unsupervised learning and supervised learning with different dimensions, such as 1D-CNN [39], 2D-CNN [36], 3D-CNN-C [6], M3D-DCNN [15], 3D-CNN-H [3] and 3D-CAE [33]. The architectures and hyperparameters of these comparative

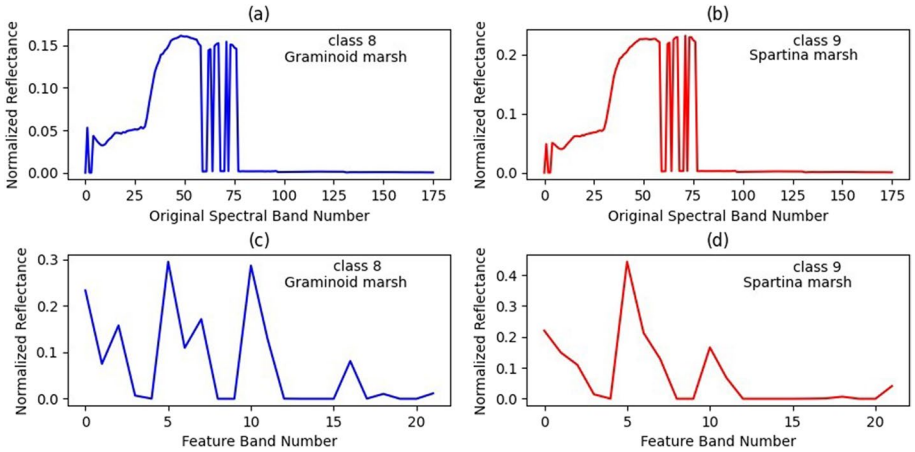

Fig. 12 Representative spectral curves of two land cover classes of the KSC. **a** Original spectral of class 8 graminoid mash. **b** Original spectral of class 9 Spartina marsh. **c** Features of class 8 after SAE-1. **d** Features of class 9 after SAE-1

models are consistent with that given in the corresponding papers. All the models are implemented using Python language and TensorFlow library. In order to verify the feature extraction ability of proposed model under the condition of small number of labeled samples, the training sample percentage of each class for IP, KSC and SV is set to 5%, 5% and 1% respectively.

The quantitative results over IP, KSC and SV datasets are listed in Tables 4, 5 and 6 respectively. It can be observed from three tables that the OA, AA and Kappa of propsed TMC-SAE outperform those of all other models for all datasets. The OA of TMC-SAE achieves 92.65% for IP, 94.41% for KSC and 98.50% for SV. The best accuracy of class 1–4, 10, 11, 13, 15 for IP, class 1, 3, 6–13 for KSC and class 1–3, 5–7, 9, 13, 15, 16 for SV is generated by the proposed TMC-SAE model. The experimental results show that there is no much lower result among the accuracy of each class of the proposed TMC-SAE even if the training sample is very few. It can be concluded that the feature extraction capability of TMC-SAE is more stronger and the above capability is enhanced by the unsupervised learning of SAE-1 and SAE-2. Figure 13 illustrates the classification maps of IP dataset with each above-mentioned model. The quality of the classification map of TMC-SAE is much better than other models especially for the classes with small number of samples.

5.4 Impact of the training sample size

In this part, the effect of the different training sample size with all models is explored. For IP and KSC datasets, the percentage of training samples is set 3%, 5%, 10%, 15%

Table 4 Classification accuracy of different models over the Indian Pines dataset

Class	Train	Test	1D-CNN	2D-CNN	3D-CNN-C	M3D-DCNN	3D-CNN-H	3D-CAE	TMC-SAE
1	2	44	65.91	18.18	77.27	25.00	52.27	54.55	84.09
2	71	1357	77.16	82.46	85.92	72.00	73.91	82.02	88.14
3	41	789	68.82	87.07	87.83	69.96	66.67	81.24	89.73
4	12	225	59.11	74.67	87.11	76.89	52.00	78.22	82.22
5	24	459	85.62	89.76	92.16	69.93	84.75	93.03	91.07
6	37	693	95.24	99.42	91.05	96.39	94.08	95.53	92.35
7	1	27	18.52	14.81	88.89	3.70	25.93	66.67	70.37
8	24	454	99.78	100.00	100.00	100.00	92.51	99.34	98.46
9	1	19	0.00	31.58	47.37	31.58	42.11	68.42	57.89
10	49	923	74.00	87.54	89.38	71.83	65.01	90.90	91.12
11	123	2332	79.97	91.55	95.45	85.12	80.06	90.18	96.4
12	30	563	81.71	70.87	87.21	69.45	61.81	83.84	84.9
13	10	195	98.97	98.97	98.46	100.00	97.95	95.38	100.00
14	63	1202	96.01	99.17	95.67	96.67	94.68	97.50	96.09
15	19	367	53.41	80.65	94.28	81.47	62.67	82.02	97.00
16	5	88	82.95	100	82.95	93.18	85.23	95.45	98.86
OA			81.00	88.92	91.74	81.54	77.99	89.17	92.65
AA			71.07	76.66	87.56	74.15	70.73	84.64	88.67
Kappa			78.31	87.32	90.57	78.86	74.78	87.65	91.61

Table 5 Classification accuracy of different models over the KSC dataset

Class	Train	Test	1D-CNN	2D-CNN	3D-CNN-C	M3D-DCNN	3D-CNN-H	3D-CAE	TMC-SAE
1	38	723	94.74	94.47	87.28	92.81	92.67	68.74	100.00
2	12	231	84.85	42.42	82.68	84.85	90.04	30.74	89.61
3	13	243	87.65	57.20	64.61	67.90	71.60	68.31	99.59
4	13	239	46.03	35.56	75.31	39.75	48.12	36.40	61.09
5	8	153	40.52	75.16	82.35	47.06	37.25	72.55	68.63
6	11	218	45.41	50.46	83.03	82.57	70.18	47.25	95.41
7	5	100	86.00	80.00	67.00	13.00	95.00	62.00	100.00
8	22	409	80.20	58.92	81.91	49.14	74.15	73.11	80.20
9	26	494	94.53	85.63	90.08	83.60	95.14	53.44	96.15
10	20	384	85.42	51.56	97.66	91.67	86.72	63.54	99.74
11	21	398	96.48	88.94	96.98	96.23	98.99	83.17	100.00
12	25	478	80.54	56.69	95.82	92.05	79.50	59.62	100.00
13	46	881	100.00	98.30	100.00	100.00	99.89	94.67	100.00
OA(%)			85.32	73.99	89.13	82.04	85.24	67.74	94.41
AA(%)			78.65	67.33	84.98	72.36	79.71	62.58	91.57
Kappa			83.64	71.00	87.91	80.00	83.57	64.24	93.76

Table 6 Classification accuracy of different models over the Salinas dataset

Class	Train	Test	1D-CNN	2D-CNN	3D-CNN-C	M3D-DCNN	3D-CNN-H	3D-CAE	TMC-SAE
1	20	1989	100	99.95	100	100	100	100	100
2	37	3689	96.43	95.57	85.56	99.68	98.33	100	100
3	20	1956	94.93	99.54	99.06	93.36	92.25	99.34	99.79
4	14	1380	98.70	96.77	99.78	98.63	96.50	95.80	99.07
5	27	2651	98.00	99.62	98.31	98.78	93.72	99.23	99.89
6	39	3920	99.31	99.77	100	100	98.21	100	100
7	36	3543	98.79	100.00	90.57	98.71	99.66	100	99.72
8	113	11,158	71.67	89.12	97.31	89.65	82.44	95.39	97.24
9	62	6141	96.26	99.47	98.81	97.06	96.50	99.02	99.98
10	33	3245	90.32	94.90	97.84	89.66	86.67	99.48	97.85
11	11	1057	83.77	98.03	97.88	100	90.09	90.62	99.81
12	19	1908	97.44	100.00	92.13	95.45	97.31	99.89	98.50
13	9	907	95.96	100.00	97.88	98.22	83.21	98.59	100
14	11	1059	92.10	99.06	98.28	98.36	90.32	99.15	96.53
15	72	7196	81.67	95.18	88.53	75.77	76.78	89.51	95.49
16	18	1789	98.73	92.07	99.30	95.14	97.60	100	100
OA(%)			88.70	95.81	95.17	92.53	90.26	97.03	98.50
AA(%)			92.39	97.44	94.98	94.68	93.00	98.32	98.89
Kappa			87.35	95.33	94.62	91.69	89.16	96.69	98.33

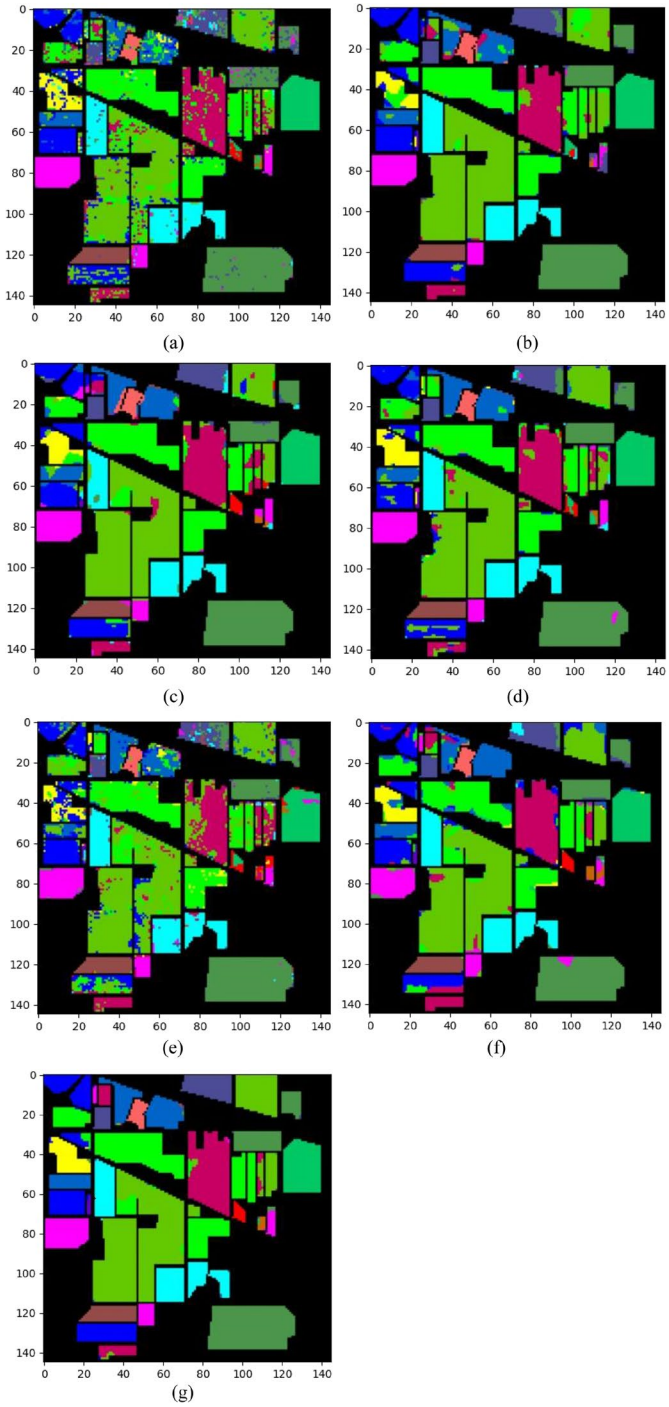

Fig. 13 Classification maps generated by different models over IP dataset. **a** 1D-CNN. **b** 2D-CNN. **c** 3D-CNN-C. **d** M3D-DCNN. **e** 3D-CNN-H. **f** 3D-CAE. **g** Proposed TMC-SAE

and 20% and for SV dataset, it is set 0.5%, 1%, 3%, 5% and 7%. Figure 14 shows the OA results of different percentage of training samples on all datasets. As we can observe in Fig. 14, for all models, higher classification results can be obtained with larger proportion of training samples. However, with the decline of the proportion of training samples, the decline of classification accuracy of different models varies greatly. For IP dataset, the OA results of 2D-CNN-N, 3D-CNN-C, 3D-CAS and TMC-SAE are similar, when the percentage of training sample is 20%. However, there is more than difference between the largest OA result (proposed TMC-SAE, 85.29%) and the smallest classification result (M3D-DCNN, 75.11%) when the percentage of training sample is reduced to 3%. The proposed TMC-SAE model generates the highest accuracies in all experiments with small number of training sample. Specifically, when the proportion of training sample is 3% and 5%, the decline of classification accuracy of the proposed TMC-SAE is the smallest. For SV dataset, when the percentage of training samples is 7%, the OA results of all methods exceed 99% except 1D-CNN. It indicates that these models can extract sufficient features for classification when there are enough training samples. When the percentage of training samples decreases, especially at 1% and 0.5%, the OA of TMC-SAE remains the highest value. It indicates that the TMC-SAE maintains better feature extraction ability in small number of training samples.

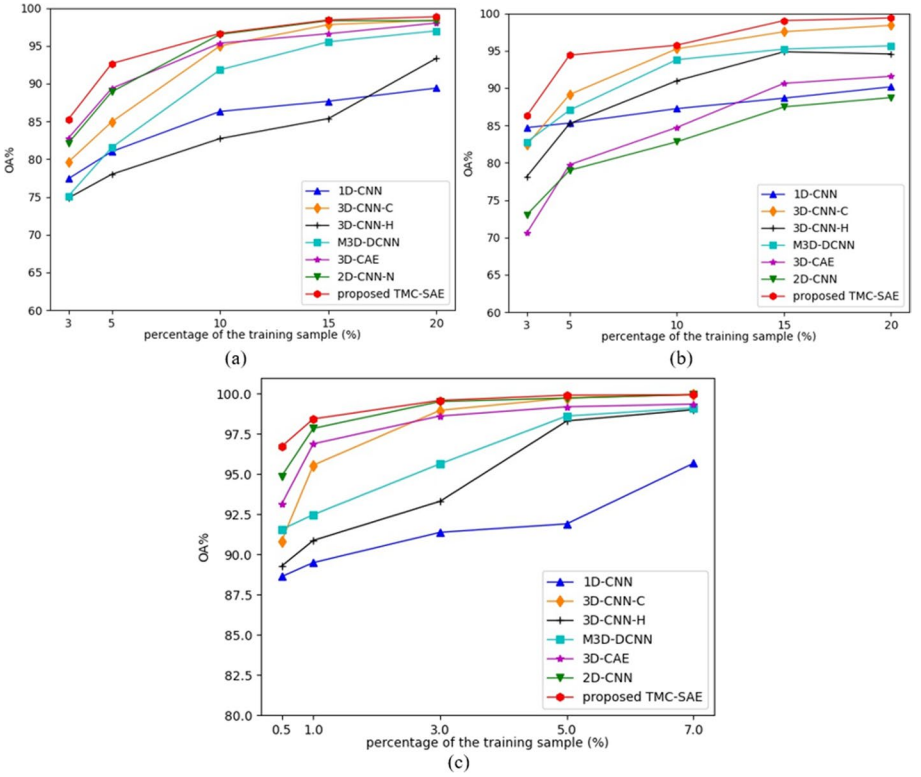

Fig. 14 Experimental results of all models with different percentages of training samples over three datasets. **a** IP. **b** KSC. **c** SV

6 Discussion and conclusion

In this paper, a new network architecture for hyperspectral remote sensing image classification is proposed. It consists of two stacked autoencoder networks SAE-1 and SAE-2. The purpose of SAE-1 based on 1D CNN is for feature extraction in spectral domain only. The asymmetric architecture improves the feature extraction ability of SAE-1 by making the number of trainable parameters in encoder more than that in decoder. The SAE-2 based on 2D and 3D CNN can extract spatial-spectral joint features from the information compressed by SAE-1. Generally, there is only one unsupervised learning in the previous network training. In this paper, the proposed TMC-SAE is divided into two independent autoencoders SAE-1 and SAE-2. This architecture increases the number of unsupervised training times to two, so that the information in unlabeled samples can be extracted more fully. The experimental results with real hyperspectral images demonstrate that the proposed TMC-SAE can achieve better classification result with a small number of training samples.

Funding This research work is supported by the project supported by Guangxi Key Laboratory of Precision Navigation Technology and Application, Guilin University of Electronic Technology (No. DH202208).

Data availability The datasets analysed during the current study are available from the corresponding author on reasonable request.

Declarations

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

References

1. Bai Y, Sun X, Ji Y, Huang J, Fu W, Shi H (2022) Bibliometric and visualized analysis of deep learning in remote sensing. Int J Remote Sens 43(15-16SI):5534–5571
2. Bao R, Xia J, Dalla Mura M, Du P, Chanussot J, Ren J (2016) Combining morphological attribute profiles via an ensemble method for hyperspectral image classification. IEEE Geosci Remote Sens Lett 13:359–363
3. Ben Hamida A, Benoit A, Lambert P, Ben Amar C (2018) 3-D deep learning approach for remote sensing image classification. IEEE Trans Geosci Remote Sens 56:4420–4434
4. Chen Y, Lin Z, Zhao X, Wang G, Gu Y (2014) Deep learning-based classification of hyperspectral data. IEEE J Sel Top Appl Earth Observ Remote Sens 7:2094–2107
5. Chen Y, Zhao X, Jia X (2015) Spectral-spatial classification of hyperspectral data based on deep belief network. IEEE J Sel Top Appl Earth Observ Remote Sens 8:2381–2392
6. Chen Y, Jiang H, Li C, Jia X, Ghamisi P (2016) Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. IEEE Trans Geosci Remote Sens 54:6232–6251

7. Du Q, Younan NH (2008) Dimensionality reduction and linear discriminant analysis for hyperspectral image classification. Knowledge-based intelligent information and engineering systems, pt 3, proceedings, 5179, pp 392–399
8. Falco N, Benediktsson JA, Bruzzone L (2015) Spectral and spatial classification of hyperspectral images based on ICA and reduced morphological attribute profiles. IEEE Trans Geosci Remote Sens 53:6223–6240
9. Fauvel M, Chanussot J, Benediktsson JA (2006) Kernel principal component analysis for feature reduction in hyperspectrale images analysis. IEEE, New York, p 238
10. Fauvel M, Benediktsson JA, Chanussot J, Sveinsson JR (2008) Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles. IEEE Trans Geosci Remote Sens 46:3804–3814
11. Feng J, Jiao L, Sun T, Liu H, Zhang X (2016) Multiple kernel learning based on discriminative kernel clustering for hyperspectral band selection. IEEE Trans Geosci Remote Sens 54:6516–6530
12. Ghassemi M, Ghassemian H, Imani M (2018) Deep belief networks for feature fusion in hyperspectral image classification. Proceedingss of the 2018 IEEE international conference on aerospace electronics and remote sensing technology (ICARES 2018)
13. Guilloteau C, Oberlin T, Berne O, Dobigeon N (2020) Fusion of hyperspectral and multispectral infrared astronomical images. 2020 IEEE 11th sensor array and multichannel signal processing workshop (SAM)
14. Haque MR, Mishu SZ (2019) Spectral-spatial feature extraction using PCA and multi-scale deep convolutional neural network for hyperspectral image classification. 2019 22nd International Conference on Computer and Information Technology (ICCIT) 2019 22nd International Conference on Computer and Information Technology (ICCIT), pp 1–6
15. He MY, Li B, Chen HH (2017) Multi-scale 3D deep convolutional neural network for hyperspectral image classification. IEEE International Conference on Image Processing ICIP. IEEE, New York, pp 3904–3908
16. Imani M, Ghassemian H (2015) Two dimensional linear discriminant analyses for hyperspectral data. Photogramm Eng Remote Sens 81:777–786
17. Jijón-Palma ME, Kern J, Amisse C, Centeno JAS (2021) Improving stacked-autoencoders with 1D convolutional-nets for hyperspectral image land-cover classification. J Appl Remote Sens 15:26506
18. Khan Z, Shafait F, Mian A (2015) Joint Group Sparse PCA for compressed hyperspectral imaging. IEEE Trans Image Process 24:4934–4942
19. Li W, Prasad S, Fowler JE, Bruce LM (2011) Locality-preserving discriminant analysis in kernel-induced feature spaces for hyperspectral image classification. IEEE Geosci Remote Sens Lett 8:894–898
20. Li W, Wu G, Zhang F, Du Q (2017) Hyperspectral image classification using deep pixel-pair features. IEEE Trans Geosci Remote Sens 55:844–853
21. Li Y, Zhang H, Shen Q (2017) Spectral-spatial classification of hyperspectral imagery with 3D convolutional neural network. Remote Sens 9:67
22. Licciardi G, Marpu PR, Chanussot J, Benediktsson JA (2012) Linear versus nonlinear PCA for the classification of hyperspectral data based on the extended morphological profiles. IEEE Geosci Remote Sens Lett 9:447–451
23. Liu B, Guo W, Chen X, Gao K, Zuo X, Wang R, Yu A (2020) Morphological attribute profile cube and deep random forest for small sample classification of hyperspectral image. IEEE Access 8:117096–117108
24. Lu B, Dao PD, Liu J, He Y, Shang J (2020) Recent advances of hyperspectral imaging technology and applications in agriculture. Remote Sens 12:2659
25. Marotz J, Kulcke A, Siemers F, Cruz D, Aljowder A, Promny D, Daeschlein G, Wild T (2019) Extended perfusion parameter estimation from hyperspectral imaging data for bedside diagnostic in medicine. Molecules 24:4164
26. Mei S, Ji J, Geng Y, Zhang Z, Li X, Du Q (2019) Unsupervised spatial-spectral feature learning by 3D convolutional autoencoder for hyperspectral classification. IEEE Trans Geosci Remote Sens 57:6808–6820
27. Nakayama K, Tonooka H (2021) Improvement of a mineral discrimination method using multispectral image and surrounding hyperspectral image. J Appl Remote Sens 15
28. Peng J, Luo T (2016) Sparse matrix transform-based linear discriminant analysis for hyperspectral image classification. Signal Image Video Process 10:761–768
29. Roy SK, Krishna G, Dubey SR, Chaudhuri BB (2020) HybridSN: exploring 3-D-2-D CNN feature hierarchy for hyperspectral image classification. IEEE Geosci Remote Sens Lett 17:277–281

30. Sellami A, Farah M, Farah IR, Solaiman B (2018) Hyperspectral imagery semantic interpretation based on adaptive constrained band selection and knowledge extraction techniques. IEEE J Sel Top Appl Earth Observ Remote Sens 11:1337–1347
31. Shimoni M, Haelterman R, Perneel C (2019) Hyperspectral imaging for military and security applications combining myriad processing and sensing techniques. IEEE Geosci Remote Sens Mag 7:101–117
32. Stuart MB, McGonigle AJS, Willmott JR (2019) Hyperspectral imaging in environmental monitoring: a review of recent developments and technological advances in compact field deployable systems. Sensors 19:3071
33. Sun Q, Liu X, Bourennane S (2021) Unsupervised multi-level feature extraction for improvement of hyperspectral classification. Remote Sens 13:1602
34. Sun Y, Qian X, Liu Y, Wang J, Lv Q, Yuan M (2021) Identification of typical solid hazardous chemicals based on hyperspectral imaging. Remote Sens 13:2608
35. Tao C, Pan H, Li Y, Zou Z (2015) Unsupervised spectral-spatial feature learning with stacked sparse autoencoder for hyperspectral imagery classification. IEEE Geosci Remote Sens Lett 12:2438–2442
36. Tun NL, Gavrilov A, Tun NM, Trieu DM, Aung H (2021) Hyperspectral remote sensing images classification using fully convolutional neural network. IEEE, pp 2166–2170
37. Wang C, Gong M, Zhang M, Chan Y (2015) Unsupervised hyperspectral image band selection via column subset selection. IEEE Geosci Remote Sens Lett 12:1411–1415
38. Weber C, Aguejdad R, Briottet X, Avala J, Fabre S, Demuynck J, Zenou E, Deville Y, Karoui MS, Benhalouche FZ et al (2018) Hyperspectral imagery for environmental urban planning. IGARSS 2018 - 2018 IEEE international geoscience and remote sensing symposium, pp 1628–1631
39. Wei H, Yangyu H, Li W, Fan Z, Hengchao L, Tianfu W (2015) Deep convolutional neural networks for hyperspectral image classification. J Sens 2015
40. Ye Z, Yan Y, Bai L, Hui M (2018) Feature extraction based on morphological attribute profiles for classification of hyperspectral image. Tenth international conference on digital image processing (ICDIP 2018), 10806
41. Yi B, Li W, Du J (2012) Classification of hyperspectral data based on principal component analysis. Information-Int Interdiscip J 15:3771–3777
42. Yu C, Li F, Chang C, Cen K, Zhao M. 2019. Deep 2D convolutional neural network with deconvolution layer for hyperspectral image classification. Springer Singapore, Singapore pp 149–56
43. Yuan H, Lu Y, Yang L, Luo H, Tang YY (2013) Spectral-spatial linear discriminant analysis for hyperspectral image classification. 2013 IEEE International conference on cybernetics (CYBCONF)
44. Yue J, Zhao W, Mao S, Liu H (2015) Spectral-spatial classification of hyperspectral images using deep convolutional neural networks. Remote Sens Lett 6:468–477
45. Zhang L, Su H, Shen J (2019) hyperspectral dimensionality reduction based on multiscale superpixelwise Kernel principal component analysis. Remote Sens 11:1219
46. Zhang J, Wei F, Feng F, Wang C (2020) Spatial-spectral feature refinement for hyperspectral image classification based on attention-dense 3D–2D-CNN. Sensors 20:5191