Hyperspectral Unmixing Based on Spectral and Sparse Deep Convolutional Neural Networks

Lulu Wan, Tao Chen, *Senior Member, IEEE*, Antonio Plaza, *Fellow, IEEE*, and Haojie Cai

***Abstract*—Hyperspectral unmixing refers to the process of obtaining endmembers and abundance vectors through linear or nonlinear models. The traditional linear unmixing model assumes that each mixed pixel can be represented by a linear combination of endmembers. Considering real-world situations, a sparse constraint is normally added to the linear unmixing model. However, the linear model does not take into account that the spectrum of mixed pixels is not simply linearly mixed. To fully study the mixing characteristics of ground object spectra before being imaged by the sensor, we propose a supervised unmixing architecture based on a one-dimensional convolutional neural network (CNN) by considering the spectral information and the sparse characteristics in the mixed pixel. Since 1-D CNN only considers feature learning, we combine the traditional root-mean-square error (RMSE) and ℓ_1 regularization in its loss function to minimize training error. The performance of our proposed unmixing model is assessed by comparing the unmixing results with three traditional linear sparse unmixing algorithms and the fuzzy ARTMAP neural network in a simulated dataset and three real datasets. The RMSE was used to verify the unmixing accuracy of the different methods. The results showed that the RMSE obtained by our proposed CNN-based method was the lowest among the methods on all three real datasets, proving the effectiveness and stability of the CNN in unmixing tasks.**

***Index Terms*—Convolutional neural networks (CNNs), hyperspectral unmixing, spectral information.**

I. Introduction

IN RECENT years, with the rapid development of hyperspectral technology and the wide accessibility of remote sensing images, hyperspectral images have been continuously applied in different applications, such as classification [1] and segmentation [2], change detection [3], [4], object recognition [5], and resource exploration [6]. However, due to the limitation of the spatial resolution of Earth satellite sensors and the influence of the complex diversity of the surface, a single pixel in a remotely-sensed image contains a variety of materials, leading to the presence of so-called mixed pixels [7]. The traditional classification method considers each pixel in the image as a category of ground objects, which is clearly inaccurate. Mixed pixels exist at the junctions of different ground objects, and the classification of such pixels is often difficult because the spectrum of mixed pixels is often a mixture of various ground objects, which does not refer to a single ground object type.

At present, hyperspectral unmixing models are mainly divided into linear [8] and nonlinear [9]–[11]. The physical meaning of the linear unmixing model is simple and very clear, and the results of the unmixing can basically meet the application requirements. Therefore, the linear model is still the most widely used model in the field of unmixing. The traditional linear unmixing model assumes that a variety of land cover spectra are weighted together to form a mixed spectrum of one pixel, so the unmixing process usually includes two parts: endmember extraction and abundance estimation. For example, vertex component analysis [12], the pixel purity index [13], and the simplex growing algorithm [14] are all based on the pure pixel assumption and need to extract endmembers from the dataset, which may be difficult if the resolution of the dataset is low or the features are highly mixed.

With the continuous development of theories such as compressed sensing and sparse representation [15] in the fields of classification and denoising, as well as the expansion and popularization of spectral libraries, adding sparse constraints to hyperspectral unmixing has become a major research focus. Currently, linear sparse unmixing algorithms (relying on spectral libraries) have achieved extraordinary results, and their representative methods include sparse unmixing using the variable splitting and augmented Lagrangian (SUnSAL) algorithm, which achieved good results. However, this method only considered sparsity, without considering spatial information [8]. Therefore, a hyperspectral sparse unmixing algorithm based on total variation (TV), namely SUnSAL-TV, was proposed [16]. This method can effectively promote the smoothness between pixels to optimize the understanding of mixing accuracy. Subsequently, a hyperspectral unmixing algorithm based on collaborative sparse regression (CLSUnSAL) was proposed [17] by utilizing

Manuscript received May 27, 2021; revised August 13, 2021, September 22, 2021, and October 21, 2021; accepted November 2, 2021. Date of publication November 9, 2021; date of current version November 30, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 62071439 and Grant 61871259, in part by the Opening Foundation of Qilian Mountain National Park Research Center (Qinghai) under Grant GKQ2019-01, in part by the Opening Foundation of Beijing Key Laboratory of Urban Spatial Information Engineering under Grant 20210209, and in part by the Opening Foundation of Geomatics Technology and Application Key Laboratory of Qinghai Province under Grant QHDX-2019-01. *(Corresponding author: Tao Chen.)*

Lulu Wan and Haojie Cai are with the Institute of Geophysics and Geomatics, China University of Geosciences, Wuhan 430074, China (e-mail: lucy@cug.edu.cn; cason@cug.edu.cn).

Tao Chen is with the Institute of Geophysics and Geomatics, China University of Geosciences, Wuhan 430074, China with the Geomatics Technology and Application Key Laboratory of Qinghai Province, Xining 810001, China, and also with the Beijing Key Laboratory of Urban Spatial Information Engineering, Beijing 100038, China (e-mail: taochen@cug.edu.cn).

Antonio Plaza is with the Hyperspectral Computing Laboratory Department of Technology of Computers and Communications Escuela Politécnica, University of Extremadura, 10071 Cáceres, Spain (e-mail: aplaza@unex.es).

Digital Object Identifier 10.1109/JSTARS.2021.3126755

the row sparsity characteristic shown by the abundance coefficients. Based on sparse unmixing, Feng *et al.* [18] proposed spatial regularized sparse unmixing. The algorithm combines spatial information and known standard spectral features with the traditional spectral unmixing model in the form of sparse regression to make full use of the role of spatial information in unmixing. In 2020, inspired by the feature classification of kernel functions, Weng used nonlinear feature mapping to apply HMC library decomposition to the kernel space to improve the accuracy of unmixing [19].

In the nonlinear spectral mixing model (LSMM) [20]–[22], due to the complex interactions between ground features and the sun, a model based on radiance theory usually aims at specific ground features and requires extensive prior knowledge to analyze the source of the mixed spectra. However, the nonlinear unmixing model takes into account the distortion and noise of the spectrum before it is collected by the sensor, and it is very difficult to simulate such mixed data artificially. Therefore, the nonlinear unmixing method has not been widely used. With the wide application of deep learning methods such as neural networks (NNs) in computer vision of remote sensing image classification, recognition, and detection, increasing attention has been given to the application of NN models in the study of hyperspectral unmixing, which is characterized by extracting high-dimensional features of data. Although the linear model is still the mainstream direction of unmixing research, for hyperspectral data, the linear model has not been able to make full use of the spectral information of the image. An NN method represented by a convolutional neural network (CNN) [23], [24] has a unique ability to mine spectral features. Applying it to unmixing can make full use of the deep spectral properties.

Although the historical application of an NN method in unmixing is far less than that of a linear method, the origin of its application in unmixing is not recent. For example, in 2007, Wu *et al.* proposed an NN unmixing method with endmember changes [25]. Based on the fuzzy ARTMAP NN model [26], the method utilized cross-spectral matching technology to dynamically adjust endmembers to improve the unmixing accuracy. Li proposed a remote sensing image unmixing algorithm based on an improved backpropagation (BP) NN [27]. This method needs to modify the output of the BP model. The main focus of the improvement is in the output layer node. Subsequently, methods based on deep autoencoders [24], [28] and convolutional autoencoder networks [29] were developed. Palsson *et al.* proposed a blind unmixing method of an NN autoencoder based on deep learning [30]. This method requires the number of endmembers in the image to be known in advance, the sparsity of abundance to be constrained by a custom activation function, and the original image to be reconstructed by a linear decoder. Subsequently, Qi *et al.* used CNNs to unmix hyperspectral images [31]. Ozkan and Akar extended and improved the deep spectral convolution network (DSCN) model [32]. The Wasserstein generative adversarial network was introduced into the network to improve the stability of the model and capture endmenber uncertainty. In general, the network is composed of an encoder and a decoder. The encoder introduces several mixed cores and trainable uncertainties based on NN architecture to obtain more accurate and stable solutions. Second, the abundance of each pixel was estimated by simulating the mixed Gaussian distribution, and finally, the optimal abundance estimation solution was determined according to the residual of optical endmember spectral estimation. Zhang *et al.* proposed a deconmixing method based on three-dimensional CNN (3-D CNN) [33], which can learn spatial and spectral information of images at the same time, effectively solving the spatial correlation problem of abundance in traditional mixed pixel decomposition methods. For the input hyperspectral pixels, the CNN could extract spectral and spatial features and, through network training, approximate the corresponding abundance of the hyperspectral pixels. However, these hyperspectral unmixing algorithms based on deep autoencoding use nonlinear activation functions in the encoder part of the network to extract spectral features and map them into abundance vectors. In the last layer, a simple linear decoder is added to obtain endmember spectra and reconstruct the original data. It is not reasonable to obtain the endmember spectra in such a simple and crude way, and the reconstructed data obtained in a linear way cannot reasonably account for the network error.

The supervised CNN-based unmixing network developed in this article does not need a cascade decoder, thereby reducing both the errors in network training and the complexity of the network. By comparing the unmixing results and accuracy obtained from the classical linear model and our newly proposed CNN model, the advantages of the deep learning method based on CNN in hyperspectral unmixing can be summarized as follows.

1) In the past, CNNs have been widely used in ground object classification. In this article, CNNs are used to deeply mine the characteristics of image features, extract the spectral features of hyperspectral images, gradually compress the high-dimensional spectral information to the low-dimensional spectral information, and finally map the element abundance vectors associated with the image pixels. In this article, the structure of the proposed CNN hyperspectral unmixing model has been adjusted to adapt to mixed pixel unmixing in different environments, which is of great significance in large-scale ground object inversion. In addition, in feature extraction of different spectral data, the structure adjustment of our proposed CNN can also become traceable according to actual requirements. Besides, we also streamlined the network as much as possible to make it more efficient.
2) Our newly developed pixel-based CNN model is an end-to-end framework with a simple and fast processing process and can improve the unmixing accuracy by training samples to learn the mapping process from mixed spectrum to abundance vector. In addition, the combination of the root-mean-square error (RMSE) and ℓ_1 regularization is used as the loss function of the network to achieve a sparse constraint on the abundance vector. The RMSE is used to measure the deviation between the predicted abundances and the true abundances, and ℓ_1 regularization is used to constrain the sparsity of the predicted abundance.

RMSE plays an important role in error evaluation, and the abundance vector is a continuous value between 0 and 1. Using RMSE for network error training is beneficial to measure the deviation between the observed value and real value. Second,

RMSE is equivalent to ℓ_2 regularization and will not ignore smaller values as the number of network calculations increases, so it is also more sensitive in the processing of pixel outliers. Third, combining ℓ_1 regularization in the loss function is more conducive to selecting the eigenvalues we need, and also to reduce the abundance smoothing caused by ℓ_2 regularization of RMSE.

The rest of this article is organized as follows. Section II introduces the classical linear unmixing method and the newly proposed unmixing models. Section III presents the obtained results. The discussion is given in Section IV, followed by the conclusions in Section V.

II. Methods

A. Linear Spectral Mixing Model (LSMM)

The LSMM has been widely used to determine and quantify the abundance of ground objects in mixed pixels [34], [35]. It assumes that the spectral response of a pixel in any given spectral band is a linear combination of all the endmembers in the pixel. For example, in a typical remote sensing scene, the mixed pixel spectrum can be expressed as $y = [y_1, \ldots, y_L]$, where L is the number of spectral bands. Meanwhile, let M be the endmember matrix, denoted as $[m_1, m_2, \ldots, m_q]$, where m_j is the column vector $L \times 1$ denoting the spectral vector of the jth endmember, and q is the number of endmembers existing in the observed image. Let $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_q]^T$ be an abundance vector of dimensions $q \times 1$, where α_j represents the percentage of the nth endmember in pixel y. The mixed pixel vectors can be modeled as follows:

$$y = M\alpha + n \tag{1}$$

where n is the $L \times 1$ vector, representing the noise and model error. According to the actual situation, we impose the following two constraints on the LSMM: the abundance nonnegative constraint (ANC) and abundance sum-to-one constraint (ASC):

$$\alpha_i \geq 0 \, (i = 1, 2, \ldots, q)$$
$$\sum_{i=1}^{q} \alpha_i = 1. \tag{2}$$

B. Sparse Unmixing Model

The sparse unmixing model finds the linear combination of endmembers of each observed pixel from a large spectral library and replaces the set M of endmembers with a known library A [36]. In other words, there is no need to extract endmembers from the image or to assume the existence of endmembers in the image. According to the decomposition mechanism of hyperspectral mixed pixels and the actual ground object distribution, the spectral number p in A is generally much larger than the number of endmembers q contained in each pixel, so the abundance vector of y in A is sparser. This sparsity takes into account the mixed spectra of actual ground objects and can improve the accuracy of the mixed pixel decomposition model and the stability of the abundance solution. The sparse unmixing model is written as follows:

$$\min_x \frac{1}{2} \|AX - Y\|_F^2 + \|\lambda X\|_0$$
$$s.t.\ X \geq 0,\ 1^T X = 1 \tag{3}$$

where Δ_F represents the Frobenius norm, λ is a regularization parameter, and X_0 represents the ℓ_0 -regularization of X. According to compressed sensing and sparse representation theory, ℓ_0-regularization can describe sparsity well, but the problem of ℓ_0 minimization is nonconvex and difficult to solve. Therefore, ℓ_1-regularization can be used to replace ℓ_0-regularization, so a relatively satisfactory sparse solution can be obtained. As a result, (3) can be rewritten as follows:

$$\min_x \frac{1}{2} \|AX - Y\|_F^2 + \|\lambda X\|_{1,1}$$
$$X_{1,1} = \sum_{j=1}^{n} \|x_j\|_1$$
$$\text{s.t. } X \geq 0,\ 1^T X = 1 \tag{4}$$

where x_j represents the j th column of the abundance matrix X. The model can be solved using the SUnSAL [8] algorithm based on alternate iterations. However, since the ℓ_1 regularization is not sufficient to describe signal sparsity, the unmixing accuracy is affected to some extent. Therefore, Iordache *et al.* adopted $\ell_{2,1}$-mixed regularization to replace ℓ_1-regularization and then proposed an unmixing algorithm based on CLSUnSAL [17]:

$$\min_x \frac{1}{2} \|AX - Y\|_F^2 + \|\lambda X\|_{2,1}$$
$$\|X\|_{2,1} = \sqrt{\sum\nolimits_{j=1}^{n} {x_j}^2}$$
$$s.t.\ X \geq 0,\ 1^T X = 1 \tag{5}$$

The unmixing algorithm based on collaborative sparsity effectively solves the problem of abundance sparsity, but it does not take into account the local spatial smoothness of the image, resulting in the lack of local spatial information, which further affects the understanding accuracy of mixing. Therefore, Iordache *et al.* proposed SUnSAL-TV [16], whose model can be expressed as

$$\min_x \frac{1}{2} \|AX - Y\|_F^2 + \lambda \|X\|_{1,1} + \lambda_{\mathrm{TV}}(X)$$
$$\text{s.t. } X \geq 0,\ 1^T X = 1 \tag{6}$$

where $\mathrm{TV}(X) \equiv \sum_{\{i,j\} \in \varepsilon} \|x_i - x_j\|_1$ promotes similar abundance coefficients between adjacent pixels, x_j represents adjacent pixel sequences of x_i in the abundance matrix X, x_i and x_j are column vectors of abundance matrix X, ε represents the horizontal and vertical neighborhood sets in X, and the sparsity of the above models is added to the abundance coefficients.

C. Fuzzy ARTMAP Neural Network (NN)

Adaptive resonance theory (ART) [37] is a self-organized NN proposed in 1976 by G. A. Carpenter for the unified mathematical theory of human psychological and cognitive activities. The

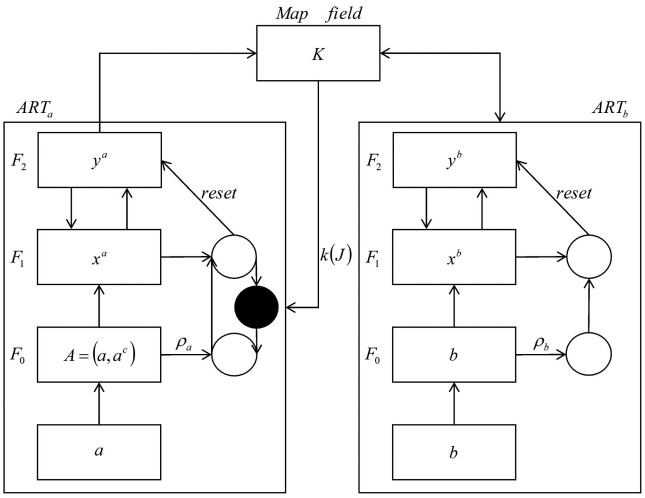

Fig. 1. Fuzzy ARTMAP network.

fuzzy ARTMAP NN combines fuzzy set theory and the ART self-organizing feedback function to realize real-time nonoffline learning [25]. The network consists of two ART modules (ARTa and ARTb) plus a mapping field (MapField). The mapping domain connects the two ART modules and can conduct real-time supervised learning for any input vector [38]. In the learning process, tracking matching rules are also implemented so that errors in the matching process do not occur repeatedly. Fig. 1 shows a graphical illustration of fuzzy ARTMAP.

The workflow of mixed pixel decomposition using the fuzzy ARTMAP NN [25], [26] is as follows.

1) Training phase: $\mathrm{ART_a}$ receives an input a in the training mode, whereas $\mathrm{ART_b}$ receives the expected output b in the training mode.
2) Test phase: During the discrimination, $\mathrm{ART_a}$ receives the input and obtains the learning result from $\mathrm{ART_b}$. The function of MapField is used to control the learning of association mapping between $\mathrm{ART_a}$ classification and $\mathrm{ART_b}$ classification. It controls the warning parameters until it can drive the network to look for better $\mathrm{ART_a}$ classification. Therefore, the size of the warning parameter controls the classification accuracy to some extent.

Fuzzy ARTMAP has superior adaptability and can maintain the same sparse and nonnegative distribution as the label abundance without changing the internal structure of the network, which is in sharp contrast to the CNN structure. The fuzzy ARTMAP NN for mixed pixel decomposition is a recognition algorithm with supervised learning [25]. This algorithm is relatively complex, and its variables and parameters are given in Table I.

TABLE I
FUZZY ARTMAP PARAMETER TABLE

Model	Parameters	Range	Value
ARTa Component	M_a	-	Band Number
ARTb Component	M_b	-	Endmembers
Selection Parameter	α	$(0,\infty)$	10^{-6}
ARTa Baseline Warning Parameter	ρ_a	[0,1]	0
ARTa Warning Parameter	ρ_b	-	-
ARTb Warning Parameter	ρ	[0,1]	0.8
Matching Pursuit	ε	tiny	0.01
ARTa Number of Coded Nodes	C_a	-	Increasing
ARTb Number of Coded Nodes	C_b	-	Increasing

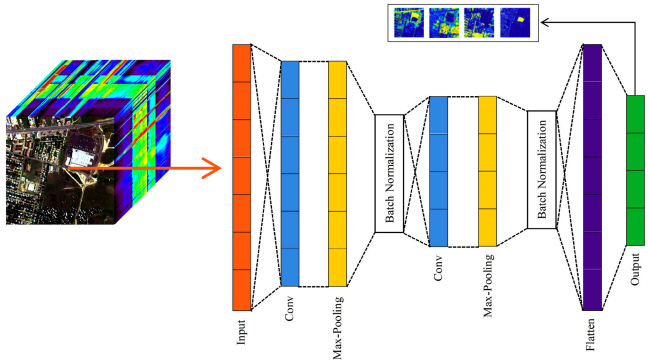

Fig. 2. Structure of the proposed CNN.

D. CNN Unmixing Model

In this section, the structure of the CNN and its application in hyperspectral unmixing are described in detail. Since machine learning was proposed, a variety of NN structures have been proposed to solve a variety of complex problems, and CNN is one of its representative models. Based on the linear unmixing model, the spectral mixing process and the sparse principle based on the spectral library are explained in this article. On this basis, the principle and role of the CNN in unmixing are studied, and the linear and NN models are compared. Combined with the advantages of the CNN in feature extraction and other aspects, it is introduced into hyperspectral unmixing to extract spectral features in depth and improve the unmixing accuracy.

A CNN is a deep learning model, which is similar to the multilayer perceptron of artificial NNs [39]–[41]. It is widely used in the field of image visual analysis. The model was introduced by LeCun [42], and it was the first to solve the problem of handwritten numbers in the MNIST dataset through a CNN. A CNN is characterized by local connections, weight sharing and downsampling, which greatly reduces the parameters that the network needs to train, thus reducing the complexity of the model, reducing the risk of overfitting, and making the model highly robust. The structure of a general CNN consists of an input layer, several continuous convolutional layers and pooling layers, a fully connected layer, and an output layer. The most commonly used activation functions include Sigmoid, Tanh, ReLU, and LeakyReLU. In this article, image features are extracted by a deep network, and then the abundance of hyperspectral unmixing is approximated through the final fully connected layer. The structure of the proposed CNN used for unmixing is shown in Fig. 2.

1) Convolution Operation: The convolutional layer is the most important layer in the CNN, and it is the key layer to extract image features [43], as shown in Fig. 3. During the convolution operation, the size of the convolution kernel (filter), namely

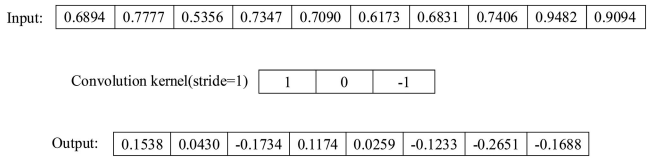

Fig. 3. Convolutional layer structure diagram.

M×M (M is the size of the convolution kernel), should be given first, and the size of the matrix is also called the receptive field. According to the data dimension features to be extracted, the depth of the convolution kernel is consistent with the depth of the input layer, so the size of a convolution kernel of $M\times M\times N$ (N is the depth of the convolution kernel) can be obtained. Assume that the size of the input image is 4×4, the size of the convolution kernel is 2×2, and the moving step is 1. When the convolution kernel moves on the input image in turn, an output image of size 3×3 can be obtained. It can be seen that the size of the convolution kernel varies with M and N. When $M = 1$, the input data is 1-D spectral data, and the size of N determines the number of spectral features extracted during each convolution operation. When N is too small, the more feature layers are obtained, but it may also cause feature redundancy, thereby reducing the unmixing accuracy; otherwise, it may cause the network to lose important features. Therefore, the value of N needs to be set according to different data dimensions.

Previously, we talked about how the convolutional layer extracts features, but the features extracted by the convolutional operation are random, so it is necessary to tell the network which features are useful. Similar to feature extraction by many NNs, CNN perceives the local spectral information through kernel function in the process of extracting spectral features, and then synthesizes the previously extracted local features at a higher level to obtain global information. The nonlinear feature extraction of spectral dimension is opposite to linear unmixing. In short, linear unmixing is the multiplication of the endmember matrix and abundance vector. CNN can extract spectral features layer by layer and map them to abundance vectors. In this process, features extracted by kernel and activation functions cannot be explained by a single linear equation.

2) Pooling operation: The pooling layer, which can also be called the downsampling layer, is the transition layer between the convolutional layers [44]. Its goal is to gradually reduce the dimensionality of the input data, reduce the number of parameters in the network, shorten the calculation time, and effectively control overfitting. The pooling and convolution layers need to set the size of the sliding window, mainly including the max-pooling layer and mean-pooling layer. In the forward propagation of max-pooling, the maximum value in the window is passed to the next layer, while the value of other pixels is directly discarded. In the BP, the gradient is directly passed to a pixel in the previous layer, whereas other pixels do not accept the gradient. The forward propagation of average pooling calculates the average value of a window and passes it to the next layer, and the BP divides the gradient of an element into n parts that are passed back to the previous layer.

3) Batch Normalization (BN): The input data need to be standardized before entering the deep NN, but in the subsequent process of feature extraction, as with convolution and pooling, it is easy to cause drastic changes in the output data as the model parameters are constantly updated. BN [42] calculates the mean and standard deviation of small batch data and constantly adjusts the output data size of the network middle layer to keep the data size of different middle layers of the whole network in a relatively stable range. Let $x_i, i = 1, \ldots, n$ be the input value of the batch processing normalized layer, and the final standardized output value is $y = \mathrm{BN}_{\gamma,\beta}(x) = \gamma\bar{x} + \beta$. The specific processing process is as follows:

$$y = \gamma\frac{x-\mu_\beta}{\sqrt{\sigma_\beta^2+\varepsilon}} + \beta$$

$$\mu_\beta = \frac{1}{n}\sum_{i=1}^{R} x_i$$

$$\sigma_\beta^2 = \frac{1}{n}\sum_{i=1}^{R}(x_i-\mu_\beta)^2 \tag{7}$$

where γ and β are learnable parameters and ε is a very small number.

4) Fully connected layer: All the fully connected layers [45] have the effect of a classifier in the CNN. In front of the convolution, pooling layer and layer activation function operations, such as mapping to the hidden layer, is the original feature space. We expand the output data of the last convolution layer, connect each node of the current layer with the node of the next layer, and set the number of categories according to the activation function and the number of nodes of the last fully connected layer. Different from hyperspectral classification and target recognition, we use the softmax activation function for the last layer, which makes the output value within the range of [0,1], satisfying both ASC and ANC conditions, and can be expressed as the abundance of hyperspectral unmixing. According to the softmax function, the abundance a_i corresponding to class i is given by the following formula:

$$a_i = e^{b_i}\Big/\sum_{j=1}^{q} e^{b_i} j = 1, 2, \ldots, q \tag{8}$$

where b_i represents the output value of the jth endmember, and q represents the number of endmembers.

5) Network Optimization Process: The use of a CNN for supervised unmixing means that a large number of labeled samples are needed in the model training process. In this article, part of the real hyperspectral data and the corresponding real abundance data were input into the network for training, and then the whole hyperspectral dataset was tested and the final abundance map was obtained. The experimental datasets we selected were all verified public datasets, and their true abundance vectors were all obtained by presumed measurement. The number of endmember categories also determines the dimension of the abundance vector. For example, in the Urban dataset, the numbers of endmembers are four and six. This is mainly based on the reality of our degree of subdivision of different ground objects.

CNN structure in this article is mainly used to extract spectral features. Although spatial features are lost in this process, spatial features are added to the subsequent loss function. Since NNs are an infinite approximation problem, the estimation from spectral dimension to abundance is not only a supervised classification problem but also an approximation fitting problem. The pooling layer in the network uses the maximum pooling layer, and the activation function used is ReLU. This activation function not only overcomes the problem of gradient disappearance but also accelerates the training speed. Therefore, when ReLU is used as the activation function, the value of the input data will not change greatly and remains in a relatively stable interval, thus ensuring the training network of the deep CNNs. Its expression is

$$f\ (x) = \begin{cases} x, x \geq 0 \\ 0, x < 0 \end{cases}. \tag{9}$$

In the network training stage, the CNN parameters were randomly initialized, the error BP algorithm was used for training, and the parameters were updated using the mini-batch strategy. To find the optimal NN structure and obtain the optimal unmixing result, the loss function between the predicted abundance and the true index in the training is minimalized

$$L = \mathrm{RMSE}\left(X, \bar{X}\right) + \lambda \ell_1 (X)$$
$$L_{\mathrm{RMSE}} = \mathrm{sqrt}\left(\frac{\left|\bar{X} - X\right|^2}{n}\right) \tag{10}$$

where X is the predicted abundance of the network, $\bar{X}$ is the corresponding true abundance value, and λ is the regularization parameter ℓ_1. Since the abundance vector we finally obtained is a continuous value, the RMSE can better measure the error between the predicted value and the true value in network training rather than calculate the degree of similarity between them. In the formula, the first term is the error term, which is used to calculate the error before the predicted abundance and the true abundance. The second term is the sparse regularization term, which is used to constrain the sparsity of the predicted abundance through ℓ_1 regularization. In the network training process, we not only need to ensure that the training error is as small as possible, but also hope that the predicted abundance has a certain sparseness, so we need to add the ℓ_1 regularization term. From (4), we know that ℓ_1 regularization is the product term of the absolute value of the predicted value X and the parameter. When calculating the gradient, the derivative of (10) can be obtained as follows:

$$\partial L \partial X = \partial L_{\mathrm{RMSE}} \partial X + \lambda \Delta \mathrm{sign}\,(X) \tag{11}$$

where sign is a symbolic function. When $X > 0$, $\mathrm{sign}(X) = 1$; when $X = 0$, $\mathrm{sign}(X) = 0$; and when $X < 0$, $\mathrm{sign}(X) = -1$. We further analyze the regularization term and let $\frac{\partial L_1}{\partial X} = \mathrm{sign}(X)$. According to $X > 0$, the update formula of $\ell_1(X)$ is $X = X - \lambda * 1$, which means that X will be reduced by a specific value every time it is updated (for example, $\lambda = 0.05$), then after several iterations, X may be reduced to zero. It can be seen from this process that ℓ_1 regularization can make the predicted value close to zero become zero faster, resulting in a sparse effect.

In practical application, both ℓ_1 and ℓ_2 regularization can be used to reduce the risk of overfitting, but ℓ_1 regularization more easily obtains sparse solutions than ℓ_2 regularization; that is, the weight vector obtained has fewer nonzero components. Second, compared with 2-D and 3-D CNNs, the pixel-based 1-D CNN used in this article only considers the spectral features of pixels in feature extraction but does not consider the spatial correlation features of images. However, the modification of the loss function is sufficient to compensate for the lack of abundance space features, and the computation time and speed of a 1-D CNN are much faster than those of a 2-D CNN. Therefore, to satisfy the abundance accuracy, the efficiency of a 1-D CNN is higher, which is also one of the important reasons why a 1-D CNN is used in hyperspectral unmixing in this article.

There is no doubt about the superiority of a CNN in feature extraction, but the network structure also greatly affects the network performance and learning efficiency. Therefore, the proposed CNN structure in this article aims to deal with different mixed image metadata and simplify the network structure as much as possible to improve the computing speed. Otherwise, the learning ability of CNN cannot be targeted only at a single type of data, that is, it is difficult for CNN structure to adapt to the characteristics of two data types at the same time, even if the same type of data also has the problem of migration. Therefore, it is very important to adjust CNN structure to adapt to various remote sensing image data.

III. Experiments and Results

A. Parameter Settings

In this article, RMSE is used to evaluate the accuracy of the mixed pixel decomposition of the four methods.

After many experimental tests, in the linear methods, i.e., SUnSAL, CLSUnSAL, and SUnSAL-TV, λ was set as 0.001, λ_{TV} was set as 0.003, and the number of iteration epochs was 4000. The set values of these parameters can not only obtain effective unmixing results but also ensure the operational efficiency of the model. Through many experiments, it is most appropriate to set the warning parameter of fuzzy ARTMAP as 0.8. Besides, to fully compare the unmixing effects of different CNN methods, 3-D CNN [33] and SPMCNN-ESPCN [46] were selected to conduct an experimental comparison with our proposed CNN methods.

The Proposed CNN needs to set different network structures for different data. In the experiment, λ in the loss function was set as 0.05, the size of the mini-batch was set as 30, and the epoch was set as 100. Specific parameters are shown in Table II, where FC is the fully connected layer. The table represents a total of 12 layers of networks, whose values represent the number of output neurons of the hidden layer. The Adam optimizer was used for network training, and the learning rate of the optimizer was set as 0.0001. Compared with other optimizers, Adam is more suitable for correcting bias errors and has good robustness in the selection of super parameters. Finally, fuzzy ARTMAP

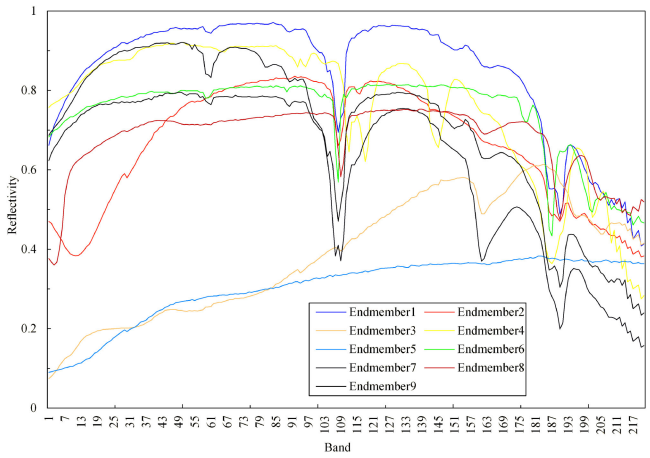

Fig. 4. Spectral curves of nine endmembers of simulated data.

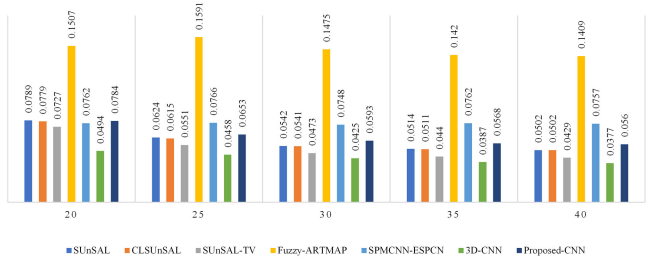

Fig. 5. RMSE with respect to different noise levels.

and the proposed CNN use 30% of the dataset for training and 100% of the dataset for testing.

B. Simulated Dataset

The dataset is 100×100 in size with 221 bands and was created to benchmark the accuracy of spectral decomposition provided in the HyperMix tool [21]. The mixed pixels use the fractal representation of some natural objects, including clouds, mountains, shorelines, and vegetation, to provide a baseline for simulating the spatial patterns often found in nature. The composite images were modeled from a linear mix of a randomly selected set of endmember features from a spectral library compiled by the United States Geological Survey, totaling 420 features. In a composite image, each scene is fixed with 9 endmembers, and the spectral curve is shown in Fig. 4. In addition, different levels of Gaussian noise (SNR = 20, 25, 30, 35, and 40 db) are added to the image.

Fig. 5 shows the RMSE with respect to different noise levels. Except for the fuzzy ARTMAP, linear sparse unmixing methods and the three CNN-based methods both show good unmixing potential. The simulated dataset was also designed to test the unmixing performance of different methods and then to evaluate the impact of different levels of noise on the unmixing performance by adding different SNRs. In general, with the increase in SNR, most of the methods show a downward trend. In principle, sparse constraints have a certain denoising capability, which makes the unmixing results stable in different noise environments. The three linear sparse unmixing algorithms all show good unmixing performance in different noise environments. In the case of a large number of endmembers, sparse regularization imposes a good constraint on the abundance of endmembers. Although the proposed CNN did not achieve the best results in this experiment, its accuracy was always stable and not much different from the results of the sparse linear decomposition.

Due to the large number of endmembers in this dataset, we only selected the abundance maps of endmembers 1, 5, and 9 for quantitative evaluation, as shown in Fig. 6. As seen from the original spectral curves in Fig. 4, except for endmembers 3 and 5, other spectral curves have similar peaks or troughs, which creates great difficulty in unmixing. The phenomenon of "foreign bodies in the same spectrum" often appears in hyperspectral images. Especially when the NN learns these spectral features, a large number of similar spectral curves are superimposed together, making it difficult for the model to distinguish these features in the learning process. Although the accuracy of fuzzy ARTMAP is the lowest and the accuracy of the proposed CNN is slightly lower than that of the linear method and 3-D CNN, from the visual effect, the noise of the two NN methods in the dark region is less than that of the linear sparse methods. This dataset is simulated using the K-means algorithm and contains a large number of linear relationships. Although noise is added, it is still not possible to simulate the nonlinear relationship in the real dataset. In contrast, the decomposition results obtained by SPMCNN-ESPCN method failed to distinguish the three kinds of endmember abundance. However, 3-D CNN and proposed CNN in this article have better results, and 3-D CNN has the lowest RMSE among all methods.

C. Real Data Experiments

We evaluated all approaches on three real-world datasets, Samson, Jasper Ridge, and Urban, as shown in Fig. 7, from different sensors with different sizes, resolutions, and the number of endmembers to fully compare the effectiveness of the different approaches.

1) Samson Dataset: The image contains 952×952 pixels, and each pixel contains 156 bands, with a wavelength range of 401–889 nm and a spectral resolution of 3.13 nm. Because the original image was too large, we extracted an area of 95×95 size for the experiment, as shown in Fig. 7(a). The image contains rock, tree, and water [as shown in Fig. 7(d)]. In the data experiment, the parameters of the CNN network structure are shown in Table II.

Table III shows the RMSE quantitative comparison results of all the methods, among which the proposed CNN method has the best result, and the RMSEs of the three kinds of ground objects are all lower than those of the other methods. The result of fuzzy ARTMAP is slightly worse than that of the CNN, which reflects the advantage of a deep CNN over a shallow network. Although the error of the linear sparse unmixing algorithm is not more than 0.06, compared with the accuracy of the NN, which is less than 0.03, the error is still too large.

Fig. 8 shows the unmixing results obtained by all methods for Samson Dataset. It can be seen that the SUnSAL, CLSUnSAL, and SUnSAL-TV linear unmixing methods have different

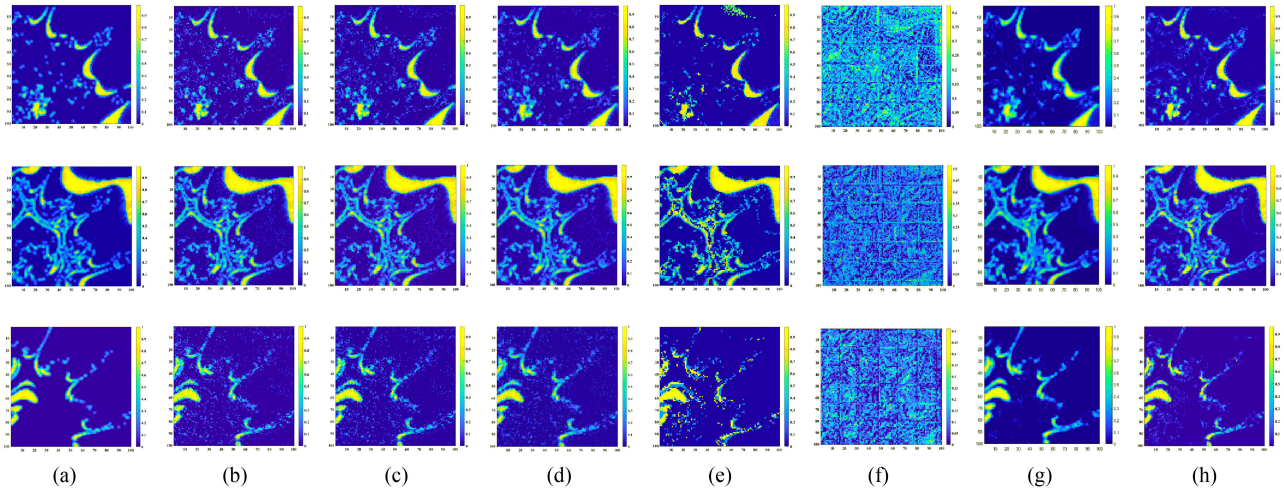

Fig. 6. Abundance maps obtained by all methods for the simulated dataset with no noise. (a) Reference abundances. (b) SUnSAL. (c) CLSUnSAL. (d) SUnSAL-TV. (f) SPMCNN-ESPCN. (g) 3-D CNN. (h) Proposed CNN.

TABLE II
CNN NETWORK STRUCTURE FOR EACH DATASET

Layer name		Input-1	Conv-2 Maxpool-3 BN-4	Conv-5 Maxpool-6 BN-7	Conv-8 Maxpool-9 BN-10	L11	L12
Output Size	Simulated Dataset	1 × 221	1 × 27R	1 × 9R	1 × R	FC	1 × R
	Samson	1 × 156	1 × 27R	1 × R	1 × R		
	Jasper Ridge	1 × 198	1 × 32R	1 × 16R	1 × 4R		
	Urban-4	1 × 162	1 × 16R	1 × 4R	1 × R		
	Urban-6	1 × 162	1 × 18R	1 × 6R	1 × R		
	Feature map	1	64	256	128		
	activation function	-	ReLU	ReLU	ReLU	-	softmax

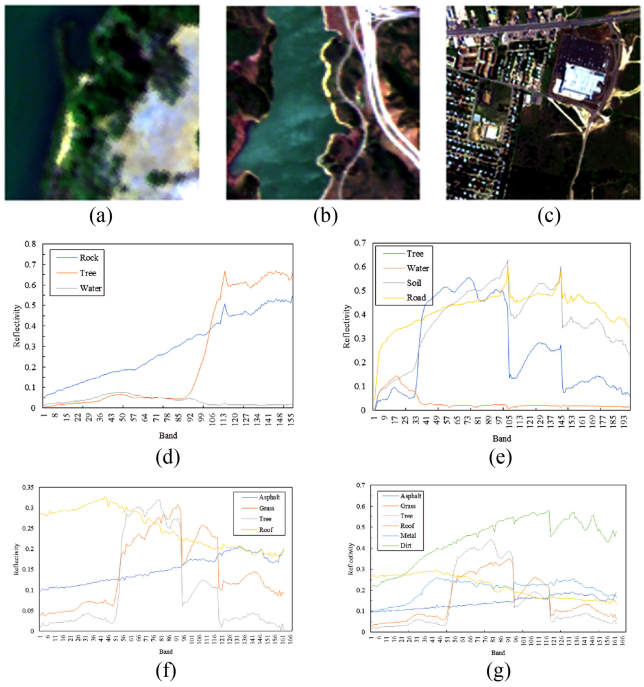

Fig. 7. Three real-world datasets used in this study. (a) Samson. (b) Jasper Ridge. (c) Urban. (d) Endmember spectrum of Samson. (e) Endmember spectrum of Jasper Ridge. (f) and (g) Endmember spectrum of Urban.

TABLE III
RMSE ON SAMSON DATA (THE BOLDED NUMBER MEANS THE LOWEST RMSE IN EACH ROW)

RMSE	Rock	Tree	Water	Mean
SUnSAL	0.0779	0.0441	0.0487	0.0588
CLSUnSAL	0.0775	0.0439	0.0486	0.0586
SUnSAL-TV	0.0804	0.0436	0.0429	0.0583
Fuzzy-ARTMAP	0.0347	0.0316	0.0124	0.0280
SPMCNN-ESPCN [46]	0.2328	0.2458	0.2458	0.2341
3D-CNN [33]	0.0159	0.0171	0.0180	0.0170
Proposed-CNN	**0.0144**	**0.0111**	**0.0094**	**0.0118**

degrees of misclassification of water, which are mainly concentrated in the upper right corner of the image, namely the tree and the shadow area. Because the light is blocked, the tree in this area appears as a black shadow on the image, and its spectrum is similar to that of water, thus increasing the abundance of water in this area. Through training and learning, the four NN algorithms have achieved good unmixing results, whereas the abundance maps obtained by the proposed CNN model are the best, which also reflects that the proposed CNN has a better capability of feature learning and extraction. On the basis of the design of the proposed CNN structure, this article adds ℓ_1 regularization to the constrained abundance sparsity in the loss function, which makes the abundance value more consistent with

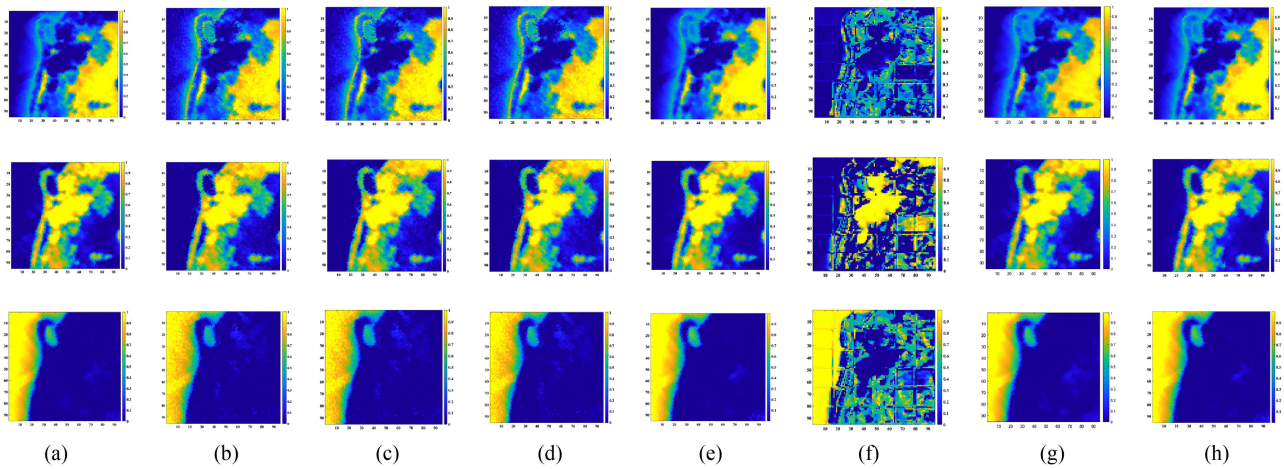

Fig. 8. Abundance maps obtained by all methods for Samson Dataset. (a) Reference abundances. (b) SUnSAL. (c) CLSUnSAL. (d) SUnSAL-TV. (e) Fuzzy ARTMAP. (f) SPMCNN-ESPCN. (g) 3-D CNN. (h) Proposed CNN.

TABLE IV
RMSE ON JASPER RIDGE DATA (THE BOLDED NUMBER MEANS THE LOWEST RMSE IN EACH ROW)

RMSE	Tree	Water	Soil	Road	Mean
SUnSAL	0.0668	0.1014	0.0697	0.0682	0.0779
CLSUnSAL	0.0311	0.0449	0.0502	0.0676	0.0502
SUnSAL-TV	0.0257	0.0426	0.0472	0.0681	0.0483
Fuzzy-ARTMAP	0.0296	0.0170	0.0415	0.0338	0.0317
SPMCNN-ESPCN [46]	0.1401	0.1572	0.1148	0.1224	0.1336
3D-CNN [33]	0.0235	**0.0144**	0.0294	0.0241	0.0235
Proposed-CNN	**0.0131**	0.0162	**0.0195**	**0.0210**	**0.0177**

the actual ground object distribution. However, there are only three kinds of ground objects in the Samson dataset. A large number of pixels in the image are pure pixels, and mixed pixels are mainly distributed at the junction of different ground objects. ℓ_1 regularization makes the abundance value of pixels close to zero according to the extracted feature part. Compared with the other two improved methods based on CNN, the RMSE obtained by the method proposed in this article is the lowest at each endmember. Especially, the result based on SPMCNN-ESPCN is the worst, while the image block boundary effect seriously affects the decomposition accuracy, and the decomposition result between adjacent blocks is obviously not smooth.

2) Jasper Ridge Dataset: The image comprises 512 $\times$ 614 pixels, the spectral range of each pixel is 380–2500 nm, a total of 224 bands, and the spectral resolution is 9.46 nm. Due to the complexity of the original image, we extracted a subimage of 100 $\times$ 100 pixels in the experiment, removed bands 1–3, 108–112, 154–166, and 220–224 caused by water vapor and atmospheric effects, and retained 198 bands. The preprocessed image is shown in Fig. 7(b), which includes four endmember ground features: tree, water, soil, and road [as shown in Fig. 7(e)].

Table IV shows the RMSE for all methods. Clearly, except for the RMSE value of water endmember, the RMSE value of the other three endmembers and the average RMSE obtained by the proposed CNN method is the lowest, indicating the advantages of the proposed method.

Fig. 9 shows the abundance maps obtained by all methods. In the figure, the brighter the region, the higher the abundance value, the darker the region, the lower the abundance value, and the lower the proportion of the endmember. First, the seven methods clearly distinguish the four kinds of ground features, which are very close to the reference abundance maps. However, in the water abundance map obtained by three linear sparse algorithms, the road area on the right side is brighter and the abundance value is higher, which is also the main reason for higher water error than other methods. Second, in the distribution of ground features, the shape of the road is relatively long and narrow. If the spatial resolution of the image is not high, the road is easily mixed with other ground features, which brings difficulty in classification and unmixing. Similar to the results of Samson dataset, the average RMSE of the proposed method is still the lowest, and 3-D CNN also achieves good results. Therefore, compared with the blocking results, the global analysis of each pixel spectrum is very important. However, the blocking effect of SPMCNN-ESPCN is likely to make the spectral features of individual pixels over-amplified. The overall error of the image block is large.

3) Urban Dataset: Urban is one of the most widely used hyperspectral image datasets in the study of hyperspectral unmixing. The data were obtained from the HYDICE sensor in October 1995. The image is of Texas in the United States, and the size is 307 $\times$ 307 pixels. The dataset has a broad spectrum with a total of 210 bands, a spectral resolution of 10 nm, and a spatial resolution of 2 m. As shown in Fig. 7(c), the number of endmembers in the image are four, namely, asphalt, grass, tree, and roof (the spectrum is shown in Fig. 7(f), and the CNN structure is shown in Table II), and six, namely asphalt, grass, tree, roof, metal, and dirt (the spectrum is shown in Fig. 7(g), and the CNN structure is shown in Table II). The bands affected by the atmosphere (1–4, 76, 87, 101–111, 136–153, and 198–210) were removed, and the remaining 162 bands were reserved for the subsequent testing of the method.

The unmixing results obtained by the different methods on the Urban dataset are shown in Figs. 10 and 11. Tables V and VI

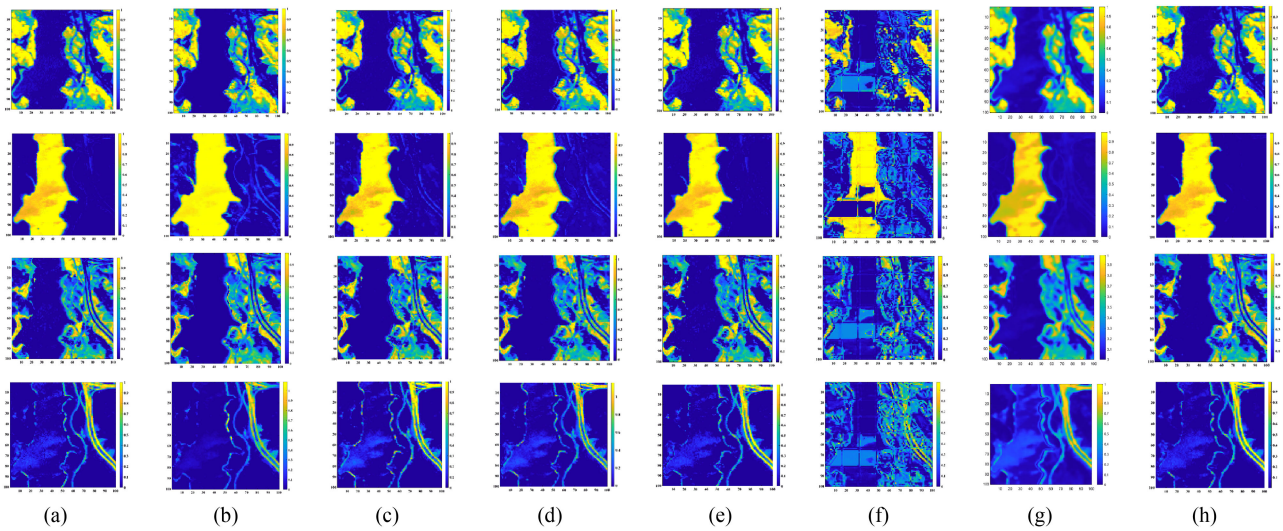

Fig. 9. Abundance maps obtained by all methods for Jasper Ridge dataset. (a) Reference abundances. (b) SUnSAL. (c) CLSUnSAL. (d) SUnSAL-TV. (e) Fuzzy ARTMAP. (f) SPMCNN-ESPCN. (g) 3-D CNN. (h) Proposed CNN.

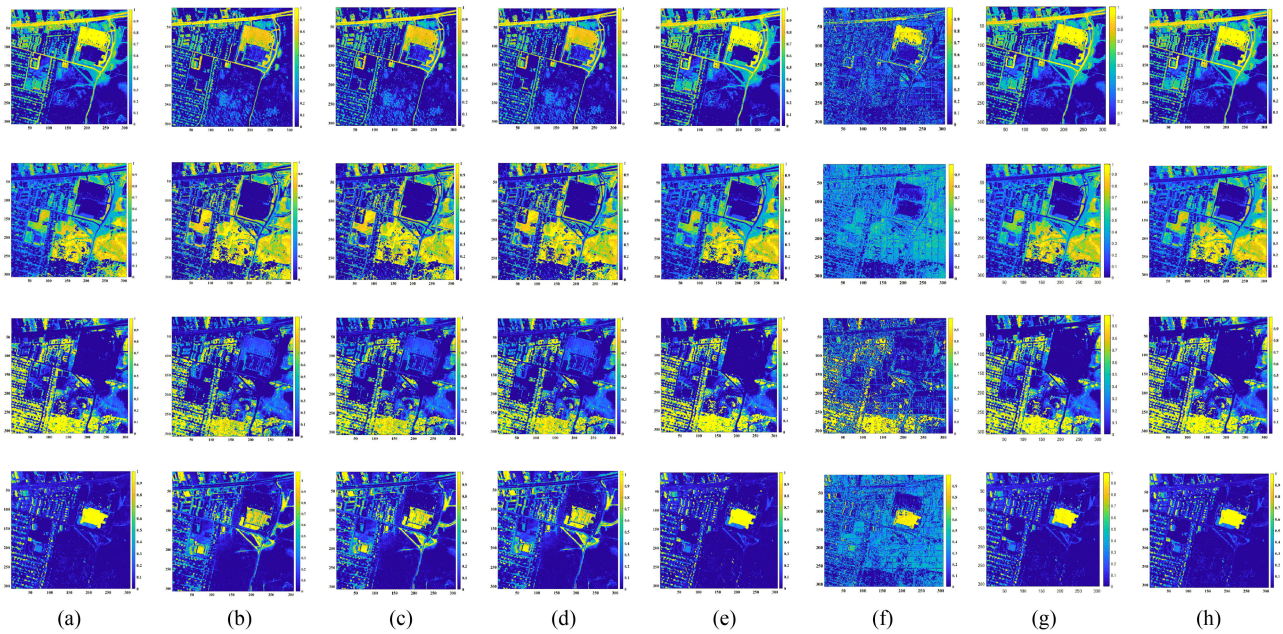

Fig. 10. Abundance maps obtained by all methods for Urban dataset (4 endmembers). (a) Reference abundances. (b) SUnSAL. (c) CLSUnSAL. (d) SUnSAL-TV. (f) SPMCNN-ESPCN. (g) 3-D CNN. (h) Proposed CNN.

show the RMSE results for all methods with these data. Clearly, the RMSE obtained from the proposed CNN method is still the lowest. The Urban dataset has two kinds of endmember numbers. Compared with the case of four endmembers, the classification of the Urban dataset by six endmembers is more detailed, which means that the mixed spectrum involved in each pixel is more complex. However, according to the RMSE in the table, the unmixing error of the proposed CNN method is the lowest in the two endmember cases, so it can be seen that the proposed CNN method can still obtain the best unmixing result with an increasing number of endmember categories, and it has good stability. Compared with the first two real datasets, the Urban dataset is clearly more complex, especially because the decomposition of impervious water surfaces and vegetation, involving many local species with a complex distribution, is very difficult. Therefore, extracting the spectrum of each fine ground object and solving linearly will not only further expand the endmember spectrum library and increase the solution difficulty but also make it more difficult to constrain the abundance sparsity. As mentioned above, the complexity of this dataset affects

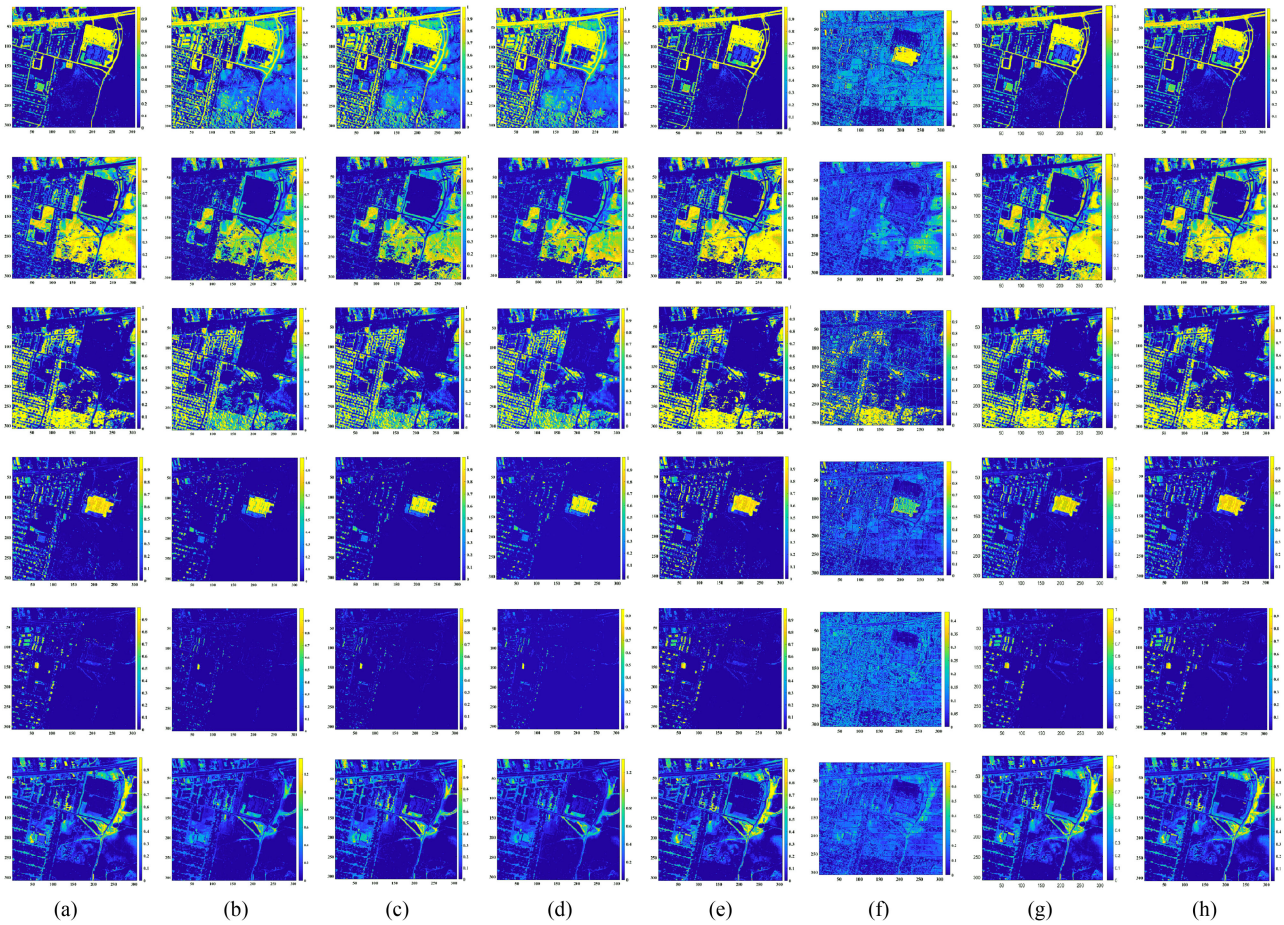

Fig. 11. Abundance maps obtained by all methods for Urban dataset (six endmembers). (a) Reference abundances. (b) SUnSAL. (c) CLSUnSAL. (d) SUnSAL-TV. (f) SPMCNN-ESPCN. (g) 3-D CNN. (h) Proposed CNN.

TABLE V
RMSE ON URBAN DATA (FOUR ENDMEMBERS) (THE BOLDED NUMBER MEANS THE LOWEST RMSE IN EACH ROW)

RMSE	Asphalt	Grass	Tree	Roof	Mean
SUnSAL	0.2440	0.1631	0.1479	0.2825	0.2167
CLSUnSAL	0.2440	0.1631	0.1479	0.2817	0.2164
SUnSAL-TV	0.2433	0.1609	0.1472	0.2818	0.2157
Fuzzy-ARTMAP	0.0366	0.0476	0.0587	0.0508	0.0491
SPMCNN-ESPCN [46]	0.1553	0.1281	0.1453	0.1403	0.1423
3D-CNN [33]	0.0177	0.0238	0.0294	0.0199	0.0227
Proposed-CNN	**0.0125**	**0.0132**	**0.0109**	**0.0123**	**0.0122**

the decomposition accuracy. In practical application, we may face a variety of ground conditions. Under the condition of six endmembers, the accuracy of SPMCNN-ESPCN is improved, but its RMSE is still higher than that of 3-D CNN and the method proposed in this article.

IV. DISCUSSION

Influence of the Number of Endmembers: To compare the complexity of the ground feature coverage in different areas in horizontal and vertical applications, the number of endmembers in the selected simulated dataset and the three real datasets vary within 3–9, and the image size also varies, as does the mixing of ground features in different images. By comparing the five unmixing results obtained from the four scenes, the RMSE of the real datasets obtained by the CNN method is always lower than that obtained by the other methods, which also proves the effectiveness of the extraction of high-dimensional information and the simulation of hyperspectral mixing based on a CNN. Although the proposed CNN does not achieve the best results in the simulated dataset, its accuracy is 0.01–0.02 lower than that of 3-D CNN, so its unmixing results still have some validity. Second, the dataset was synthesized by artificial linear simulation. Although noise was added in the later stage, it could still not fully represent the real data results. However, the NN mainly solves the deep feature mixing problems, so the NN method may complicate simple linear problems and extract too many spectral features, thus making the decomposition results worse than the linear method.

Comparisons Among Different CNN-Based Methods: Similar to 3-D CNN [33] and SPMCNN-ESPCN [46] methods, the network structure in this article is mainly improved based on a CNN. In particular, 3-D CNN is superior to 1-D CNN in extracting spatial features of images. Although the mixed pixel is due to the spatial resolution of the image is not enough to distinguish each type of ground object to a certain extent, the

TABLE VI
RMSE ON URBAN DATA (SIX ENDMEMBERS) (THE BOLDED NUMBER MEANS THE LOWEST RMSE IN EACH ROW)

RMSE	Asphalt	Grass	Tree	Roof	Metal	Dirt	Mean
SUnSAL	0.3326	0.1878	0.1750	0.1443	0.0960	0.1509	0.1955
CLSUnSAL	0.3313	0.1883	0.1756	0.1443	0.0960	0.1506	0.1953
SUnSAL-TV	0.3309	0.1865	0.1749	0.1443	0.0968	0.1507	0.1949
Fuzzy-ARTMAP	0.0714	0.0700	0.0732	0.0444	0.0416	0.0379	0.0619
SPMCNN-ESPCN [46]	0.1114	0.1417	0.1026	0.0795	0.0772	0.0718	0.0941
3D-CNN [33]	0.0322	0.0425	0.0310	0.0256	0.0300	**0.0198**	0.0310
Proposed-CNN	**0.0242**	**0.0146**	**0.0118**	**0.0134**	**0.0169**	0.0225	**0.0178**

final information we get is the mixed spectrum, so the feature analysis of the spectral dimension is the focus of the study of unmixing. However, both 2-D and 3-D CNN focus on the analysis and extraction of spatial features, which is the reason why 2-D and 3-D CNN are not selected for the study of unmixing. Decomposition results based on different datasets also confirm the superiority and adaptability of our proposed CNN method in unmixing. In the previous analysis of different datasets, we also mentioned the blocking influence of SPMCNN-ESPCN method. Just like the network migration problem, the features of the mixed pixels between blocks are different. When some image blocks are randomly selected for network training, the features of some blocks are likely to be lost, resulting in poor decomposition results of the whole image. Although 3-D CNN and the method proposed in this article also require random selection of some data for training, it is more difficult to select adjacent data from random selection points in the whole image compared with random selection of image blocks. After all, there are certain similarities between adjacent data, and it is most important for us to select mixed pixels with different features, so that the network can learn more features.

Comparison of the Linear and NN Methods: The essence of linear unmixing is to solve the $Y = AX + B$ problem, and it is very dependent on the extraction of the pure pixel spectrum. On the simulated dataset, the accuracy of fuzzy ARTMAP is clearly lower than that of the other four methods. Compared with the CNN, the two NNs have essential differences in training and learning. The results also show that the CNN is more effective than fuzzy ARTMAP in learning complex data using training matching. On the Samson and Jasper Ridge datasets, all five methods have achieved good results, and fuzzy ARTMAP and the CNN are superior to the three linear methods. However, compared with the Urban dataset, the linear unmixing method is clearly worse than the other two methods. From the perspective of images, its size is larger, and the ground objects are more complex; that is, the same objects have more seriously different spectra. The mixed pixels in the image cannot be completely solved by simple spectral superposition of pure pixels. However, the fuzzy ARTMAP and CNN unmixing methods use a small amount of spectral data and labels as the training set to train the network and then use the whole spectrum for testing to obtain better unmixing results. With its superior ability to extract high-dimensional features, the CNN method maps features into abundance vectors, and its performance is better than the tracking matching learning between data and tags in fuzzy ARTMAP. Compared with the linear decomposition model, the CNN model and other three NN models are all nonlinear decomposition models. Although the simulated data we use are linear synthesis, the proposed CNN models we propose have achieved good results based on the joint experiments of real data sets. Therefore, although we cannot simulate nonlinear mixing mode in data, CNN shows superior performance in processing unknown noise in real mixed pixels.

Influence of the Network Structure: The CNN is widely used in the fields of scene classification, recognition, and monitoring, and its ability to extract spatial and spectral features is utilized to gradually compress high-dimensional spectral data and finally map it into abundance vectors. Compared with the CNN, fuzzy ARTMAP has no process of feature extraction and data compression and no influence of the activation function. Therefore, the two methods are also completely different in spectral feature learning, allowing the advantages of the CNN in feature extraction to be compared with those of fuzzy ARTMAP. Table II shows the specific network structure of the four groups of data. Referring to the literature[47], [48], the output units of the hidden layer were set as multiples of the number of endmember categories according to the different datasets, and the output units of the hidden layer were reduced successively. In this way, redundant information can be removed while extracting spectral features so that the network dimension can be gradually reduced and important features are retained. Finally, features were mapped to abundance vectors through the full connection layer and the softmax activation function. Different from the linear methods, the NN model relies on the network structure to extract the spectral features of the pixel to achieve the classification of the endmember, so it is more difficult to understand in the physical sense. Compared with fuzzy ARTMAP, the CNN's convolutional layer, pooling layer, and full connection layer construct the network's low hidden layer and high hidden layer, which can extract the network features of higher level.

V. CONCLUSION

Linear unmixing can be regarded as blind source signal separation or a nonnegative matrix decomposition problem, but it often results in a large number of solutions inconsistent with the actual situation. Therefore, the sparse constraint, some-to-one constraint, and the nonnegative constraint of abundance are generally added to the linear model, but the nature of linear mixing of the spectrum is not changed. Fuzzy ARTMAP and CNNs (as well as many other deep learning networks) learn the mapping between input and output and apply the learned mapping to the test data. Compared with fuzzy ARTMAP, the feature extraction ability of a CNN is more flexible. The CNN based on 1-D pixels

presented in this article removes the redundant information in spectral signatures and adds sparse constraints to the abundance vector through a loss function combined with ℓ_1 regularization in linear sparse unmixing, making the unmixing result more consistent with the actual situation.

REFERENCES

[1] Z. Lin, Y. Chen, X. Zhao, and G. Wang, "Spectral-Spatial classification of hyperspectral image using autoencoders," *Proc. 9th Int. Conf. Inf. Commun. Signal Process.*, 2013, pp. 1–5.

[2] L. Zhang, M. Lan, J. Zhang, and D. Tao, "Stagewise unsupervised domain adaptation with adversarial self-training for road segmentation of remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2021.3104032.

[3] X. Wang, S. Liu, P. Du, H. Liang, J. Xia, and Y. Li, "Object-based change detection in urban areas from high spatial resolution images based on multiple features and ensemble learning," *Remote Sens.*, vol. 10, no. 2, Feb. 2018, Art. no. 276.

[4] D. Zhu, T. Chen, Z. Wang, and R. Niu, "Detecting ecological spatial-temporal changes by remote sensing ecological index with local adaptability," *J. Environ. Manage.*, vol. 299, 2021, Art. no. 113655.

[5] Y. X. Lin *et al.*, "Hyperspectral image enhancement by two dimensional quaternion valued singular spectrum analysis for object recognition," *Remote Sens.*, vol. 13, no. 3, Feb. 2021, Art. no. 405.

[6] C. W. Calvin, E. F. Littlefield, and C. Kratt, "Remote sensing of geothermal-related minerals for resource exploration in Nevada," *Geothermics*, vol. 53, pp. 517–526, Jan. 2015.

[7] Y. Zhong, R. Feng, and L. Zhang, "Non-local sparse unmixing for hyperspectral remote sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 1889–1909, Jun. 2014.

[8] M. D. Iordache, J. M. Biousas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2011.

[9] M. Wang, M. Zhao, J. Chen, and S. Rahardja, "Nonlinear unmixing of hyperspectral data via deep autoencoder networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 3, pp. 1467–1471, Sep. 2019.

[10] Y. Qu and H. R. Qi, "uDAS: An untied denoising autoencoder with sparsity for spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1698–1712, Mar. 2019.

[11] B. Palsson, M. O. Ulfarsson, and J. R. Sveinsson, "Convolutional autoencoder for spectral-spatial hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 535–549, Jan. 2020.

[12] J. M. P. Nascimento and J. M. B. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.

[13] C. I. Chang, C. H. Wen, and C. C. Wu, "Relationship exploration among PPI, ATGP and VCA via theoretical analysis," *Int. J. Comput. Sci. Eng.*, vol. 8, no. 4, pp. 361–367, Oct. 2013.

[14] L. Zhao, J. Zheng, and X. Li, and L. Wang, "Kernel simplex growing algorithm for hyperspectral endmember extraction," *J. Appl. Remote Sens.*, vol. 8, no. 1, Jul. 2014, Art. no. 083594.

[15] L. Zhang, L. Song, B. Du, and Y. Zhang, "Nonlocal low-rank tensor completion for visual data," *IEEE Trans. Cybern.*, vol. 51, no. 2, pp. 673–685, Feb. 2021.

[16] M. D. Iordache, J. M. Biousas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4484–4502, Nov. 2012.

[17] M. D. Iordache, J. M. Biousas-Dias, and A. Plaza, "Collaborative sparse regression for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 341–354, Jan. 2014.

[18] R. Y. Feng, L. Z. Wang, and Y. F. Zhong, "Joint local block grouping with noise-adjusted principal component analysis for hyperspectral remote-sensing imagery sparse unmixing," *Remote Sens.*, vol. 11, no. 10, May 2020, Art. no. 1223.

[19] R. L. Weng, "Kernel sparse representation for hyperspectral unmixing based on high mutual coherence spectral library," *Int. J. Remote Sens.*, vol. 41, no. 4, pp. 1286–1301, Feb. 2020.

[20] Y. B. Deng, R. R. Chen, and C. S Wu, "Examining the deep belief network for subpixel unmixing with medium spatial resolution multispectral imagery in urban environments," *Remote Sens.*, vol. 11, no. 13, Jul. 2019, Art. no. 1566.

[21] Y. C. Su, J. Li, A. Plaza, A. Marinoni, P. Gamba, and S. Chakravortty, "DAEN: Deep autoencoder networks for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4309–4321, Jul. 2019.

[22] O. Savas, K. Berk, and A. G. Bozdagi, "EndNet: Sparse autoencoder network for endmember extraction and hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 482–496, Jan. 2019.

[23] X. Gao, T. Chen, R. Niu, and A. J. Plaza, "Recognition and mapping of landslide using a fully convolutional DenseNet and influencing factors," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 7881–7894, 2021.

[24] T. Liu, T. Chen, R. Niu, and A. J. Plaza, "Landslide detection mapping employing CNN, ResNet, and DenseNet in the Three Gorges Reservoir, China," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 11417–11428, 2021, doi: 10.1109/JSTARS.2021.3117975.

[25] K. Wu, L. F. Wei, and X. M. Wang, "Adaptive pixel unmixing based on a fuzzy ARTMAP neural network with selective endmembers," *Soft Comput.*, vol. 20, no. 12, pp. 4723–4732, Dec. 2016.

[26] A. M. Filippi, C. Brannstrom, I. Dobreva, D. M. Cairns, and D. Kim, "Unsupervised fuzzy ARTMAP classification of hyperspectral hyperion data for Savanna and agriculture discrimination in the Brazilian Cerrado," *GISci. Remote Sens.*, vol. 46, Jan. 2009, Art. no. 1.

[27] G. F. Li, "Based on the improved BP neural network of remote sensing image decomposition of mixed pixels," *Geospatial Inf.*, vol. 7, no. 3, pp. 65–67, 2009.

[28] M. Zhao, L. Yan, and J. Chen, "LSTM-DNN based autoencoder network for nonlinear hyperspectral image unmixing," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 2, pp. 295–309, Feb. 2021.

[29] P. Y. Jia, M. Zhang, and Y. Shen, "Deep spectral unmixing framework via 3D denoising convolutional autoencoder," *IET Image Process.*, vol. 15, no. 7, pp. 1399–1409, May 2021.

[30] B. Palsson, J. Sigurdsson, J. R. Sveinsson, and M. O. Ulfarsson, "Hyperspectral unmixing using a neural network autoencoder," *IEEE Access*, vol. 6, pp. 25646–25656, 2018.

[31] L. Qi, J. Li, and Y. Wang, "Deep spectral convolution network for hyperspectral image unmixing with spectral library," *Signal Process.*, vol. 176, Nov. 2020, Art. no. 107672.

[32] S. Ozkan and G. B. Akar, "Improved deep spectral convolution network for hyperspectral unmixing with multinomial mixture kernel and endmember uncertainty," *arXiv:1808.01104*, 2018.

[33] X. R. Zhang, Y. J. Sun, J. Y. Zhang, P. Wu, and L. Jiao, "Hyperspectral unmixing via deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 11, pp. 1755–1759, Nov. 2018.

[34] S. Moussaoui, C. Carteret, D. Brie, and A. Mohammad-Djafari, "Bayesian analysis of spectral mixture data using Markov chain Monte Carlo methods," *Chemometrics Intell. Lab. Syst.*, vol. 81, no. 2, pp. 137–148, Apr. 2006.

[35] H. C. Li, S. Liu, and X. R. Feng, "Sparsity-constrained coupled nonnegative matrix-tensor factorization for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5061–5073, 2020.

[36] C. Li *et al.*, "Sparse unmixing of hyperspectral data with bandwise model," *Inf. Sci.*, vol. 512, pp. 1424–1441, Feb. 2020.

[37] G. A. Carpenter and S. Grossberg, "The ART of adaptive pattern recognition by a self-organizing neural network," *Computer*, vol. 21, no. 3, pp. 77–88, Mar. 1988.

[38] P. Farhad, L. C. Peng, and H. Qi, "A reinforced fuzzy ARTMAP model for data classification," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 7, pp. 1643–1655, Jul. 2019.

[39] I. Ahmed, M. Ahmad, and F. A. Khan, "Comparison of deep-learning-based segmentation models: Using top view person images," *IEEE Access*, vol. 8, pp. 136361–136373, 2020.

[40] L. P. Zhang and B. Du, "Deep learning for remote sensing data a technical tutorial on the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, Jun. 2016.

[41] I. Priyadarshini and V. Puri, "A convolutional neural network (CNN) based ensemble model for exoplanet detection," *Earth Sci. Informat.*, vol. 14, pp. 735–747, Feb. 2021.

[42] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature: Int. Weekly J. Sci.*, vol. 521, no. 7553, pp. 436–444, May 2015.

[43] A. Shakya, M. Biswas, and M. Pal, "Parametric study of convolutional neural network based remote sensing image classification," *Int. J. Remote Sens.*, vol. 42, no. 7, pp. 2663–2685, Apr. 2021.

[44] J. Zhang, J. Z. Liu, and Z. Z. Wang, "Assessing deep convolutional neural networks and assisted machine perception for urban mapping," *Remote Sens.*, vol. 13, no. 8, Apr. 2021, Art. no. 1523.

[45] J. T. Yao, T. Mitran, X. B. Kong, R. Lal, Q. Chu, and M. Shaukat, "Landuse and land cover identification and disaggregating socio-economic data with convolutional neural network," *Geocarto Int.*, vol. 35, no. 10, pp. 1109–1123, Jul. 2020.
[46] D. He, Y. F. Zhong, and X. Y. Wang, "Deep convolutional neural network framework for subpixel mapping," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9518–9539, Nov. 2020.
[47] F. Palsson, J. Sigurdsson, J. R. Sveinsson, and M. O. Ulfarsson, "Neural network hyperspectral unmixing with spectral information divergence objective," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2017, pp. 755–758.
[48] F. Khajehrayeni and H. Ghassemian, "A supervised nonlinear spectral unmixing method by means of neural networks," in *Proc. 5th Iranian Conf. Signal Process. Intell. Syst.*, Dec. 2019, pp. 1–6.

Lulu Wan received the B.S. degree in geographic information science from Huazhong Agricultural University, Wuhan, China, in 2018. She is currently working toward the M.S. degree in earth exploration and information technology with China University of Geosciences, Wuhan, China.

Her current main research interests include hyperspectral unmixing based on convolutional neural networks.

Tao Chen (Senior Member, IEEE) received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan China, in 2008.

From 2015 to 2016, he was a Visiting Scholar with the University of New South Wales, Sydney, NSW, Australia. He is currently an Associate Professor with the Institute of Geophysics and Geomatics, China University of Geosciences, Wuhan, China. He has authored or coauthored more than 50 scientific papers including IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, *Journal of Environmental Management*, *Remote Sensing*, *Environmental Earth Sciences*, and *Environmental Science and Pollution Research*, and guest edited two journal special issues. His research interests include image processing, machine learning, and geological remote sensing.

Antonio Plaza (Fellow, IEEE) received the M.Sc. and Ph.D. degrees in computer engineering from the Department of Technology of Computers and Communications, University of Extremadura, Badajoz, Spain, in 1999 and 2002, respectively.

He is currently a Full Professor and the Head of the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, University of Extremadura, Badajoz, Spain. He has authored more than 600 publications and guest edited ten journal special issues. He has reviewed more than 500 manuscripts for over 50 different journals.

Dr. Plaza was the Editor-in-Chief for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING from 2013 to 2017. From 2018 to 2020, he was included in the Highly Cited Researchers List (Clarivate Analytics).

Haojie Cai received the B.S. degree in geoinformatics in 2018 from the China University of Geoscience, Wuhan, China, where he is currently working toward the M.S. degree in earth exploration and information technology.

His current research interests include landslide detection and remote sensing applications.