3D autoencoder algorithm for lithological mapping using ZY-1 02D hyperspectral imagery: a case study of Liuyuan region

Junchuan Yu⊙,[a,]* Liang Zhang,[b] Qiang Li,[c] Yichuan Li,[a] Wei Huang,[a] Zhiwei Sun,[d] Yanni Ma,[a] and Peng He[a]

[a]China Aero Geophysical Survey and Remote Sensing Center for Land and Resources, Beijing, China

[b]China University of Geosciences, Beijing, China

[c]Shenyang Geotechnical Investigation and Surveying Research Institute Co., Ltd., Shenyang, China

[d]Beijing GEOWAY Spatial Co., Ltd., Beijing, China

Abstract. A hyperspectral image (HSI) contains hundreds of spectral bands, which provide detailed spectral information, thus offering an inherent advantage in classification. The successful launch of the Gaofen-5 and ZY-1 02D hyperspectral satellites has promoted the need for large-scale geological applications, such as mineral and lithological mapping (LM). In recent years, following the success of computer vision, deep learning methods have shown their advantage in solving the problem of hyperspectral classification. However, the combination of deep learning and HSI to solve the problem of geological mapping is insufficient. We propose a new 3D convolutional autoencoder for LM. A pixel-based and cube-based 3D convolutional neural network architecture is designed to extract spatial–spectral features. Traditional and machine learning methods are employed as competing methods, trained on two real hyperspectral datasets, and evaluated according to the overall accuracy, F1 score, and other metrics. Results indicate that the proposed method can provide convincing results for LM applications on the basis of the hyperspectral data provided by the ZY-1 02D satellite. Compared with traditional methods, the combination of deep learning and hyperspectral can provide more efficient and highly accurate results. The proposed method has better robustness than supervised learning methods and shows great promise under small sample conditions. As far as we know, this work is the first attempt to apply unsupervised spatial–spectral feature learning technology in LM applications, which is of great significance for large-scale applications. © *The Authors. Published by SPIE under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.* [DOI: 10.1117/1.JRS.15.042610]

1 Introduction

As one of the hottest topics in the remote sensing field, hyperspectral technology plays a significant role in Earth observation. Hyperspectral image (HSI) contains hundreds of spectral bands, which provide detailed spectral information, and thus has an inherent advantage in geological applications. Generally, most minerals and rocks have obvious spectral characteristics in the range of 400 to 2500 nm.[1] Spectral analysis of typical rocks and minerals and building a spectral database establish a good foundation for lithological mapping (LM).[2,3] Different geological bodies and formations, which vary in terms of mineral composition, weathering characteristics, alteration, and tectonic setting, also lead to a different spectral signature in hyperspectral data.[4,5] Therefore, richer spectral information and higher spatial resolution correspond to

*Address all correspondence to Junchuan Yu, yujunchuan@mail.cgs.gov.cn

its greater advantage in expressing of different types of geological bodies. Although geological mapping based on airborne hyperspectral has been performed for many years, the high cost of data acquisition has made the application and promotion of this technology more difficult. Benefiting from the development of hyperspectral satellite technology, Gaofen-5,[6,7] ZY-1 02D[8] has been successfully launched, thus providing sufficient data guarantee for large-scale LM.[9] Both Gaofen-5 and ZY-1 02D have a width of 60 km, a spatial resolution of 30 m, and hundreds of spectral bands, which are helpful for developing accurate, efficient, and low-cost geological mapping applications.

In the past few decades, various methods based on HIS have been proposed for geological mapping. Traditional lithology and mineral mapping methods can be summarized into three categories: image enhancement methods, spectral feature analysis methods, and object-oriented-based methods. Image enhancement methods such as minimum noise fraction rotation (MNF),[10,11] principal component analysis (PCA),[12] and band ratio[13,14] method aim to enhance the relevant feature of litho-units through transformation processing such as dimensionality reduction. Such methods are simple and effective, but they are mainly used to enhance the expression of lithology-related features and cannot directly achieve classification. Spectral feature analysis methods can be further subdivided into two types, spectral feature extraction (SFE) methods[15–17] and spectral matching (SM) methods.[18,19] As a method with clear physical meaning, the SFE method is based on the analysis of the diagnostic spectral features of typical mineral or litho-units and artificially identifies rules to achieve LM. However, because the algorithm mainly uses manual identification rules to implement LM, it is less efficient when applied to large-scale, multiple targets, and complex scenarios.[20] In the definition of the SM methods, rock types are distinguished by comparing the consistency of the target spectrum with that of the reference spectrum. Typical SM methods, such as spectral angle mapper (SAM)[21,22] and spectral information divergence (SID)[23] are the most commonly used. The unmixing method can also be regarded as an extension of this type of method.[24] Compared with the SFE methods, it is easier to implement, but it is not sensitive enough to minor diagnostic spectral signatures,[25] and the choice of reference spectrum is also important to accuracy. Object-wise methods should apply super-pixel segmentation,[26,27] such as simple linear iterative clustering to the target images. Although this kind of method overcomes the salt-and-pepper effect to some extent, the recognition accuracy is still affected by the initial super-pixel precision and the insufficient utilization of spectral features. The above-mentioned traditional methods have their own advantages, but the problems of insufficient spatial–spectral feature combination, weak feature extraction ability, and low efficiency are difficult to avoid.

With advances in machine learning technology, a series of learning-based methods have been proposed for LM, such as support vector machine[28,29] and random forest.[30] Compared with traditional methods, the learning-based methods can capture more effective features through supervised learning, thereby overcoming the problem of artificial threshold setting in complex scenes. In recent years, deep learning technology has achieved remarkable progress in hyperspectral application.[4,31,32] However, few application cases of LM using deep learning methods combined with HSI. As shown in Refs. 27 and 33, a convolutional neural network (CNN) was used to address LM problems based on HSI in the supervised scenario, thereby providing a good case for the application of deep learning in LM. Considering that the accuracy of supervision-based classification algorithms is constrained by the amount and representativeness of training samples and that applying locally trained models to large-scale and complex scenarios is difficult, we believe that using self-learning methods to solve LM problem is a better choice. Most recently, the autoencoder-like architecture has been developed in hyperspectral unmixing and became a new trend in self-learning methods. In Refs. 34–36, denoising and sparseness autoencoder are introduced to estimate the abundance of endmembers. In Refs. 37 and 38, the 3D-CNN autoencoders are further employed for hyperspectral unmixing to improve the classification accuracy. Inspired by these autoencoder applications, we introduce a novel end-to-end 3D convolutional autoencoder for LM. The encoder with pixel-based or cube-based CNNs is proposed to explore the spatial–spectral contextual features of HSI, and the decoder is designed to estimate the endmember of the litho-units. The novelty of this method lies in the use of a 3D autoencoder, which implements LM through self-learning and avoids the use of large training sets. In our experiments, SAM, SID, and a simple CNN network are used as competing methods on the same data

set and trained using the same parameter settings as the proposed method. Overall accuracy, F1-score, and other metrics are employed to evaluate the performance. As far as we know, this work is the first attempt to apply unsupervised spatial–spectral feature learning technology in LM application based on HSI.

The remainder of this paper is organized as follows: The proposed method is described in Sec. 2. The working area and data are described in Sec. 3. Experiments and results are presented in Sec. 4, and the conclusion and discussions are presented in Sec. 5.

2 Proposed Methods

2.1 Problem Formulation

Geological bodies and formations formed in diverse geological environments show differences in mineral composition, weathering characteristics, and alteration, thereby leading to distinct spectral signatures in hyperspectral data. Reasonably, the LM problem can be considered as an unmixing problem, which aims to estimate the proportions of each spectral endmember that representing certain litho-units. The formula can be expressed as

$$M = \Psi(EA) + N, \tag{1}$$

in which M is indicative of a mixture pixel of reflectance, E represents the endmembers of litho-units, A denotes their proportions, N is the additive vector, and Ψ represents the implicit non-linear function applied to the linear transform. The problem investigated in this paper is the estimation of the proportion matrix A by given the endmembers matrix E. It is worth noting that the proportion non-negative and sum-to-one constrain are two physical constraints[39] that restrict the model from estimating the correct result following the physical rules. The former demonstrates all elements of the estimation result, which indicates that the proportions of each litho-unit must be nonnegative, and the latter requires that the sum of proportion in each pixel equals one.

Autoencoder is an unsupervised training network composed of an encoder and a decoder, which can learn the data patterns and reconstruct input information with a minimum reconstruction error. Under certain constraints, the decoder part can be regarded as a reconstruction of HSI based on pure litho-units matrix and their proportions matrix. Hence, given the spectral endmembers of typical litho-units as the weight of the reconstructed layer, the proportion matrix of each litho-unit can be estimated as the activations of the last hidden layer for each input spectrum.

2.2 3D-CNN Autoencoder

Compared with other remote sensing data, hyperspectral data contains richer information in channel dimension. Thus, we used 3D convolution in hyperspectral data processing, which can be expressed as

$$v_{lf}^{xyz} = \sigma\left(\sum_m \sum_{h=0}^{H_k-1} \sum_{w=0}^{W_k-1} \sum_{d=0}^{C_k-1} w_{lfm}^{hwd} v_{(l-1)m}^{(x+h)(y+w)(z+d)}\right) + b_{lf}, \tag{2}$$

in which v_{lf}^{xyz} represents the value of a unit at position (x, y, z) on the f'th feature map in the l'th layer; m indexes the sets of the feature map in the preceding layer $(l-1)$; H_k, W_k, C_k denote the height, width, and channel of the kernel, respectively; w_{lfm}^{hwd} stands for the weight at position (h, w, d) connected to the f'th feature map; and b and σ are the bias and the activation function, respectively.

Inspired by the application of autoencoder for spectral unmixing problems, we introduce a pixel-based 3D convolutional autoencoder (PBA) and a cube-based 3D convolutional autoencoder (CBA). The main architectures of PBA and CBA are the same, but the inputs are different. The PBA takes a spectrum vector at each pixel as input, whereas the CBA takes the hyperspectral
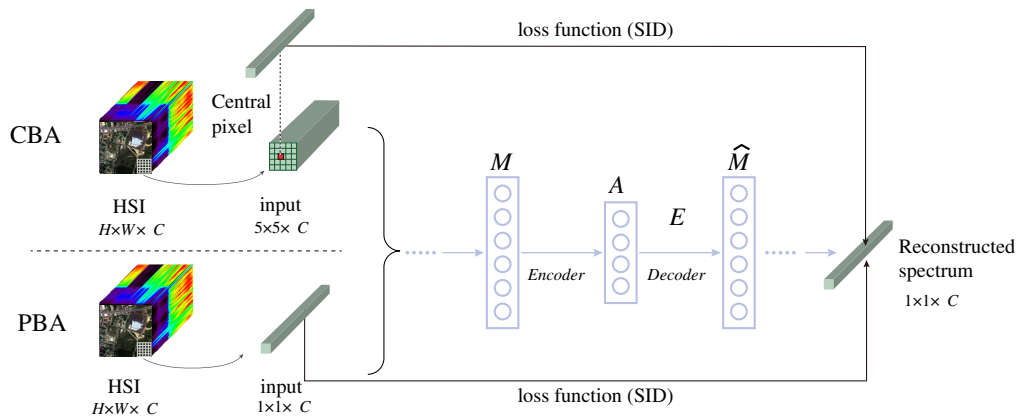
Fig. 1 Schematic diagram of autoencoder with cube-base and pixel-based input spectrum.
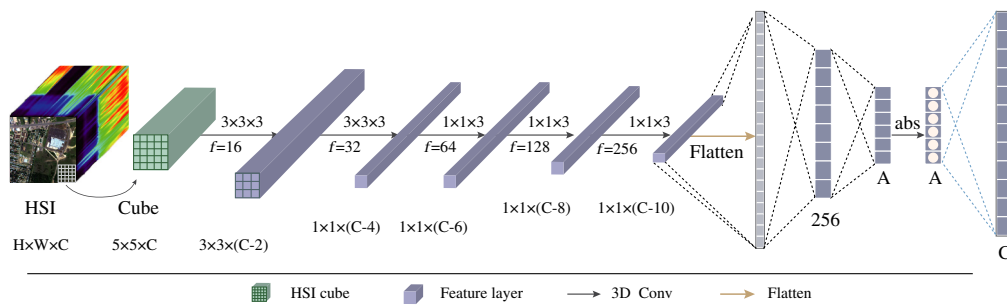

Fig. 2 Architecture of the proposed 3D-CNN autoencoder.

cube $(S \times S \times C)$ as input data to obtain joint spatial–spectral information, where S denotes the spatial window size and C is the number of spectral bands. In theory, the size of S is not fixed, it can be set according to the size of the data. In this case, S is set to 5. Generally, as an end-to-end network, the input and output of the autoencoder are the same, as in PCA. However, the output of the CBA is the central pixel of the input hyperspectral cube, as shown in Fig. 1. Similarly, the number of convolutional layers of the model's encoder is also changeable. The focus of this paper is to use the autoencoder to solve the LM problem, which is why a general encoder architecture is adopted in this case. Considering that several high-computational-cost fully connected layers (FC layers) are present in the network, we adopted five 3D convolutional layers as the encoder in this case, which can maintain sufficient feature extraction capabilities without increasing the computational burden as much as possible.

As shown in Fig. 2, an encoder with five 3D convolutional layers is designed to extract the spatial and spectral information of HSI. For the decoder part, we first use two FC layers to increase the nonlinearity of the model, and then use another FC layer to reconstruct the input spectrum. To follow the non-negative and sum-to-one constrain mentioned earlier, the absolute transformation (Abs in Table 1) and softmax activation is added before the final FC layer of the decoder. The detailed parameters of CBA are shown in Table 1.

2.3 Comparison Model

SAM and SID, which are two traditional spectral feature analysis methods, are employed as comparison methods. A supervised model with a basic CNN architecture is also used to evaluate the performance of the proposed model. Through a comparison of the angle difference between the spectrum of each pixel and the pure litho-units, the closest one is selected as the classification result. For the supervised model of deep learning in this paper, we use a basic CNN structure to obtain the probability of each category. The architecture is shown in Table 2. The network is

Table 1 The architecture of the proposed CBA model.

Model	Layer	Kernel size	Filters	Activation	Feature size
CBA (5 × 5 × C)	Conv3D-1	(3, 3, 3)	16	ReLU	(3, 3, C-2, 16)
	Conv3D-2	(3, 3, 3)	32	ReLU	(1, 1, C-4, 32)
	Conv3D-3	(1, 1, 3)	64	ReLU	(1, 1, C-6, 64)
	Conv3D-4	(1, 1, 3)	128	ReLU	(1, 1, C-8, 128)
	Conv3D-5	(1, 1, 3)	256	ReLU	(1, 1, C-10, 256)
	Flatten	—	—	—	(C-10)×256
	FC-layer-1	—	256	ReLU	256
	FC-layer-2	—	N	—	N
	Abs	—	—	Softmax	N
	FC-layer-3	—	C	ReLU	C

mainly composed of two convolution layers and two FC layers. The former is responsible for extracting the image feature and the latter is responsible for flattening the feature to 1D. Moreover, the dropout technique is utilized to prevent overfitting.

2.4 Loss Function

We use SID as the loss function of our unsupervised methods and use categorical cross-entropy as the supervised methods' loss function. SID measures the difference between the input spectrum vector and the reconstructed spectrum by the decoder. It is a measure of the similarity evaluation of two spectral curves by using the relative entropy of spectral information. The SID of input spectrum x and reconstructed spectrum y can be expressed as

$$SID(x, y) = D(x\|y) + D(y\|x). \tag{3}$$

Cross-entropy is used to evaluate the difference between the predicted result and the reference ground truth (GT) in the supervised methods. The formula can be expressed as

$$L = \frac{1}{N}\sum_i L_i = -\frac{1}{N}\sum_i \sum_{c=1}^{M} y_{ic} \log(p_{ic}), \tag{4}$$

Table 2 The architecture of the basic CNN model.

Model	Layer	Kernel size	Filters	Activation	Feature size
CNN	Conv3D-1	(3, 3)	32	ReLU	(5, 5, 32)
	Conv3D-2	(3, 3)	64	ReLU	(5, 5, 64)
	Dropout (0.25)	—	—	—	(5, 5, 64)
	Flatten	—	—	—	(1600,)
	Dense-1	—	60	ReLU	(60,)
	Dropout (0.5)	—	—	—	(60,)
	Dense-2	—	5	Softmax	(5,)

in which N represents the number of samples, M represents the number of categories, and y_{ic} is an indicator variable (0 or 1), which is 1 if the category c is the same as the category of sample i and 0 otherwise. p_{ic} denotes the predicted probability that the sample i belongs to category c.

2.5 *Evaluation*

The performance of the proposed model is quantitatively measured according to the agreements and differences between the predicted results and GTs. The most common metrics overall accuracy, recall, F1-score, and precision were used as the evaluation index to evaluate the compared methods. For reference, a general analysis of the accuracy metrics for classification tasks can be found in Ref. 40. These metrics are defined as follows:

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{5}$$

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{6}$$

$$\text{Overall Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{7}$$

$$\text{F1 score} = \frac{2}{\frac{1}{\text{Recall}} + \frac{1}{\text{Precision}}}, \tag{8}$$

where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative, respectively.

3 Working Area and Data

3.1 *Working Area*

The study area is located in Liuyuan town, northwestern Gansu Province, which is located in the middle east of the Dongtianshan-Beishan metallogenic belts, the southern margin of the Central Asian Orogenic Belt. In its long history of geological development, the Liuyuan area has experienced complex tectonic movements and magmatic activities and has excellent metallogenic conditions. The geological composition of the study area is relatively simple. According to the 1:50000 geological map, the main litho-units of the working area can be divided into five categories. Hercynian acidic intrusive rocks with different compositions occupy 50% of the study area. The Ordovician Huaniushan Formation is the main strata in the study area, which is mainly composed of plagioclase, basalt, and metamorphic sandstone. However, due to the small scale of the reference geological map, it cannot accurately reflect the edge of the geological bodies. With the aid of the high spatial-resolution and high spectral-resolution of ZY-1 02D's images, the boundaries of different geological bodies can be observed in true-color images (Fig. 3). Further, in the HSI after MNF transformation, the difference between litho-units is enhanced and can be better distinguished through color and texture. Finally, in reference to the geological map, the final GT is labeled through manual interpretation based on MNF-transformed data. Five typical litho-unit's spectrum were collected as the endmember matrix for the extraction of the proportion matrix.

3.2 *Data Acquisition*

We apply our method to two hyperspectral datasets with GT to evaluate the performance of the proposed method.

The Urban dataset is an airborne HSI obtained by HYDICE and is widely used in classification research. It contains 307×307 pixels and 210 bands in the range from 0.4 to 2.5 μm.
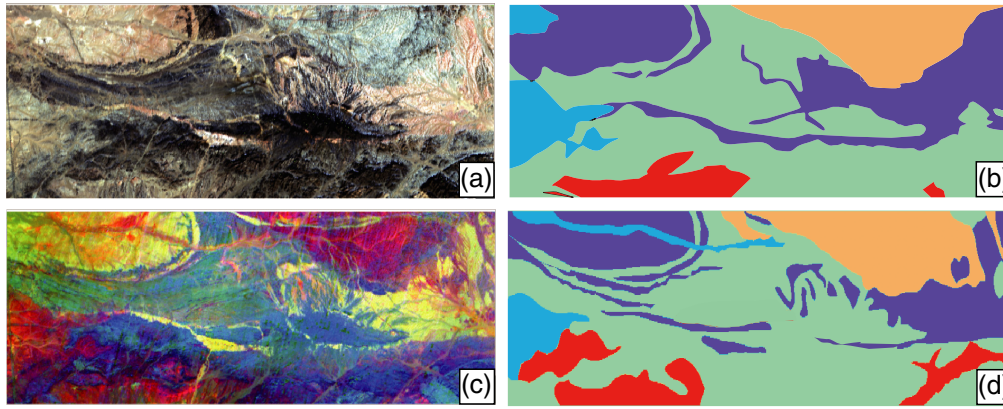
Fig. 3 (a) HSI of the working area; (b) the 1:50000 geological map; (c) the MNF-transformed HIS of the working area; (d) the GT labeled through manual interpretation.

Table 3 Technical specifications of ZY-1 02D hyperspectral sensor.

Sensor	Module	Spectral range (nm)	Bandwidth (nm)	Spatial resolution (m)	Bands
ZY-1 02D	VNIR	395 to 1040	~10	30	76
	SWIR	1005 to 2510	~20	30	90

Forty-eight bad bands are removed, and the remaining 162 bands are used for classification. The data contain six endmembers, namely, asphalt, grass, tree, roof, metal, and dirt.

ZY-1 02D was launched on September 12, 2019, and is China's first self-built commercial hyperspectral satellite. ZY-1 02D will play an important role in large-scale monitoring and quantitative application by virtue of its wide spectrum range and high spatial and spectral resolution characteristics. ZY-1 02D carries a visible and near-infrared (VNIR) multi-spectral imager and a hyperspectral sensor. As shown in Table 3, it covers a range of 0.4 to 2.5 μm and has 166 spectral bands. The spectral resolution is 10 nm for VNIR and 20 nm for shortwave infrared (SWIR). The spatial resolution is 30 m and the swash width is 60 km.

3.3 Data Processing

The experimental ZY-1 02D images were obtained in the Liuyuan area of Gansu Province on February 7, 2020. Generally, before the mineral information extraction step, the original hyperspectral data needs to be processed into reflectance data. The preprocessing of ZY-1 02D mainly includes the following steps:

1. Removal of bad bands and overlapping bands.
2. Conversion of DN to radiance by using the absolute calibration coefficient.
3. Strip noise removal by using spectral moment matching methods.
4. Correction of radiance data to reflectance by using the FLAASH atmospheric correction model.
5. Geometric correction and orthorectification.

4 Experiments and Results

4.1 Experimental Setting

The Urban and the ZY-1 02D datasets with GT are used in this experiment to evaluate the classification performance of different methods. To test and verify the robustness of the proposed method, we randomly select 1/10, 1/50, and 1/100 of the original data to construct three datasets while considering the balance of the sample number of each category. In each dataset, 70% of

these data is used as training data, and 30% is used as validation data. The experiment was performed in a TensorFlow (1.13.1) framework on an NVIDIA Tesla V100 GPU and optimized by the adaptive moment estimation (Adam) algorithm (initial learning rate as 0.001). During the training process, the SID objective function is used in the CBA and PBA models, whereas the categorical cross-entropy objective function is used in the basic CNN model. Moreover, 30 epochs are sufficient for training, and the batch size was set to 32.
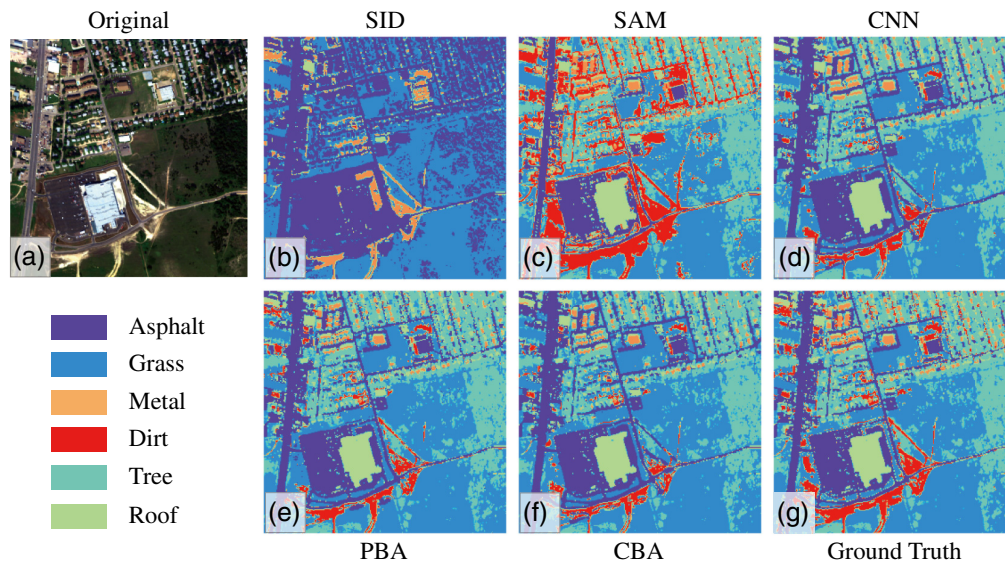
4.2 Results

4.2.1 Urban

Evaluating proposed methods on public datasets provides more convincing results. The classification performance of different methods on the Urban datasets is shown in Fig. 4. Evidently, the performance of the learning-based methods is better than those of the traditional SID and SAM methods, which have obvious misclassification issues. Interestingly, SID, which is used as the loss function in learning-based methods, obtains the worst result among all methods [Fig. 4(b)]. This finding illustrates that the importance of feature learning ability of convolutional layers for classification. Based on the visual evaluation, it is difficult to distinguish which is the best among CNN, PBA, and CBA. To quantitatively evaluate the performance of each model in the classification task, we adopted precision, recall, overall accuracy, and F1-score as the evaluation metrics. The data in Table 4 clearly show that the proposed PBA and CBA methods are better than the other comparison methods on all metrics, under all kinds of sampling ratio conditions. Moreover, as shown in Fig. 6, as the sampling ratio decreases, the F1-score of the CNN method drops significantly, whereas the performance of the proposed method is relatively stable. This finding illustrates the potential of unsupervised feature learning methods in classification applications based on a small sample.

4.2.2 ZY-1 02D

The ZY-1 02D HSI from Liuyuan is introduced to further evaluate the performance of comparison methods for the application of LM. Figure 5 shows the reference endmember spectrum of litho-units and endmember spectrum reconstructed by the proposed method. The overall shape of the reconstructed spectrum is consistent with that of the original spectrum. Although a slight difference in amplitude is observed after 2000 nm, the position of the absorption feature is


Fig. 4 Comparison of classification results of the Urban dataset using different methods. (a) Urban HSI; (b) SID; (c) SAM; (d) CNN; (e) PBA; (f) CBA; and (g) GT.

Table 4 Quantitative evaluation result of each model with different sampling ratios of training data.

| Dataset	Model	Accuracy	Precision	Recall	F1-score
1/10	PBA	0.92	0.92	0.92	**0.91**
	CBA	0.89	0.91	0.89	**0.89**
	CNN	0.88	0.89	0.89	0.88
	SID	0.48	0.67	0.48	0.38
	SAM	0.80	0.88	0.80	0.81
1/50	PBA	0.86	0.87	0.86	**0.84**
	CBA	0.87	0.89	0.87	**0.85**
	CNN	0.82	0.83	0.87	0.79
	SID	0.48	0.67	0.48	0.38
	SAM	0.80	0.88	0.80	0.81
1/100	PBA	0.88	0.86	0.88	**0.86**
	CBA	0.86	0.87	0.86	**0.84**
	CNN	0.80	0.79	0.86	0.77
	SID	0.48	0.67	0.48	0.38
	SAM	0.80	0.88	0.80	0.81
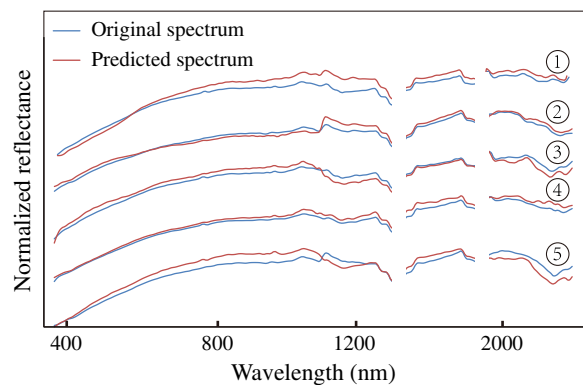
Fig. 5 Comparison between the reference endmember spectrum of litho-units and endmember spectrum reconstructed by the proposed method.

relatively consistent. Correct reconstruction of the original data indicated that the autoencoder effectively extracted the spatial–spectral features from the HSI.

We first compared the performance of CNN and autoencoder methods under different sampling ratios. In Fig. 6, we can see that, as the sampling ratio decreases, the F1-score of the CNN method presents the same decline pattern in the Urban dataset. When the sampling ratio is set to 1/100, the F1-score of CNN is lower than 0.55, whereas the F1-score of the proposed methods is stable at around 0.65 (Fig. 7).

As we mentioned earlier, the composition and formation environments are complex, resulting in variations in the spectrum. Therefore, obtaining enough representative samples to achieve LM through supervised learning methods is difficult. The application of supervised learning methods
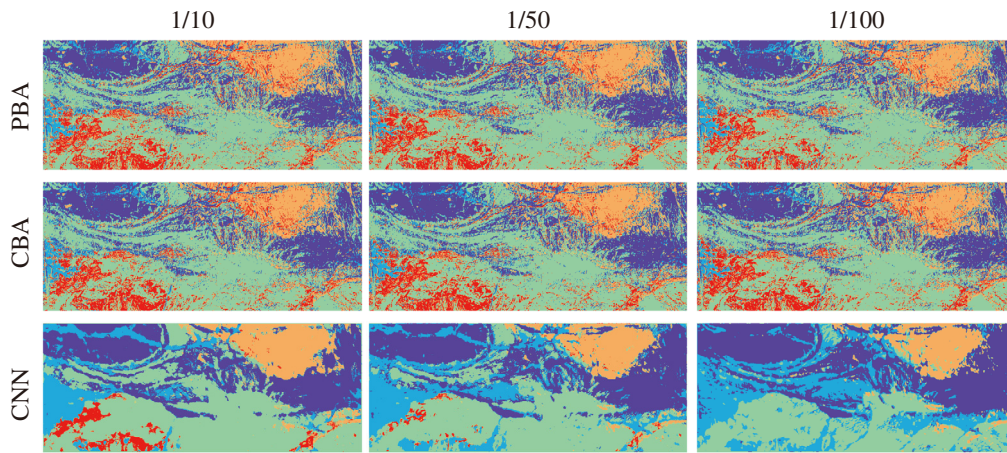
Fig. 6 LM results with different sampling ratio for LM of the working area.
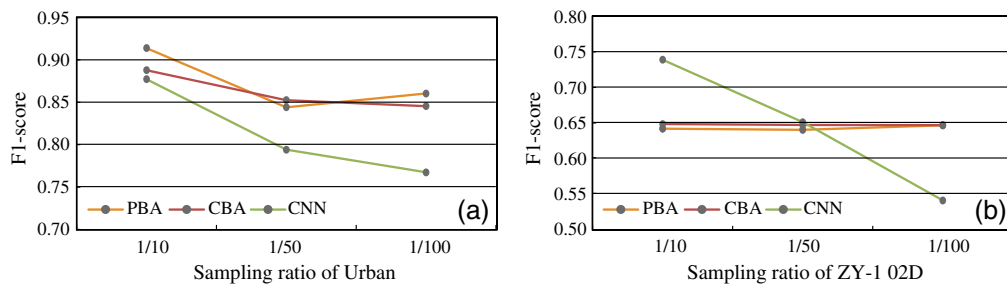
Fig. 7 F1-Score with different sampling ratios of the Urban dataset and the ZY-1 02D dataset. (a) F1-score of the Urban dataset and (b) F1-score of the ZY-1 02D dataset.

to small data is more likely to lead to overfitting, thereby leading to difficulty in maintaining robustness in larger and more complex scenarios.

Table 5 shows that when the sampling ratio is set to 1:50, the classification results of the proposed method are nearly equal to the results of the CNN method. However, CNN has obvious misclassification in the prediction of categories with a small volume of samples [Fig. 8(d)]. Figure 9 shows that the prediction accuracy of the CNN method differs for each category, whereas the CBA model has relatively good robustness in the prediction of all categories.

In this case of LM, the performance of CBA is slightly better than that of PBA (Table 5, Fig. 9) because CBA takes the HSI cube as input to obtain more spatial information. We believe that CBA is more suitable for scenes with larger classification targets; otherwise, there is no guarantee that its performance will be better than that of PBA (Table 4). Although the proposed method still has room for improvement in classification accuracy due to the inaccuracy of manual labeling, it has better robustness than the supervised learning method. Combined with

Table 5 Performance of different methods for LM with 1/50 sampling ratio of training data.

Model	Accuracy	Precision	Recall	F1-score
PBA	0.64	0.65	0.64	0.64
CBA	0.64	0.67	0.64	0.65
CNN	0.65	0.76	0.64	0.65
SID	0.24	0.41	0.24	0.25
SAM	0.39	0.61	0.39	0.37
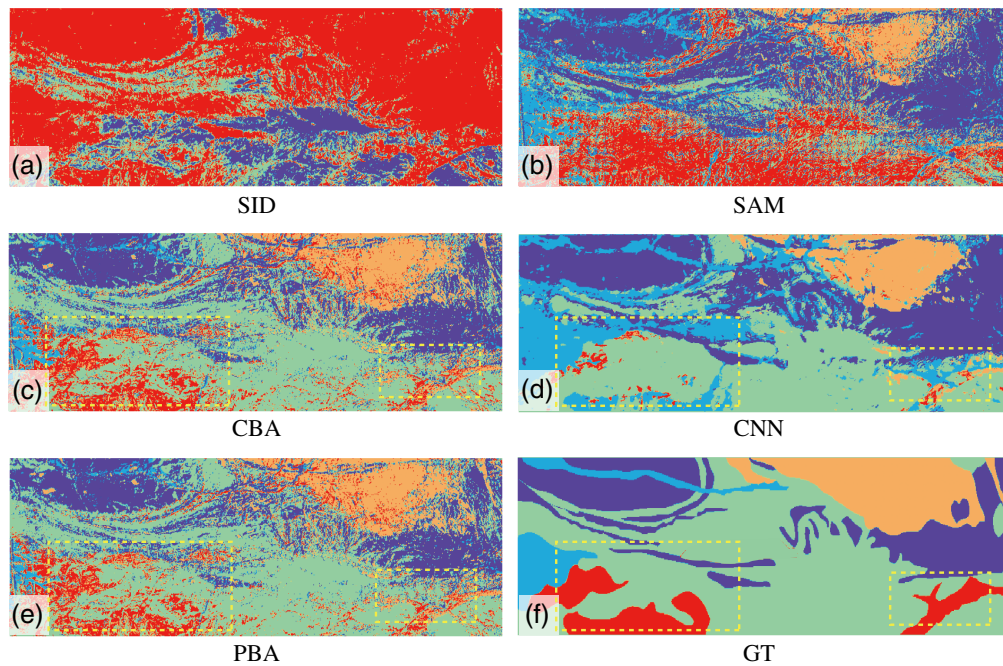
Fig. 8 Comparison of classification results for LM of the working area at 1/50 sampling ratio. (a) SID; (b) SAM; (c) CBA; (d) CNN; (e) PBA; and (f) GT.
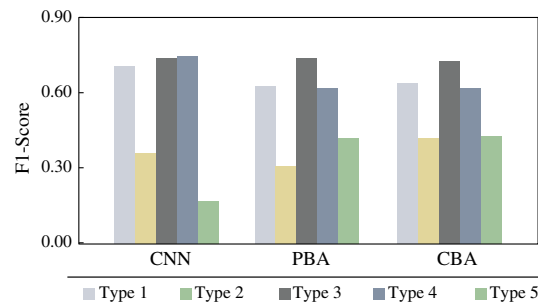
Fig. 9 Comparison of F1-score of each type of litho-units predicted by different models with 1/50 sampling ratio.

the high-quality HSI of ZY-1 02D, the proposed method shows great potential in the large-scale applications under small sample conditions.

5 Conclusion

In this work, we present a new 3D convolutional autoencoder for LM. The pixel-based and cube-based 3D convolutional architecture is designed as an encoder to extract spatial–spectral features. An FC layer with non-negative and sum-to-one constrain is employed to extract the endmember of the litho-units. The traditional methods SAM and SID and machine learning methods CNN are employed as competing methods and trained on both airborne and spaceborne hyperspectral datasets. The experimental results indicate that the proposed method can provide convincing results for LM applications on the basis of hyperspectral data provided by the ZY-1 02D satellite. Compared with traditional methods, the combination of deep learning and hyperspectral data can provide more efficient and highly accurate results. The proposed method has better robustness than the supervised learning methods and shows great promise under small sample

conditions, which is of great significance for large-scale applications based on newly spaceborne HSI payloads.

Acknowledgments

References

1. G. R. Hunt et al., "Visible and near infrared spectra of minerals and rocks. IX. Basic and ultrabasic igneous rocks" (1974).
2. R. Kokaly et al., "USGS spectral library version 7: US geological survey data series 1035" (2017).
3. I. Longhi et al., "Spectral analysis and classification of metamorphic rocks from laboratory reflectance spectra in the $0.4 - 2.5$ μm interval: a tool for hyperspectral data interpretation," *Int. J. Remote Sens.* **22**(18), 3763–3782 (2001).
4. R. N. Clark and T. L. Roush, "Reflectance spectroscopy: quantitative analysis techniques for remote sensing applications," *J. Geophys. Res. Solid Earth* **89**(B7), 6329–6340 (1984).
5. R. N. Clark, "Chapter1: Spectroscopy of rocks and minerals, and principles of spectroscopy," in *Manual of Remote Sensing, Remote Sensing for the Earth Sciences*, A. N. Rencz, Ed., Vol. **3**, pp. 3–58, John Wiley and Sons, New York (1999).
6. Y. Liu et al., "Development of visible and short-wave infrared hyperspectral imager onboard GaoFen-5 satellite," *J. Remote Sens.* **24**, 333–344 (2020).
7. Y. Yang et al., "A temperature and emissivity separation algorithm for Chinese Gaofen-5 satellite data," in *IEEE Int. Geosci. and Remote Sens. Symp.*, pp. 2543–2546 (2018).
8. Y. Zhong et al., "Advances in spaceborne hyperspectral remote sensing in China," *Geo-spat. Inf. Sci.* **24**(1), 95–120 (2021).
9. J. Yu et al., "An effective cloud detection method for Gaofen-5 images via deep learning," *Remote Sens.* **12**(13), 2106 (2020).
10. A. A. Green et al., "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. Geosci. Remote Sens.* **26**(1), 65–74 (1988).
11. M. Black et al., "Automated lithological mapping using airborne hyperspectral thermal infrared data: a case study from Anchorage Island, Antarctica," *Remote Sens. Environ.* **176**, 225–241 (2016).
12. A. Alberti et al., "Landsat TM data processing for lithological discrimination in the Caracole area (Namibe Province, SW Angola)," *J. African Earth Sci.* **17**(3), 261–274 (1993).
13. S. Dasgupta and S. Mukherjee, "Remote sensing in lineament identification: examples from western India," in *Developments in Structural geology and Tectonics*, pp. 205–221, Elsevier (2019).
14. M. W. Mwaniki, M. S. Matthias, and G. Schellmann, "Application of remote sensing technologies to map the structural geology of central Region of Kenya," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **8**(4), 1855–1867 (2015).
15. G. R. Hunt, "Spectral signatures of particulate minerals in the visible and near infrared," *Geophysics* **42**(3), 501–513 (1977).
16. R. P. Gupta, *Remote Sensing Geology*, Springer (2017).
17. C. Carli, G. Serventi, and M. Sgavetti, "VNIR spectral characteristics of terrestrial igneous effusive rocks: mineralogical composition and the influence of texture," *Geol. Soc. London Special Publ.* **401**(1), 139–158 (2015).
18. N. Xu et al., "Mineral information extraction for hyperspectral image based on modified spectral feature fitting algorithm," *Spectrosc. Spectral Anal.* **31**(6), 1639–1643 (2011).
19. R. Jain and R. U. Sharma, "Airborne hyperspectral data for mineral mapping in southeastern Rajasthan, India," *Int. J. Appl. Earth Obs. Geoinf.* **81**, 137–145 (2019).

20. A. Ghulam, R. Amer, and T. M. Kusky, "Mineral exploration and alteration zone mapping in Eastern Desert of Egypt using ASTER data," in *ASPRS Annu. Conf.* (2010).

21. X. Zhang and P. Li, "Lithological mapping from hyperspectral data by improved use of spectral angle mapper," *Int. J. Appl. Earth Obs. Geoinf.* **31**, 95–109 (2014).

22. C. Hecker et al., "Assessing the influence of reference spectra on synthetic SAM classification results," *IEEE Trans. Geosci. Remote Sens.* **46**(12), 4162–4172 (2008).

23. C.-I. Chang, "An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis," *IEEE Trans. Inf. Theory* **46**(5), 1927–1932 (2000).

24. J. B. Adams, M. O. Smith, and P. E. Johnson, "Spectral mixture modeling: A new analysis of rock and soil types at the Viking Lander 1 site," *J. Geophys. Res. Solid Earth* **91**(B8), 8098–8112 (1986).

25. J. Yu and B. Yan, "Efficient solution of large-scale domestic hyperspectral data processing and geological application," in *Int. Workshop Remote Sens. Intell. Process.*, pp. 1–4 (2017).

26. W. Wang et al., "Deep learning based lithology classification using dual-frequency Pol-SAR data," *Applied Sciences* **8**(9), 1513 (2018).

27. Y. Vasuki et al., "An interactive image segmentation method for lithological boundary detection: a rapid mapping tool for geologists," *Comput. Geosci.* **100**, 27–40 (2017).

28. M. Chakouri et al., "Geological and mineralogical mapping in Moroccan central Jebilet using multispectral and hyperspectral satellite data and machine learning," *Int. J. Adv. Trends Comput. Sci. Eng.* **9**(4), 5772–5783 (2020).

29. N. Rani, V. R. Mandla, and T. Singh, "Performance of image classification on hyperspectral imagery for lithological mapping," *J. Geol. Soc. India* **88**(4), 440–448 (2016).

30. M. Belgiu and L. Drăguţ, "Random forest in remote sensing: a review of applications and future directions," *ISPRS J. Photogramm. Remote Sens.* **114**, 24–31 (2016).

31. X. Wang et al., "Caps-TripleGAN: GAN-assisted CapsNet for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.* **57**(9), 7232–7245 (2019).

32. X. Wang et al., "CVA 2 E: a conditional variational autoencoder with an adversarial training process for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.* **58**(8), 5676–5692 (2020).

33. B. Ye et al., "Application of lithological mapping based on advanced hyperspectral imager (AHSI) imagery onboard Gaofen-5 (GF-5) satellite," *Remote Sensing* **12**(23), 3990 (2020).

34. O. Savas, K. Berk, and A. G. Bozdagi, "EndNet: sparse autoencoder network for endmember extraction and hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.* **57**, 482–496 (2017).

35. Y. Qu and H. Qi, "uDAS: an untied denoising autoencoder with sparsity for spectral unmixing," *IEEE Trans. Geosci. Remote Sens.* **57**, 1698–1712 (2019).

36. Y. Su et al., "DAEN: deep autoencoder networks for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.* **57**, 4309–4321 (2019).

37. X. Zhang et al., "Hyperspectral unmixing via deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.* **15**, 1755–1759 (2018).

38. F. Khajehrayeni and H. Ghassemian, "Hyperspectral unmixing using deep convolutional autoencoders in a supervised scenario," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, **13**, 567–576 (2020).

39. D. C. Heinz and C.-I. Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.* **39**(3), 529–545 (2002).

40. M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manage.* **45**, 427–437 (2009).

Junchuan Yu received his PhD from China University of Geosciences in 2013. He currently works in the Department of Satellite Application Research at China Aero Geophysical Survey and Remote Sensing Center for Natural Resources. His research interests include hyperspectral remote sensing, deep learning, and geological application.

Liang Zhang received his BS degree from the China University of Geosciences, Beijing, in 2019. He is currently pursuing a PhD in surveying and mapping in China University of Geosciences, Beijing. His research interests include intelligent processing of remote sensing images using deep learning methods and neural network algorithm.

Qiang Li received his MS degree from Wuhan University. He currently working in Shenyang Geotechnical Investigation and Surveying Research Institute Co., Ltd. At present, he mainly engaged in photogrammetry and remote sensing.

Yichuan Li received his PhD from China University of Geosciences in 2015. She currently working in China Aero Geophysical Survey and Remote Sensing Center for Natural Resources, engaged in hyperspectral remote sensing and deep learning.

Wei Huang is a computer engineer who works in China Aero Geophysical Survey and Remote Sensing Center for Natural Resources. His major research activities include parallel computing and computer architecture.

Zhiwei Sun graduated from Lanzhou Jiaotong University with a master's degree in cartography and geographic information system. At present, he mainly engaged in photogrammetry and remote sensing.

Yanni Ma graduated from China University of Geosciences (Beijing) with a bachelor's degree in GIS and a master's degree in surveying and mapping. At present, she works in the China Aero Geophysical Survey and Remote Sensing Center for Natural Resources, engaged in deep learning and remote sensing image information extraction.

Peng He is a PhD student at China University of Geosciences in structural geology. She currently works in the China Aero Geophysical Survey and Remote Sensing Center for Natural Resources. Her research interests are remote sensing of environment and geological disasters.