Hyperspectral Unmixing for Additive Nonlinear Models With a 3-D-CNN Autoencoder Network

Min Zhao, *Graduate Student Member, IEEE*, Mou Wang, *Student Member, IEEE*, Jie Chen, *Senior Member, IEEE*, and Susanto Rahardja, *Fellow, IEEE*

***Abstract*— Spectral unmixing is an important task in hyperspectral image processing for separating the mixed spectral data pertaining to various materials observed aiming at analyzing the material components in observed pixels. Recently, nonlinear spectral unmixing has received particular attention in hyperspectral image processing, as there are many situations in which the linear mixture model may not be appropriate and could be advantageously replaced by a nonlinear one. Existing nonlinear unmixing approaches are often based on specific assumptions on the nonlinearity and can be less effective when used for scenes with unknown nonlinearity. This article presents an unsupervised nonlinear spectral unmixing method that addresses a general model that consists of a linear mixture part and an additive nonlinear mixture part. The structure of a deep autoencoder network, which has a clear physical interpretation, is specifically designed to achieve this purpose. Moreover, a convolutional neural network (CNN) is used to capture the spectral-spatial priors from hyperspectral data. Extensive experiments with synthetic and real data illustrate the generality and effectiveness of this scheme compared with state-of-the-art methods.**

***Index Terms*— 3-D-convolutional neural network (CNN), autoencoder network, hyperspectral imaging, nonlinear spectral unmixing.**

I. Introduction

HYPERSPECTRAL imaging is a continuously growing field of study that has received considerable attention over the past decade. Hyperspectral data provide high spectral resolution over a wide spectral range that typically extends from the infrared spectrum through the visible spectrum. This rich spectral information facilitates the discrimination of different materials in the observed scene. As a result, hyperspectral imaging has been widely adopted for a wide range of applications, such as land use analysis, pollution monitoring, wide-area reconnaissance, and field surveillance [2].

However, the spectral content of individual pixels in hyperspectral images often represent a mixture of several materials from the imaged scene due to multiple factors, such as the low spatial resolution of hyperspectral imaging devices, the diversity of materials in the imaged scene, and multiple reflections of photons from several objects. Therefore, separating spectra of individual pixels into a set of spectral signatures (endmembers), and determining the fraction abundances associated with each endmember is an essential task required for analyzing remotely sensed data. This process is denoted as spectral unmixing or mixed pixel decomposition [3]. Spectral unmixing methods have been developed for this purpose based on both linear and nonlinear mixture models.

Among the presently available spectral mixture models, the linear mixture model (LMM) is the most widely used. In LMM, the incident light is assumed to be reflected by each component present in the scene only once prior to collection by the camera sensor, and the observed spectrum is thus a linear combination of the endmembers [3]. Many conventional unmixing methods based on LMM have been proposed. In [4], the total variation regularization is imposed to add spatial information of the hyperspectral image and [5] considered the endmember variability problem. While the LMM is simple and physically interpretable, numerous complex conditions arise where the incoming light may undergo complex interactions among the individual materials in the scene, resulting in higher-order photon interactions that introduce nonlinear effects in the mixed spectra. Consequently, the analysis of data collected under these conditions requires nonlinear unmixing (NAE) methods [6]. A considerable number of studies have recently focused on addressing NAE problems. For example, bilinear models [7] have been developed to address conditions of second-order scattering interactions that may occur on complex vegetated surfaces, by adding extra bilinear interaction terms to the linearly composited spectrum. Such models include the Fan model [8] and the generalized bilinear model (GBM) [9]. The polynomial post-nonlinear mixture model (PNMM) applies a polynomial function to the linearly mixed data to approximate the nonlinearity of photon interactions occurring in an imaged scene [10]. A bidirectional reflectance model has been developed to describe the photon interactions of intimately mixed particles based on the fundamental principles of radiative transfer theory. This model is generally referred to as the intimate mixture model or Hapke model [11]. The multimixture pixel (MMP) model further extended the intimate mixture model by integrating it with the LMM model [12]. The above cases have been generalized by considering a linear-mixture/nonlinear-fluctuation (K-Hype) model, where the nonlinear fluctuation was

Manuscript received February 6, 2021; revised May 2, 2021 and June 9, 2021; accepted June 9, 2021. Date of publication August 2, 2021; date of current version January 17, 2022. The work of Jie Chen was supported in part by the National Key Research and Development Program of China under Grant 2018AAA0102200 and in part by 111 Project under Grant B18041. The work of Min Zhao and Mou Wang were supported in part by the Innovation Foundation for Doctor Dissertation of Northwestern Polytechnical University. *(Corresponding author: Jie Chen.)*

The authors are with the School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: minzhao@mail.nwpu.edu.cn; mangmou21@mail.nwpu.edu.cn; dr.jie.chen@ieee.org; susantorahardja@ieee.org).

Digital Object Identifier 10.1109/TGRS.2021.3098745

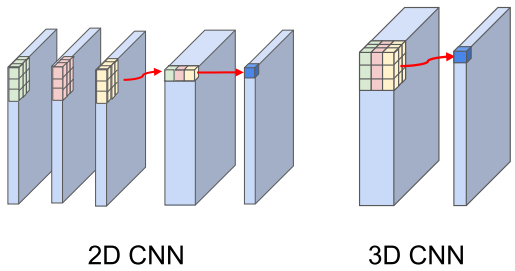


Fig. 1. Difference between 2-D CNN and 3-D CNN. 2-D CNN captures the spectral-spatial information, respectively, whereas 3-D CNN captures the spectral and spatial information simultaneously.

described by a function defined in a reproducing kernel Hilbert space (RKHS) [13]. Further extensions of this model have also been proposed with spatial regularization [14] and neighborhood-dependent contributions [15]. The multilinear mixing model (MLM) considers an infinite number of photon interactions by introducing a probability of photon undergoing further interactions [16]. The work [17] uses a graph-based model to describe the multiple photon interactions. However, most of the above models rely heavily on specific assumptions regarding the inherent nonlinearity of the spectral unmixing, and they are therefore not well suited to scenes with unknown nonlinearity characteristics. In addition, while the K-Hype model based on the RKHS presented above and other kernel-based algorithms provide flexible nonlinear modeling, the selection of appropriate kernels and kernel parameters has been demonstrated to be a nontrivial issue that restricts the application of these approaches. Finally, all of these algorithms assume that the endmembers are known prior, and therefore focus strictly on evaluating the abundance fractions.

In recent years, deep learning has demonstrated its superior performance in addressing various nonlinear problems compared to classical methods. Researchers have also investigated the use of deep neural networks (DNNs) in hyperspectral image analysis. Particular attention has been focused on the hyperspectral image classification problem [18]–[21]. However, despite the recognized potential of neural networks for solving inverse problems, only a handful of studies have applied neural networks for addressing the spectral unmixing problem. Among these, classifier models have been applied to spectral unmixing [22], [23], but this approach requires a training set with known abundance ground-truths or endmembers, which must often be generated by theoretical models. In addition, autoencoder networks have also been applied to the blind spectral unmixing problem. An autoencoder is a network that learns to compress an input into a shortcode which can be uncompressed into something that is close to the original input. Internally, it has a hidden layer that describes that short dimensional code used to represent the input data for reconstructing the output data. This is ideally suited for conducting spectral unmixing because this process can also be considered as finding a low dimensional representation (abundance fractions) of hyperspectral data. For example, approaches employing autoencoder networks have exhibited good performance in determining both endmembers and abundance fractions [24]–[31]. Recently, some convolutional neural network (CNN)-based frameworks are also used for spectral unmixing task due to their ability of extracting spatial structures from hyperspectral images [32], [33].

TABLE I
COMPARISON OF THE PROPOSED METHOD AND OTHER EXISTING AUTOENCODER-BASED UNMIXING ALGORITHMS

Reference number	Supervised/unsupervised	Linear/nonlinear	DNN/CNN
[23]	supervised	nonlinear	3D-CNN
[24]	unsupervised	linear	DNN
[25]	unsupervised	linear	DNN
[26]	unsupervised	linear	DNN
[27]	unsupervised	linear	DNN
[28]	unsupervised	linear	DNN
[29]	unsupervised	linear	DNN
[30]	unsupervised	linear	DNN
[32]	unsupervised	linear	2D-CNN
[33]	supervised	linear	3D-CNN
[34]	unsupervised	nonlinear (post-nonlinear)	DNN
Proposed	unsupervised	additive general nonlinearity	3D-CNN

However, these approaches are specifically designed to preprocess the input data or address the linear unmixing problem, and therefore fail to make use of the superior potential of neural networks for addressing nonlinear problems, while linear unmixing is readily addressed using classical methods. The work [34] considered a post-nonlinear spectral mixture, where the post-nonlinearity was modeled by the decoder part of the autoencoder. In there, pretraining and learning rate adjustment techniques were required to ensure the effectiveness of the decoder, and the nonlinear model represented by the decoder was not sufficiently general to cover multiple nonlinear cases. However, this work processes pixels independently and ignores the spatial information of hyperspectral image.

Our work addresses the drawbacks in previous works by reexamining the nonlinear mixture models and restrictions of existing unmixing schemes. A 3-D-CNN is utilized to jointly learn the spectral-spatial structures. The superiority of 3-D-CNN is shown in Fig. 1. Accordingly, this article presents a new CNN-based autoencoder network structure for blind nonlinear spectral unmixing. The difference between the proposed method and related works are highlighted in Table I. The highlights of this work are summarized as follows.

1) A general spectral mixture model that consists of a linear mixture component and an additive nonlinear mixture component is proposed. The significance of an endmember in the nonlinear mixture component is weighted according to its associated abundance fraction. A DNN is proposed to represent this nonlinear part and generalizes the existing related models.
2) The form of the inherent nonlinearity of our nonlinear mixture component is learned from the data itself, rather than relying on an assumed form. The structure of the decoder is designed with particular care so that the nonlinear interactions are imposed on endmembers weighted by abundances, which has a clear physical interpretation and covers several existing artificial models. Endmembers and abundance fractions are extracted from the outputs and weights of the particular layers of

TABLE II
RELATING TYPICAL NONLINEAR MODELS WITH THE GENERIC FORM (8) (NOISE VECTOR $\mathbf{n}$ IS OMITTED FOR SAVING SPACE)

	Model expression	Form of Ψ	Note
Bilinear model	$\mathbf{x} = \mathbf{Ma} + \sum_{i=1}^{R}\sum_{j=i+1}^{R} a_i\mathbf{m}_i \odot a_j\mathbf{m}_j$	$\Psi = \sum_{i=1}^{R}\sum_{j=i+1}^{R} a_i\mathbf{m}_i \odot a_j\mathbf{m}_j$	$\odot$ denotes the element-wise product
Post-nonlinear model	$\mathbf{x} = \mathbf{Ma} + \mathbf{Ma} \odot \mathbf{Ma}$	$\Psi = \mathbf{Ma} \odot \mathbf{Ma}$	Note $\mathbf{Ma} = \mathbf{m}_1 a_1 + \cdots + \mathbf{m}_R a_R$
K-Hype model	$x_i = \mathbf{m}_i\mathbf{a} + \psi(\mathbf{m}_{\lambda_i})$	$[\Psi]_i = \sum_{i=1}^{B} \beta_i\, \kappa(\mathbf{m}_{\lambda_i}, \mathbf{m}_{\lambda_j})$	ψ is in a RKHS with kernel κ β_i are coefficients to be determined $\mathbf{a}$ is ignored in Ψ
Multilinear mixing model	$\mathbf{x} = \mathbf{Ma} + p\mathbf{Ma}(\mathbf{Ma}-1)/(1-p\mathbf{Ma})$	$\Psi = p\mathbf{Ma}(\mathbf{Ma}-1)/(1-p\mathbf{Ma})$	p is the probability of interactions

the network. Extra regularizations are also imposed to enhance the unmixing performance.

3) Our encoder is a 3-D-CNN-based architecture. 3-D convolutional filters are used to capture the spectral-spatial structures of neighbor pixels simultaneously. Unlike other NAE methods which mainly rely on spectral information in the processing, our proposed method jointly analyzes spectral–spatial priors of a hyperspectral image.

The remainder of this article is organized as follows: Section II presents the formulation of the nonlinear mixture model. Section III presents the design of the proposed autoencoder scheme for unmixing. Section IV validates the proposed method with experiments using synthetic and real data. Section V concludes the work and provides the perspective of the future work.

II. PROBLEM FORMULATION

Notation: Normal font x and X denote scalars. Boldface small letters $\mathbf{x}$ denote vectors. All vectors are column vectors. Boldface capital letters $\mathbf{X}$ denote matrices. Considering an observed pixel data $\mathbf{x} \in \mathbb{R}^B$ with B denoting the number of spectral bands, and $\mathbf{M} = [\mathbf{m}_1, \ldots, \mathbf{m}_R]$ denotes the $(B \times R)$ endmember matrix with endmembers $\mathbf{m}_i$, R represents the number of endmembers. $\mathbf{a} = [a_1, a_2, \ldots, a_R]^\top$ is the abundance vector associated with a pixel. The operator blkdiag$\{\cdots\}$ forms a matrix of size $BR \times R$ using vectors $\{\mathbf{y}_i\}_{i=1}^{R} \in \mathbb{R}^B$ such that

$$\text{blkdiag}\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_R\} = \begin{bmatrix} \mathbf{y}_1 & \mathbf{0}_B & \cdots & \mathbf{0}_B \\ \mathbf{0}_B & \mathbf{y}_2 & \cdots & \mathbf{0}_B \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0}_B & \mathbf{0}_B & \cdots & \mathbf{y}_R \end{bmatrix} \in \mathbb{R}^{BR\times R} \quad (1)$$

with $\mathbf{0}_B$ denoting all zero vectors of length B. Considering using such a matrix $\mathbf{Y} = \text{blkdiag}\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_R\}$ as the weight matrix of a layer of a DNN, regular matrix product maps the input $\mathbf{h} = [h_1, \ldots, h_R]^\top$ to the output of the form

$$\mathbf{Yh} = \left[h_1\mathbf{y}_1^\top, h_2\mathbf{y}_2^\top, \ldots, h_R\mathbf{y}_R^\top\right]^\top. \quad (2)$$

The usefulness of such an operation will be clear when we relate (23) and (26) to (9).

The operator col$\{\mathbf{y}_1, \ldots, \mathbf{y}_N\}$ stacks its vector arguments $\{\mathbf{y}_i\}_{i=1}^{N}$ on the top of each other to generate a connected vector given by

$$\text{col}\{\mathbf{y}_1, \ldots, \mathbf{y}_N\} = [\mathbf{y}_1^\top, \ldots, \mathbf{y}_N^\top]^\top = \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_N \end{pmatrix}. \quad (3)$$

We first consider the LMM where each observed pixel is assumed to be a linear combination of the endmembers weighted by their associated abundances

$$\mathbf{x} = \mathbf{Ma} + \mathbf{n} \quad (4)$$

where $\mathbf{a}$ denotes the corresponding abundance vector, and $\mathbf{n} \in \mathbb{R}^B$ is an additive noise vector. As the abundances represent relative fractions of each material, they are required to satisfy the abundance nonnegative constraint (ANC), (5), and abundance sum-to-one constraint (ASC), (6), that are

$$\forall i : a_i \geq 0 \quad (5)$$

$$\sum_{i=1}^{R} a_i = 1. \quad (6)$$

In this work, we consider the following general mixing mechanism:

$$\mathbf{x} = \mathbf{Ma} + \Psi(\mathbf{M}, \mathbf{a}) + \mathbf{n} \quad (7)$$

which consists of a linear mixture of endmembers $\mathbf{M}$ with abundance fractions $\mathbf{a}$, and a nonlinear fluctuation Ψ that defines the interactions of $\mathbf{M}$ parameterized by $\mathbf{a}$. Several existing typical nonlinear models are summarized in Table II. We revised the mixture model defined in (7) to provide a more tractable form as follows:

$$\begin{aligned} \mathbf{x} &= \mathbf{Ma} + \Psi(a_1\mathbf{m}_1, a_2\mathbf{m}_2, \ldots, a_R\mathbf{m}_R) + \mathbf{n} \\ &= \mathbf{Ma} + \Psi(\mathbf{M}\,\text{diag}(\mathbf{a})) + \mathbf{n} \end{aligned} \quad (8)$$

where Ψ represents the nonlinear interaction between the endmembers $\mathbf{M}$ weighted by associated abundance $\mathbf{a}$. It is clear that (8) is a general model that yields the models defined in Table II under different choices of Ψ. We refer to this model as a *gerenalized linear-mixture/nonlinear-fluctuation model*. This form suggests that the nonlinear interactions of material signatures are in proportion to the abundance fractions of each material. This is reasonable, because, for instance, a material with a negligible abundance will have limited contribution to either the linear component, or the nonlinear component of $\mathbf{x}$. Several existing nonlinear models can be considered as specific cases of (8) under different definitions of Ψ. Typical nonlinear mixing models and the relations between these algorithms and (8) are summarized in Table II. With the exception of K-Hype, these algorithms are designed manually to capture the assumed nonlinearities. The linear-mixture/nonlinear-fluctuation model used by the K-Hype algorithm is relatively more general, and has some similarities with (8). However, in addition to the nontrivial issue associated with the selections of kernels and

kernel parameters discussed above, this model suffers from the use of a nonlinear fluctuation function that is independent of the abundance fractions. Hence, the endmembers contribute equivalently to the nonlinear component of the observed spectrum.

The present model clearly addresses this restriction by explicitly including the abundance fractions of the endmembers. In addition, the restriction associated with the selections of kernels and kernel parameters is addressed in the proposed approach by not assigning Ψ in (8) with any specific form. Instead, we devise a method to learn it from the data itself via an autoencoder network.

In order to facilitate to present the structure of the autoencoder network, we write (7) in the following equivalent form:

$$\mathbf{x} = \mathcal{T}(\mathbf{M}_D\,\mathbf{a}) + \Psi(\mathbf{M}_D\,\mathbf{a}) + \mathbf{n} \tag{9}$$

where

$$\mathbf{M}_D = \text{blkdiag}\{\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_R\} \in \mathbb{R}^{BR\times R} \tag{10}$$

and $\mathcal{T} : \mathbb{R}^{BR} \mapsto \mathbb{R}^{B}$ is a step-wise summation operator, i.e., for a given vector $\mathbf{y} \in \mathbb{R}^{BR}$

$$\mathcal{T}(\mathbf{y}) = \left(\sum_{i=1}^{R} y_{B\times(i-1)+1}, \ldots, \sum_{i=1}^{R} y_{B\times(i-1)+B}\right)^{\top}. \tag{11}$$

With the above notation, we have

$$\mathbf{M}_D\mathbf{a} = \text{col}\{a_1\mathbf{m}_1, a_2\mathbf{m}_2, \ldots, a_R\mathbf{m}_R\} \tag{12}$$

and

$$\mathcal{T}(\mathbf{M}_D\mathbf{a}) = \mathbf{M}\mathbf{a}. \tag{13}$$

Then the linear component and the nonlinear component share the same input $\mathbf{M}_D\mathbf{a}$. Our previous study on the above model can be found in [1].

III. Proposed Approach

In this section, we present a thorough presentation of the proposed method that solves the NAE problem using deep autoencoder networks.

A. General Structure

The structure of an autoencoder network consists of two parts, namely an encoder and a decoder. Encoder f_{E} compresses the input $\mathbf{x}$ into a low dimensional representation $\mathbf{h} \in \mathbb{R}^R$, that is

$$\mathbf{h} = f_{\text{E}}(\mathbf{x}) \tag{14}$$

with $f_E : \mathbb{R}^{B\times 1} \to \mathbb{R}^{R\times 1}$. Recall that B represents the number of bands and R denotes the number of endmembers. Note that R is assumed priorly known in our work, and it is estimated by using HySime [35] method. As the hyperspectral data is assumed to be low rank, we suppose $R < B$ in our work. Decoder f_D uncompresses the hidden representation vector $\mathbf{h}$ to reconstruct the original input data, that is

$$\hat{\mathbf{x}} = f_D(\mathbf{h}) \tag{15}$$

with $f_D : \mathbb{R}^{R\times 1} \to \mathbb{R}^{B\times 1}$. The network trains the parameters and representations by minimizing the average reconstruction error (RE) between the input $\mathbf{x}$ and its reconstructed counterpart $\hat{\mathbf{x}}_i = f_D(f_{\text{E}}(\mathbf{x}_i))$ given by

$$\mathcal{L}(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{N}\sum_{i=1}^{N} \left\|\hat{\mathbf{x}}_i - \mathbf{x}_i\right\|^2 \tag{16}$$

where N is the number of pixels. With the output of the encoder $\mathbf{h} \in \mathbb{R}^R$, in this work the decoder is designed to reconstruct the input $\mathbf{x}$ with the following specific structure:

$$\hat{\mathbf{x}} = \mathcal{T}(\mathbf{V}^{(1)}\,\mathbf{h}) + \Phi(\mathbf{V}^{(1)}\,\mathbf{h}) \tag{17}$$

where $\mathbf{V}^{(1)}$ are weights of the first layer of the decoder, as to be defined in (22), Φ is the nonlinear function constructed by the nonlinear part of our decoder, and it is expected that after the learning process, Φ mimics the generative model Ψ. Comparing this structure to (9), the decoder mimics the output in accordance to this model. Therefore, after the network parameters are learned with data, blind unmixing of the same input data can be conducted by

$$\text{Abundance estimation} : \mathbf{h} \Rightarrow \hat{\mathbf{a}} \tag{18}$$

$$\text{Endmember extraction} : \mathbf{V}^{(1)} \Rightarrow \widehat{\mathbf{M}}_D. \tag{19}$$

In order to utilize the spectral-spatial priors and achieve better unmixing performance, we divide the hyperspectral cube into N overlapping 3-D-patches with the size of $n \times n \times B$, where n denotes the spatial window size. The central pixel is regarded as the target pixel to be decomposed. Then these patches are entered into our unmixing framework. Both encoder and decoder can either be shallow or deep, but generally, it is believed that deep networks possess a superior modeling capability. The schema of the proposed autoencoder network is illustrated in Fig. 2(a) in order that readers can better understand the proposed structure. We elaborate the design of encoder and decoder in Sections III-B and III-C.

B. Encoder

In this work, a 3-D-CNN-based network is designed as the encoder with a specific structure presented in the left part of Fig. 2(a) (marked with pink color). First, we use two 3-D convolutional layers (Conv) with the spatial size of 3-D convolutional kernels set to 3×3 to fully capture the spectral-spatial features, and the spatial size of each patch is decreased from $n \times n$ to 1×1. Then three 3-D convolutional layers with kernel size of $1 \times 1 \times 8$, $1 \times 1 \times 8$ and $1 \times 1 \times 6$ are utilized to further capture the spectral-spatial features. The number of kernels of the 3-D convolutional layers gradually narrows down, and the number of kernels of the last layer is R with the output feature dimension $R \times 1$. No specific constraints are imposed on the encoder in order to fully use the capacity of the network and reduce the information loss. Instead of using pooling layers to reduce the size of features, we use the stride strategy to compress the data, Except for the last hidden layer, other layers adopt the same activation functions ϕ, such as Sigmoid, rectified linear unit (ReLU), and Leaky ReLU (LReLU). We have conducted extensive experiments to

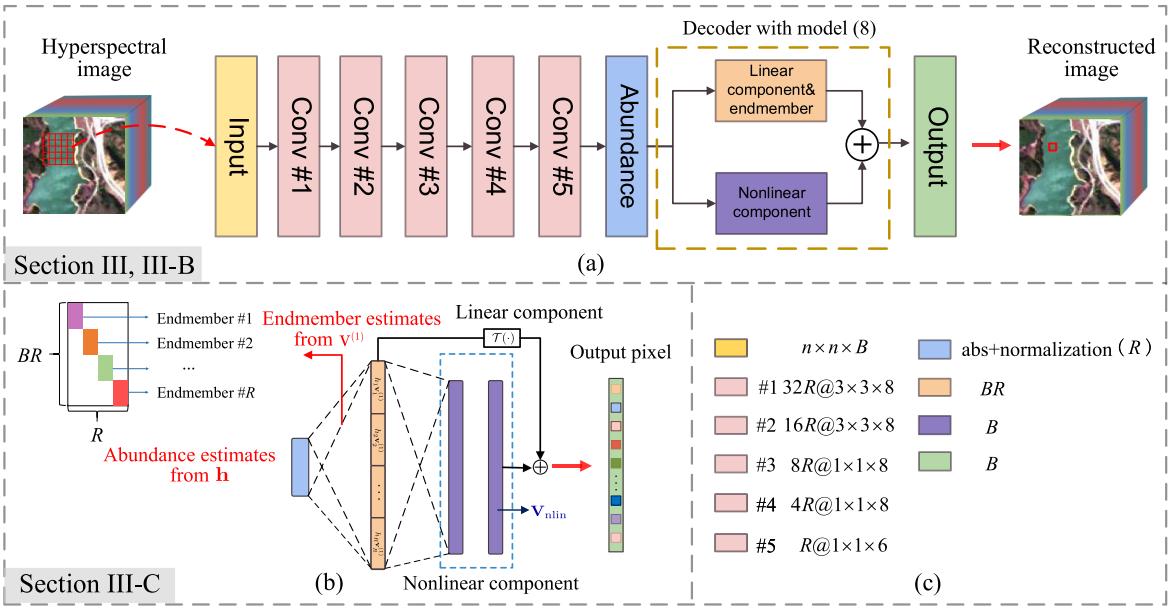


Fig. 2. (a) Diagram of the proposed scheme. (b) Detail structure of the decoder. As presented in Section III-C, $\mathbf{h}$ is the output of the utility layer ((20)), $\mathbf{V}^{(1)}$ are the weights of the first-layer of the decoder, and $\mathbf{V}_{\text{nlin}}$ denotes the weights of the nonlinear part of the decoder. (c) Size of layers.

validate the enhanced performance of LReLU. Hence, LReLU is preferred in this work.

The nonnegativity and sum-to-one constraints imposed on abundance vector $\mathbf{a}$ should be carefully addressed. In order to meet the ANC, the work [27] uses a threshold to enforce the vector to be nonnegative, and the work [29] uses a nonnegative autoencoder to guarantee the ANC over the whole network. The former strategy deactivates a large number of nodes in the network, and the capability of network is thus not fully utilized. The latter strategy imposes strong constraints on the network and makes it difficult to design the network. For the ASC, the works [27] and [29] add a regularization to encourage the ASC, and [25] uses a normalization operator on $\mathbf{a}$. In this work, we address the ANC and ASC using the strategy proposed in our previous work [34]. Absolute value rectification is used to enforce $\mathbf{h}$, the output of the encoder network (abundance estimation), to be nonnegative. The negative values in $\mathbf{h}$ thus becomes a positive value, and nonnegative values remain unchanged. Then this nonnegative vector is normalized by sum of its entries to satisfy sum-to-one, namely

$$h_i = \frac{|h_i|}{\sum_{i=1}^{R} |h_i|} \tag{20}$$

where h_i is the ith element of the abundance vector $\mathbf{h}$. The optimal points $x^* = y^*$ such that $y^* = \arg\min_{x,y=|x|} f(|x|)$ and $x^* = \arg\min_{x\geq 0} f(x)$. Thus, the absolute value can be used in the network to ensure the nonnegativity of parameters. The sum-to-one constraint can also be satisfied by using the layer defined by (20) with the same reason. We found from extensive experiments that using absolute value provides better results than using ReLU and SoftMax in this part.

C. Decoder

The decoder is designed to reconstruct the input with a linear structure and a parallel nonlinear structure. The specific setting of this structure is shown in Fig. 2(b). Recalling the operators and symbols defined in (9) to (13) and the decoder structure given by (17), the first layer of the decoder is then designed by

$$\mathbf{o}^{(1)} = \mathbf{V}^{(1)}\mathbf{h} \tag{21}$$

where $\mathbf{V}^{(1)}$ is defined as weights of the first layer of the decoder. Endmember extracted by linear algorithms, like the vertex component analysis (VCA) algorithm can be used to initialize the learning process of $\mathbf{V}^{(1)}$. $\mathbf{V}^{(1)}$ is constrained with the following form:

$$\mathbf{V}^{(1)} = \text{blkdiag}\{\mathbf{v}_1^{(1)}, \ldots, \mathbf{v}_R^{(1)}\}. \tag{22}$$

Consequently, the product $\mathbf{V}^{(1)}\mathbf{h}$ equals to

$$\mathbf{V}^{(1)}\mathbf{h} = \text{col}\{h_1\mathbf{v}_1^{(1)}, \ldots, h_R\mathbf{v}_R^{(1)}\}. \tag{23}$$

Vectors $\{h_i\mathbf{v}_i^{(1)}\}_{i=1}^{R}$ are generated as estimates of the endmembers weighted by the associated abundances. The output $\mathbf{o}^{(1)}$ of this layer is used as the input of $\mathcal{T}(\cdot)$ so that $\mathcal{T}(\mathbf{o}^{(1)})$ generates the linear component of the spectrum, and $\mathbf{o}^{(1)}$ is also used as the input of the nonlinear component defined by a fully connected (FC) network without bias weights. This nonlinear component of the decoder is designed to represent the nonlinear interactions among the endmembers weighted by the associated abundances. Studies show that a neural network with two hidden layers can represent arbitrary nonlinear relation among the input [36]. In our scheme, we use two hidden FC layers to learn the nonlinear relation among the endmembers, since very high-order photon interactions, though may exist, are usually weak in practice. To avoid overfitting, a parameter norm penalty is added to the weights of the nonlinear component. We shall elaborate on this parameter penalty in the next subsection. This network thus learns the nonlinearity from the data and models all nonlinear interactions among $\{h_i\mathbf{v}_i^{(i)}\}_{i=1}^{R}$. Finally, the outputs of these two

parallel structures are added to reconstruct the estimate $\hat{\mathbf{x}}$ by

$$\hat{\mathbf{x}} = f_D(\mathbf{h}) \tag{24}$$
$$= \hat{\mathbf{x}}_{\text{lin}} + \hat{\mathbf{x}}_{\text{nlin}} \tag{25}$$
$$= \mathcal{T}(\mathbf{o}^{(1)}) + \Phi(\mathbf{o}^{(1)}). \tag{26}$$

The energy of component $\hat{\mathbf{x}}_{\text{nlin}}$ allows to indicate where the nonlinear effects spatially appear, which can be useful in many applications.

D. Objective Function

Several components are considered to formulate the objective function of the proposed autoencoder. The mean-square error between the input and reconstructed data is employed for the data fitting

$$J_{\text{data}}(\mathbf{W}) = \mathcal{L}(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{N}\sum_{i=1}^{N} \| f_D(f_{\text{E}}(\mathbf{x}_i)) - \mathbf{x}_i \|^2. \tag{27}$$

Blind unmixing problem with both endmember and abundance unknown can be a difficult inverse problem. Regularization is often imposed to condition the problem with reasonable prior information. In this work, we first consider the regularity of the nonlinear function Ψ, as proposed in [13]. Thus the ℓ_2-norm of the weights of the nonlinear part of the decoder (denoted by $\mathbf{V}_{\text{nlin}}$) given by

$$J_{\text{reg}}(\mathbf{V}_{\text{nlin}}) = \|\mathbf{V}_{\text{nlin}}\|^2 \tag{28}$$

is used as the regularization to drive the weights to decay and avoid over-fitting. Furthermore, a first-order total variation norm (TV-norm) regularization given by

$$J_{\text{smth}}(\mathbf{V}^{(1)}) = \sum_{i=1}^{R}\sum_{j=1}^{B-1} \left| \left[\mathbf{v}_i^{(1)}\right]_{j+1} - \left[\mathbf{v}_i^{(1)}\right]_j \right| \tag{29}$$

is imposed on $\{\mathbf{v}_i^{(1)}\}_{i=1}^R$. Because $\{\mathbf{v}_i^{(1)}\}_{i=1}^R$ are the estimates of the endmembers, such a regularization encourages the smoothness of the endmembers and reduces the estimation noise. Finally, the objective function is formulated by

$$J(\mathbf{W}) = J_{\text{data}}(\mathbf{W}) + \lambda\, J_{\text{reg}}(\mathbf{V}_{\text{nlin}}) + \gamma\; J_{\text{smth}}(\mathbf{V}^{(1)}) \tag{30}$$

where positive parameters λ and γ control the strengths of the two regularization terms.

IV. Experiments

In this section, the proposed unmixing scheme was implemented and its performance was compared with several typical state-of-the-art unmixing methods, using synthetic data and real airborne image data. Note that the general network structure and number of layers are the same for all experiments.

The performance of abundance estimation was measured by the root mean square error (RMSE) defined by

$$\text{RMSE} = \sqrt{\frac{1}{NR}\sum_{i=1}^{N} \left\|\mathbf{a}_i - \hat{\mathbf{a}}_i\right\|^2} \tag{31}$$

where N represents the number of pixels, $\mathbf{a}_i$ and $\hat{\mathbf{a}}_i$ denote the true and estimated abundance vectors of the ith pixel.

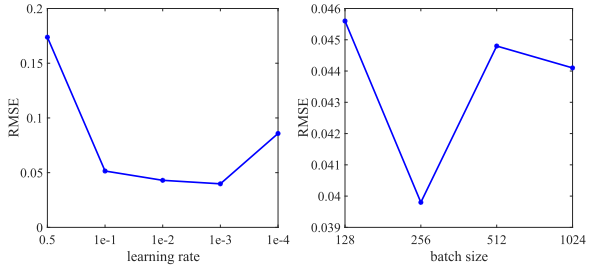


Fig. 3. RMSE as functions of the learning rate and batch size.

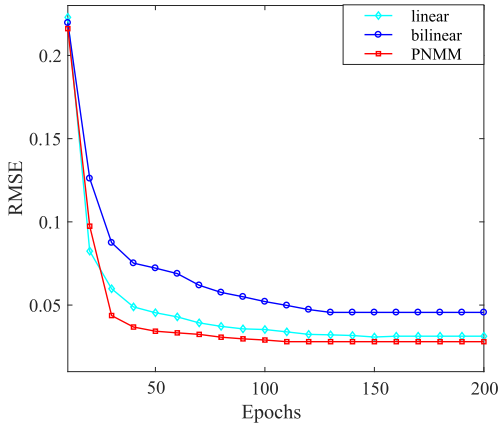


Fig. 4. Convergence curves during 200 epochs.

The accuracy of the endmember estimation was evaluated using the spectral angle distance (SAD) and the spectral information divergence (SID) given by

$$\left\{ \begin{array}{l} \text{SAD} = \cos^{-1}\left(\dfrac{\mathbf{m}^\top \hat{\mathbf{m}}}{\|\mathbf{m}\|\|\hat{\mathbf{m}}\|}\right) \\ \text{SID}(\mathbf{m}|\hat{\mathbf{m}}) = \sum_j \mathbf{p}_j \log\left(\dfrac{\mathbf{p}_j}{\hat{\mathbf{p}}_j}\right) \end{array} \right\} \tag{32}$$

where $\mathbf{m}$ represents an endmember and $\hat{\mathbf{m}}$ represents its estimate, $\mathbf{p} = (\mathbf{m}/\mathbf{1}^\top\mathbf{m})$ is the probability distribution vector of each endmember, and $\hat{\mathbf{p}} = (\hat{\mathbf{m}}/\mathbf{1}^\top\hat{\mathbf{m}})$.

The following typical algorithms were compared:

1) **The endmember extraction with VCA and abundance estimation with K-Hype [13]:** VCA is a classic geometric method used for endmember extraction. The K-Hype algorithm considers the linear-mixture/nonlinear-fluctuation model and approximates the nonlinearity by the kernel trick.
2) **The endmember extraction with VCA and abundance estimation with multilinear model (MLM) [16]:** MLM is based on a Markov chain interpretation of the reflection process of a single light ray. A probability parameter is used to describe the possibility of interacting with the next material.
3) **The endmember extraction with N-finder (N-FINDR) [37] and abundance estimation with nonlinear unmixing by variable splitting and augmented Lagrangian (NUSAL) [38]**: N-FINDR is a classic method that is used to extract endmember. NUSAL is a kernel-based method for NAE by variable splitting and augmented Lagrangian. The method also assumes

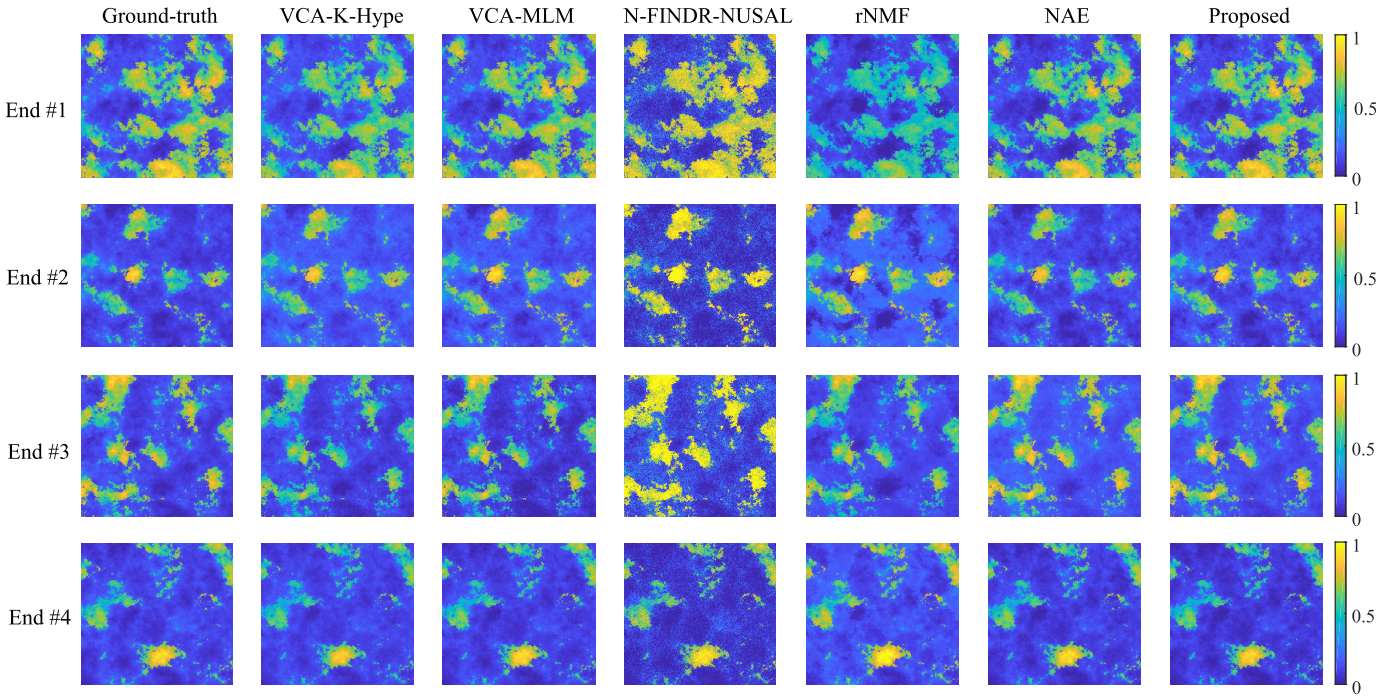


Fig. 5. Abundance maps of the data with the PNMM mixture under SNR = 20 dB. From left to right columns: ground-truth, estimated results of VCA-K-Hype, VCA-MLM, N-FINDR-NUSAL, rNMF, NAE and the proposed method, respectively. From top to bottom: different endmembers.

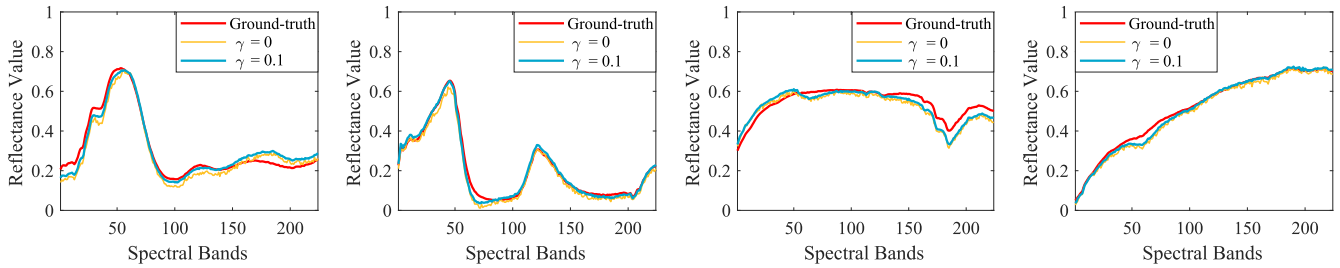


Fig. 6. Illustration of extracted four endmembers from the data with the linear model mixture under SNR = 10 dB. Red curves represent the ground-truth. The yellow and blue curves represent the extracted endmembers with $\gamma = 0$ and $\gamma = 0.1$, respectively. Proper regularization increases the smoothness of the estimated endmembers.

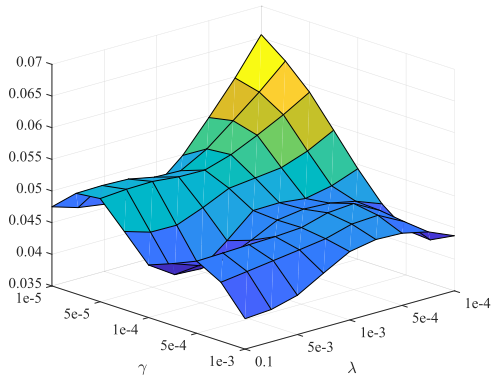


Fig. 7. RMSE as a function of the regularization parameters for the proposed method.

a linear mixing model corrupted by an additive term whose expression can be adapted to account for multiple scattering nonlinearities.

4) **The robust nonnegative matrix factorization (rNMF) [39]:** rNMF is an NMF-based nonlinear method that determines the endmembers and abundances simultaneously via a block-coordinate descent algorithm that involves majorization-minimization updates.
5) **A deep autoencoder network for NAE [34]:** NAE is a novel scheme for blind nonlinear unmixing based on a deep autoencoder network that addresses the post-nonlinear mixture problem.

Note that the linear endmember extraction algorithms were used for the first three methods that are focused on the abundance estimation. These geometrical algorithms still provide sufficiently good results when the nonlinearity degree in data is moderate, as they are able to extract vertices from distorted data clouds [40]. Our experiments will also confirm their performance. All unsupervised nonlinear unmixing methods, namely rNMF, NAE and our proposed method, are initialized by the same VCA results.

A. Experiments With Synthetic Data

1) Data Description: The synthetic data were generated with the LMM and two nonlinear models. The endmembers used to generate the data were extracted from the U.S. Geological Survey (USGS) digital spectral library. These spectra

TABLE III
ABUNDANCE RMSE COMPARISON OF THE SYNTHETIC DATA

	SNR=10dB			SNR=20dB			SNR=30dB		
	linear	bilinear	PNMM	linear	bilinear	PNMM	linear	bilinear	PNMM
VCA-K-Hype	0.0876	0.0897	0.0761	0.0493	0.0797	0.0443	0.0419	0.0517	0.0378
VCA-MLM	0.0792	0.0851	0.0734	0.0350	0.0539	0.0378	0.0289	0.0451	0.0296
N-FINDR-NUSAL	0.1181	0.1354	0.1094	0.0412	0.0797	0.1084	0.0300	0.0507	0.1064
rNMF	0.1698	0.1573	0.1589	0.0669	0.0641	0.0681	0.0556	0.0603	0.0586
NAE	0.0791	0.0783	0.0676	0.0529	0.0642	0.0423	0.0388	0.0632	0.0318
Proposed method	**0.0547**	**0.0690**	**0.0504**	**0.0313**	**0.0456**	**0.0284**	**0.0219**	**0.0398**	**0.0221**

Boldface numbers denote the lowest RMSEs.

TABLE IV
ENDMEMBER SAD COMPARISON OF THE SYNTHETIC DATA

	SNR=10dB			SNR=20dB			SNR=30dB		
	linear	bilinear	PNMM	linear	bilinear	PNMM	linear	bilinear	PNMM
VCA-K-Hype/MLM	3.8528	4.3132	**5.0008**	0.8501	3.2297	5.4322	0.1758	3.1177	5.3712
N-FINDR-NUSAL	19.1240	19.6878	18.8068	6.5615	6.7955	7.8875	1.8875	2.2535	5.8264
rNMF	4.0501	4.3968	6.8123	3.1724	4.5322	6.2772	3.0019	2.6911	5.2633
NAE	3.8898	4.5667	5.1826	1.1031	3.4624	5.6240	0.5974	3.4668	5.5974
Proposed method	**3.6393**	**4.2688**	5.0218	**0.8497**	**3.1142**	**5.0416**	**0.1719**	**3.0457**	**5.2680**

Boldface numbers denote the lowest SADs.

TABLE V
ENDMEMBER SID COMPARISON OF THE SYNTHETIC DATA

	SNR=10dB			SNR=20dB			SNR=30dB		
	linear	bilinear	PNMM	linear	bilinear	PNMM	linear	bilinear	PNMM
VCA-K-Hype/MLM	0.0096	**0.0155**	**0.0127**	**0.0013**	0.0116	0.0164	0.0003	0.0112	0.0158
N-FINDR-NUSAL	0.2181	0.2222	0.2437	0.0388	0.0377	0.0572	0.0032	0.0149	0.0192
rNMF	0.0156	0.0157	0.0512	0.0063	0.0238	0.0333	0.0040	0.0140	0.0190
NAE	0.0104	0.0166	0.0136	0.0020	0.0165	0.0187	**0.0002**	0.0161	0.0171
Proposed method	**0.0086**	0.0156	0.0131	**0.0013**	**0.0107**	**0.0157**	**0.0002**	**0.0101**	**0.0154**

Boldface numbers denote the lowest SIDs.

consist of 224 contiguous bands. The LMM is given by (4). The bilinear mixture model

$$\mathbf{x} = \mathbf{M}\mathbf{a} + \sum_{i=1}^{R-1} \sum_{j=i+1}^{R} a_i a_j (\mathbf{m}_i \odot \mathbf{m}_j) + \mathbf{n} \tag{33}$$

and the post-nonlinear mixing model (PNMM)

$$\mathbf{x} = \mathbf{M}\mathbf{a} + \mathbf{M}\mathbf{a} \odot \mathbf{M}\mathbf{a} + \mathbf{n} \tag{34}$$

were used as the two nonlinear models. In this experiment, four pure material spectra ($R = 4$) were considered and the abundance fractions were generated from Hyperspectral Imagery Synthesis (HYDRA) toolbox. A total number of 256×256 pixels were generated to evaluate the performance. Zero-mean Gaussian noise was added with the signal-to-noise ratio (SNR) set to 10, 20, and 30 dB, respectively. Our proposed scheme was implemented using PyTorch and Torch. We used Adam optimizer to train the network. Adam is a simple and computationally efficient algorithm for gradient-based optimization of stochastic objective functions. The specific parameters are given in the following subsections.

2) Results: Our proposed method is an unsupervised method and not based on training and test data, the unmixing results are obtained by using all data as input. Rules of manual inspection guide most choices of parameters. The learning rate was set to 1×10^{-3}. The batch size was set to 256 in this experiment. Fig. 3 shows the performance of our proposed with respect to learning rate and batch size with bilinear model SNR = 30 dB. Note that a larger batch size leads to more accurate descent directions but increases the possibility of reaching a local optimum, while a small batch size may result in difficulties in convergence. The number of training epochs was set to 200. The spatial window size n was set to 5. By using a grid search strategy, we manually set the parameter λ to 1×10^{-1}, and the smoothing regularization parameter γ to 1×10^{-3}. Fig. 4 shows the convergence curves during the learning process with the data under SNR = 20 dB.

Tables III–V report the RMSE, SAD and SID results of the compared methods under different models and SNR settings. It is clear that our proposed method achieves the best abundance estimation performance, and sufficiently good endmember estimation performance with both linear and nonlinear models. Note that when the mixtures are affected by moderate nonlinearities, geometrical endmember extraction algorithms based on linear model can still provide sufficiently good results for nonlinear mixtures, in particular when constraints on simplex volumes are imposed [40]. Among compared algorithms, MLM is an NAE method with a specific assumption on the nonlinearity. Both K-Hype and NUSAL are kernel-based methods, and the selection of the kernel and its parameters notably affect their performance. The proposed method builds a model by learning the nonlinearity from the observed

Fig. 8. Laboratory-created data for unmixing performance evaluation (RGB images). (a)–(c) Pure quartz sand with a diameter of 0.3 mm of three colors. They serve as pure materials for providing endmembers. (d) and (e) Mixtures of sand with spatial patterns. Square regions of 60-by-60 pixels in the center of each subfigures are clipped out and used in experiments.

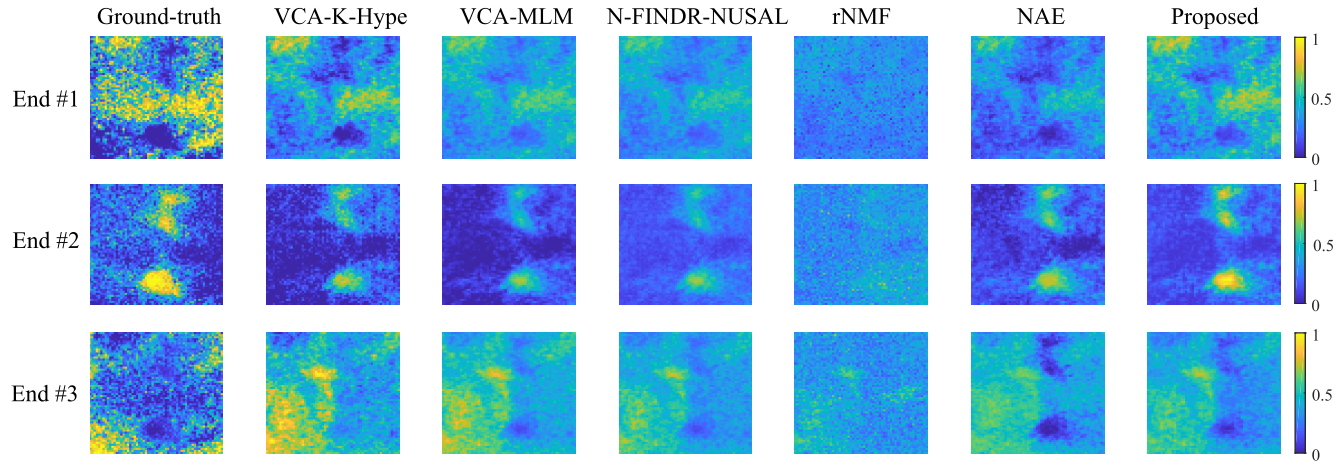


Fig. 9. Abundance maps of the first mixture of the laboratory-created data. From left to right columns: ground-truth, estimated results of VCA-K-Hype, VCA-MLM, N-FINDR-NUSAL, rNMF, NAE, and the proposed method, respectively. From top to bottom: abundance maps of red quartz sand, green quartz sand and blue quartz sand, respectively.

data, and therefore the issue of the kernel selection is then avoided. Compared to the state-of-the-art unsupervised NAE methods, namely rNMF and NAE with the same initialization, our proposed method almost always improves the abundance estimation accuracy. Moreover, benefiting from the fact that the low-dimensional vector generated from encoder maintains the main information and gets rid of redundant information and noise, the proposed method is robust to noise. Fig. 5 presents the abundance maps of our method and compared algorithms. In order to understand the effect of the smoothing regularization, we show in Fig. 6 the extracted endmembers with γ set to 0 and 1×10^{-1} in the linear case with SNR $= 10$ dB. Removing this regularization ($\gamma = 0$) leads to noisy estimated endmember curves. The usefulness of this smoothing effect is clearly illustrated. Fig. 7 shows the sensitivity of the proposed method with the regularization parameters λ and γ with the synthetic data of the bilinear model under SNR $= 30$ dB. It can be seen the method exhibits satisfactory RMSE within a reasonable range around the optimal parameter values.

B. Experiments With Laboratory-Created Data

In order to perform quantitative evaluation of unmixing performance with real data, we designed several experimental scenes with known ground-truth in our laboratory. Our data were collected by the GaiaField and GaiaSorter systems in our laboratory. Our GaiaField (Sichuan Dualix Spectral Image Technology Co. Ltd., GaiaField-V10) is a push-broom imaging spectrometer with HyperSpectral Image Asia (HSIA)-OL50 lens, covering the visible and near infrared (NIR) wavelengths ranging from 400 to 1000 nm, with a spectral resolution up to 0.58 nm. GaiaSorter sets an environment that isolates external lights and is endowed with a conveyer to move samples for the push-broom imaging.

Two nonuniform mixtures of colored quartz sand with spatial patterns were used. The experimental settings were strictly controlled so that pure material spectral signatures and material compositions were known. The data consist of 256 spectral bands. Different colors of quartz sands with uniform size were used as pure materials shown in Fig. 8(a)–(c). The mixtures are shown in Fig. 8(d) and (e). To calculate the ground truth, the aligned high-resolution RGB images of these scenes were captured and linked to hyperspectral pixels using the spatial resolution ratio, and then the percentage of each colored sand in a low-resolution hyperspectral pixel could be analyzed with the help of the associated RGB image. In our experiments, a subimage of 60-by-60 was clipped out from the center of each subfigure. More details can be seen in [41].

In this set of experiments, the learning rate was also set to 1×10^{-3}, and the Adam optimizer was used to train the network. The batch size of this experiment was set to 100 and the number of training epochs was set to 200. The parameter λ was set to 1×10^{-4}, and γ was set to 1×10^{-6}. Figs. 9 and 10 illustrate the estimated abundance maps of these algorithms. The ground-truth abundance maps are shown in the first columns of Figs. 9 and 10. The abundance maps estimated using the compared algorithms and our proposed method are shown alongside. The proposed algorithm results have sharper abundance maps, and the general spatial patterns of the estimated maps are more consistent with the ground-truth.

TABLE VI
RMSE, SAD AND SID COMPARISON OF UNMIXING RESULTS OF THE LABORATORY-CREATED DATA

		VCA-K-Hype	VCA-MLM	N-FINDR-NUSAL	rNMF	NAE	Proposed
RMSE	Mixture 1	0.1957	0.2050	0.2029	0.2315	0.2035	**0.1879**
	Mixture 2	0.1764	0.2198	0.1946	0.2212	0.1797	**0.1721**
SAD	Mixture 1	10.7889	—	**9.2097**	19.8765	10.7815	9.2958
	Mixture 2	9.3823	—	12.2489	15.5301	9.3736	**9.3094**
SID	Mixture 1	0.0995	—	**0.0325**	0.2128	0.0994	0.0935
	Mixture 2	0.0326	—	0.0568	0.1233	**0.0325**	**0.0325**

Boldface numbers denote the lowest results.

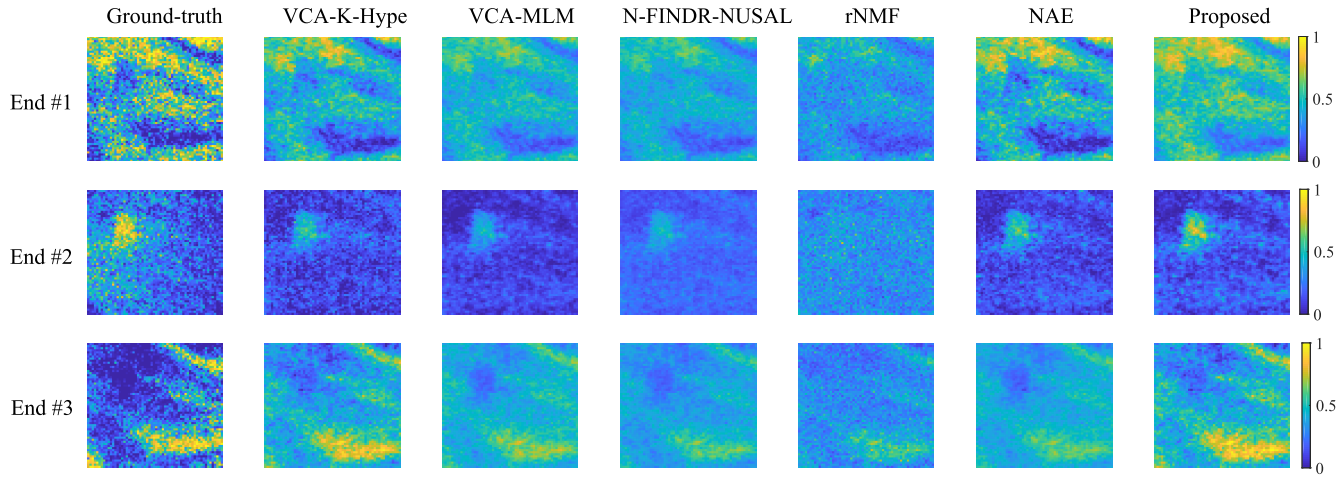


Fig. 10. Abundance maps of the second mixture of the laboratory-created data. From left to right columns: ground-truth, estimated results of VCA-K-Hype, VCA-MLM, N-FINDR-NUSAL, rNMF, NAE, and the proposed method, respectively. From top to bottom: abundance maps of red quartz sand, green quartz sand and blue quartz sand, respectively.

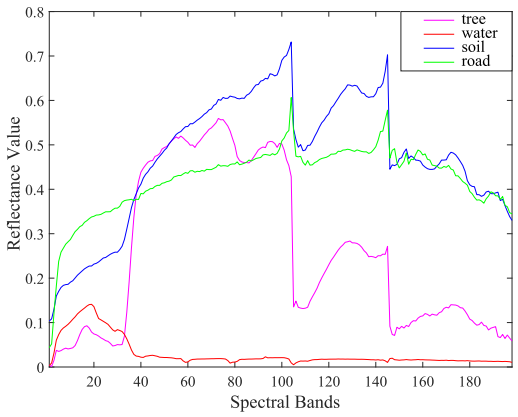


Fig. 11. Extracted endmembers from the Jasper Ridge data by the proposed algorithm.

The quantitative RMSE, SAD and SID results are compared in Table VI. We observe that the proposed algorithm achieves the lowest RMSEs and sufficiently good endmember estimation performance. These unmixing results with labeled real data highlight the superior performance of the proposed method.

C. Experiments With Real Airborne Data

Two real airborne images, namely, Jasper Ridge dataset and Urban dataset, were used to validate the proposed scheme.

1) Jasper Ridge: It is an image of Jasper Ridge Biological Preserve, which was recorded by Analytical Imaging and Geophysics (AIG) in 1999. There are 512×614 pixels in it. We use a subimage with 100×100 pixels in our work. The first pixel corresponds to the pixel (105, 269) in the original image, which is also used in many other works [31], [33]. Each pixel was recorded at 224 channels ranging from 380 to 2500 nm with spectral resolution up to 9.46 nm. After removing the channels affected by water vapor and the atmospheric environment, 198 channels were kept. The number of endmembers was set to 4, including tree, water, soil, and road.

The learning rate was set to 1×10^{-3} for this dataset. The batch size used in this experiment was set to 256, and the number of training epochs was set to 200. The parameter λ was set to 1×10^{-3}, and γ was set to 1×10^{-6} in this experiment. Fig. 11 illustrates the extracted endmembers by the proposed algorithm. Fig. 12 illustrates the estimated abundance maps of the four endmembers obtained by these algorithms. We observe that the proposed algorithm provides a shaper and clearer map of different materials. Fig. 13 shows the energy of the nonlinear components estimated by these algorithms. These maps demonstrate that nonlinear components are active at the boundary or transition parts of different regions, e.g., at the water shore. The proposed algorithm provides a clear map of nonlinear components with several particular locations emphasized.

Note that this real data is extensively used in hyperspectral unmixing, however, no ground-truth information is available for a quantitative performance evaluation of abundance. Thus,

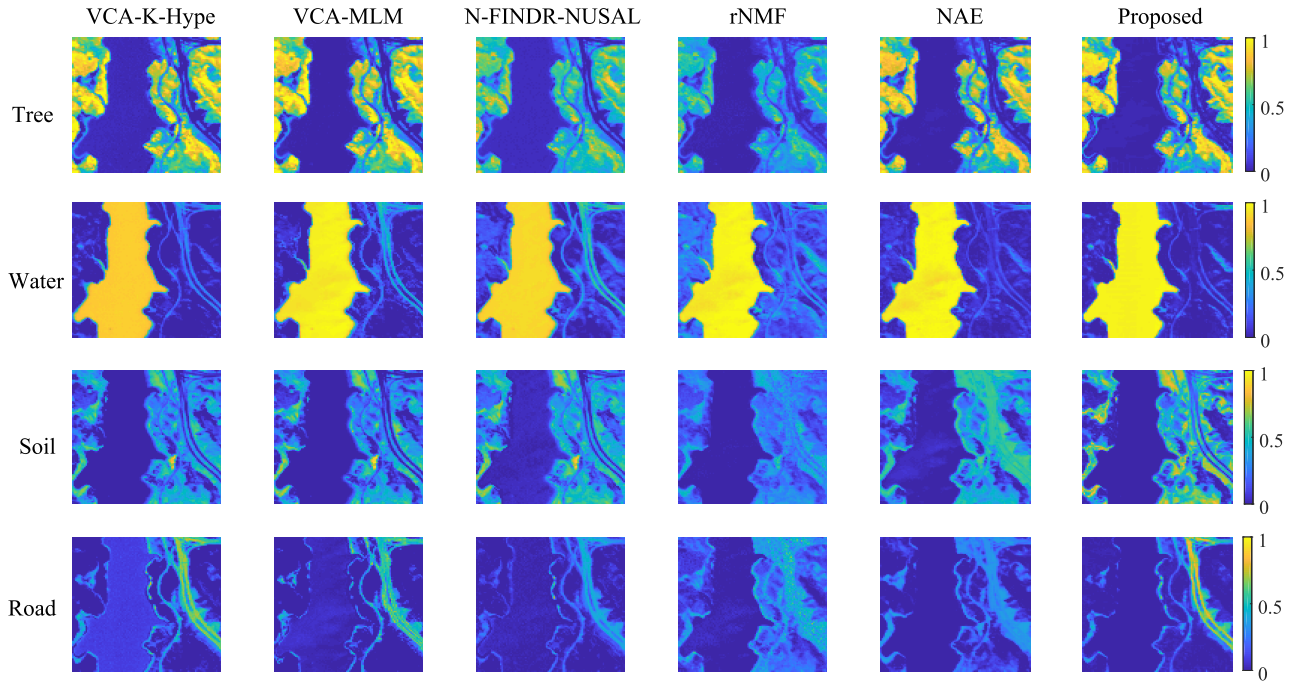


Fig. 12. Estimated abundance maps of Jasper Ridge data. From left to right: VCA-K-Hype, VCA-MLM, N-FINDR-NUSAL, rNMF, NAE and the proposed method. From top to bottom: tree, water, soil and road, respectively.

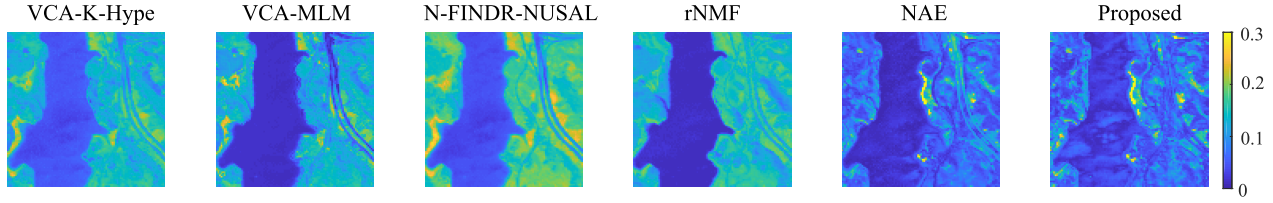


Fig. 13. Energy of the nonlinear components of the Jasper Ridge data. From left to right: VCA-K-Hype, VCA-MLM, N-FINDR-NUSAL, rNMF, NAE, and the proposed method.

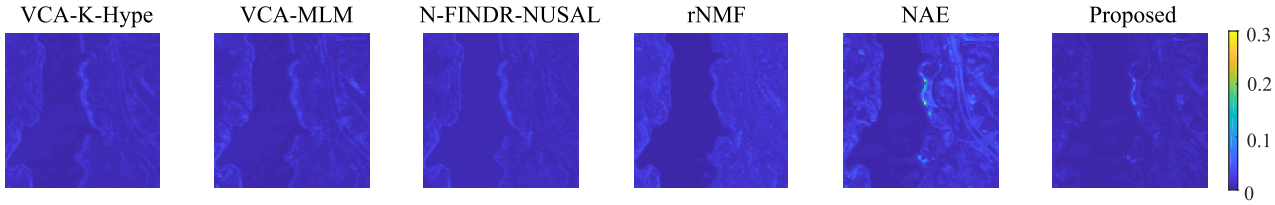


Fig. 14. Maps of RE of the Jasper Ridge data. From left to right: VCA-K-Hype, VCA-MLM, N-FINDR-NUSAL, rNMF, NAE, and the proposed method, respectively.

the RE defined by

$$\mathrm{RE} = \sqrt{\frac{1}{NR}\sum_{i=1}^{N} \|\mathbf{x}_i - \widehat{\mathbf{x}}_i\|_2} \tag{35}$$

is used for a quantitative comparison, though RE may not be proportional to the abundance estimation accuracy. The RE results of different algorithms are reported in Table VII, and the RE maps are illustrated in Fig. 14. We observe that our method leads to the lowest RE in the mean sense and the spatial distribution.

2) Urban: It was captured by the Hyperspectral Digital Imagery Collection Experiment (HYDICE) in October 1995. Its location is an urban area at Copperas Cove, TX, U. It has 307×307 pixels, with each pixel covering 2×2 m^2 area. All the pixels were used to evaluate the unmixing performance. The data consist of 210 spectral bands ranging from 400 to 2500 nm with spectral resolution up to 10 nm. After removing channels [1–4, 76, 87, 101–111, 136–153, 198–210] affected by dense water vapor and the atmosphere, 162 channels remained. Five prominent endmembers exist in this data, namely, asphalt, grass, tree, roof, and dirt.

In this experiment, the same network, learning rate, and optimizer were used to conduct the unmixing study. The batch size was set to 200, with the number of epoch set to 200. The parameter λ was set to 1×10^{-4} and γ was set to 1×10^{-6}. The estimated abundance maps of five endmembers are shown in Fig. 15. These figures clearly indicate that our proposed method provides a smoother and clearer map. Fig. 16 shows the energy of the nonlinear components estimated by these

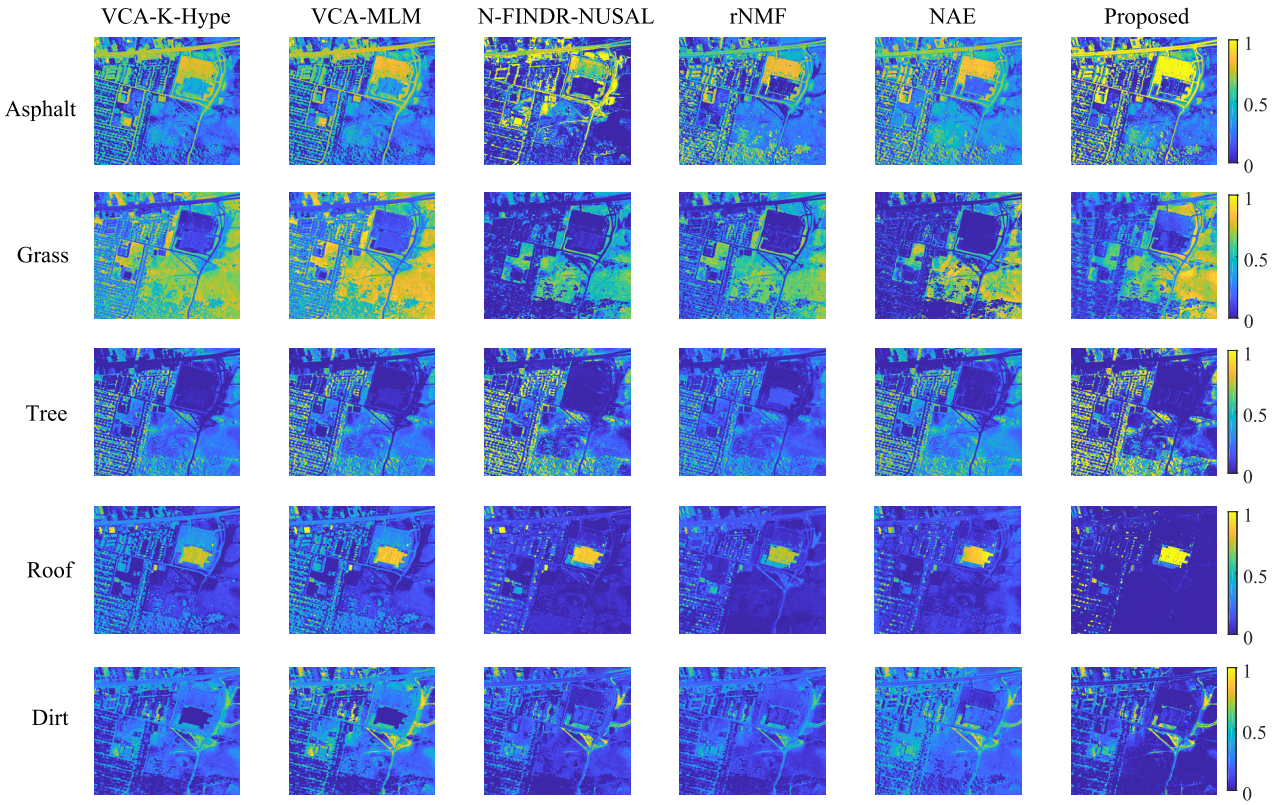


Fig. 15. Estimated abundance maps of Urban data. From left to right: VCA-K-Hype, VCA-MLM, N-FINDR-NUSAL, rNMF, NAE and the proposed method. From top to bottom: asphalt, grass, tree, roof, and dirt.

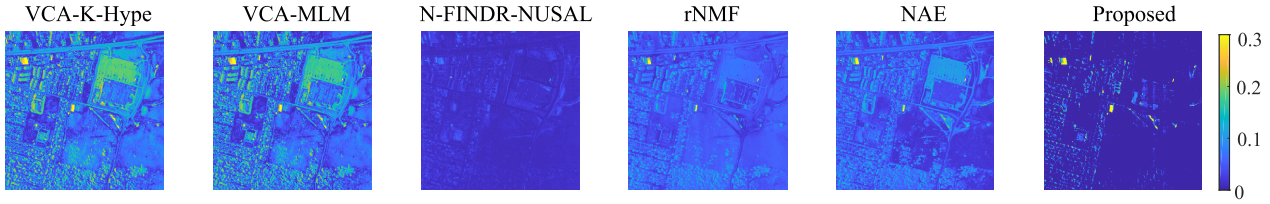


Fig. 16. Energy of the nonlinear components of the Urban data. From left to right: VCA-K-Hype, VCA-MLM, N-FINDR-NUSAL, rNMF, NAE, and the proposed method.

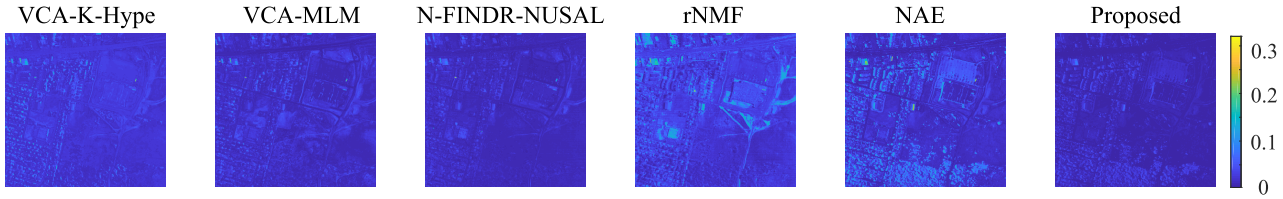


Fig. 17. RE maps of the Urban data. From left to right: VCA-K-Hype, VCA-MLM, N-FINDR-NUSAL, rNMF, NAE, and the proposed method.

algorithms. These maps demonstrate that nonlinear components are active at vegetated regions and boundary or transition parts of different regions. The proposed algorithm provides a clearer map of nonlinear components. The RE results achieved by different algorithms are reported in Table VIII, and the RE maps are shown in Fig. 17. We observed that our method leads to the lowest RE. Fig. 18 shows the extracted endmembers by the proposed method.

D. Effect of Using the Training Strategy

The proposed method is based on the autoencoder structure without requiring separate training datasets. The estimated parameters are given once the network is learned with the data reconstruction process. However, it is possible to only use a part of the data to learn the autoencoder and then apply the learned encoder to estimate the abundances for the rest data. This strategy may further reduce the complexity of the method. To validate this strategy, experiments with real data were conducted. We randomly selected 20% data to train the model, and then the whole image was fed into the trained model to obtain the unmixing results. Fig. 19 shows the estimated abundance maps. These results are similar to the abundance maps of our proposed method in Figs. 12 and 15. These results show that our model is stable and the training strategy can be useful for spectral unmixing.

TABLE VII

RE COMPARISON OF THE JASPER RIDGE DATA

Algorithm	VCA-K-Hype	VCA-MLM	N-FINDR-NUSAL	rNMF	NAE	Proposed
RE	0.0131	0.0138	0.0132	0.0135	0.0154	**0.0118**

Boldface numbers denote the lowest RE value.

TABLE VIII

RE COMPARISON OF THE URBAN DATA

Algorithm	VCA-K-Hype	VCA-MLM	N-FINDR-NUSAL	rNMF	NAE	Proposed
RE	0.0144	0.0139	0.0123	0.0176	0.0211	**0.0116**

Boldface numbers denote the lowest RE value.

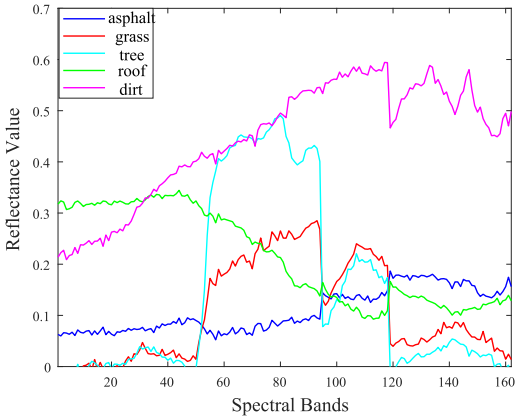
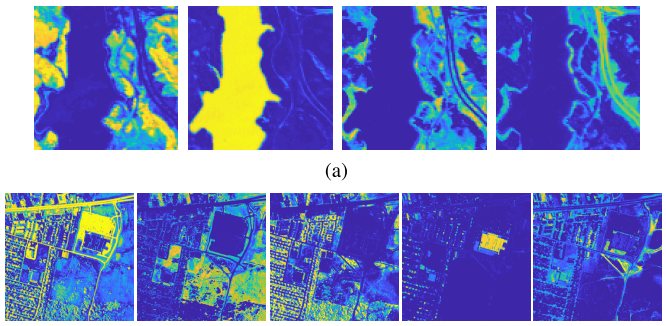


Fig. 18. Extracted endmembers from the Urban data by the proposed algorithm.



(a)

(b)

Fig. 19. Abundance maps of using a model trained with 20% data. (a) Maps of Jasper Ridge data, from left to right: tree, water, soil, and road. (b) Maps of Urban data, from left to right: asphalt, grass, tree, roof, and dirt.

TABLE IX

TIME CONSUMING OF DIFFERENT IMAGES (RESPECTIVELY, USING FULL DATA AND TRAINING STRATEGY)

	Synthetic data	Laboratory-created data	Jasper Ridge	Urban
time with full data(s)	851	182	246	675
training time(s)	326	74	112	254
test time(s)	71	15	21	38
total time with training(s)	397	89	133	292

E. Execution Time

In this section, we conduct experiments to evaluate the execution time of our proposed method. Note that all these experiments are carried out on the same hardware sources (Intel Xeon E5-2650 v3 and an NVIDIA Tesla k80). Time consuming of synthetic data, laboratory-created data, Jasper Ridge dataset, and Urban dataset are shown in the first row of Table IX. We can see that the running time of the proposed framework mainly depends on the size of the image. Moreover, the second and third rows show the computing time with the training strategy where 20% data were selected for constructing the autoencoder. We observe that compared to learn the autoencoder with full data, this strategy reduces the computation time since that inferring (test) process requires much less time.

V. CONCLUSION

This article presented an unsupervised nonlinear spectral unmixing method based on a CNN autoencoder network. The proposed approach benefits from the 3-D-CNN-based networks jointly capture the spectral–spatial information of the hyperspectral image. Our framework applied a general mixture model consisting of a linear mixture component and an additive nonlinear mixture component, which can utilize the universal modeling ability of deep neural networks to learn the inherent nonlinearity of the nonlinear mixture component from the data itself. We compared our proposed method with four state-of-the-art NAE methods with both synthetic and real data, especially the laboratory-created data with known ground truth. Experiment results demonstrated the superior performance of our proposed method, results with different SNRs validate that our proposed method is robust to noise, and various kinds of nonlinearity data confirm the generalization ability of our deep framework. Moreover, our proposed method showed better RMSE results compared with other compared methods, indicating that our method can extract more accurate abundance maps. It also achieved good endmember extraction results. Future work will integrate nonlocal information of the image into the autoencoder network to further enhance its performance.

REFERENCES

[1] M. Zhao, M. Wang, J. Chen, and S. Rahardja, "Hyperspectral unmixing via deep autoencoder networks for a generalized linear-mixture/nonlinear-fluctuation model," 2019, *arXiv:1904.13017*. [Online]. Available: http://arxiv.org/abs/1904.13017

[2] J. M. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[3] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.

[4] J. Yao, D. Meng, Q. Zhao, W. Cao, and Z. Xu, "Nonconvex-sparsity and nonlocal-smoothness-based blind hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 28, no. 6, pp. 2991–3006, Jun. 2019.

[5] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1923–1938, Apr. 2019.

[6] R. Heylen, M. Parente, and P. Gader, "A review of nonlinear hyperspectral unmixing methods," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 1844–1868, Jun. 2014.

[7] C. C. Borel and S. A. W. Gerstl, "Nonlinear spectral mixing models for vegetative and soil surfaces," *Remote Sens. Environ.*, vol. 47, no. 3, pp. 403–416, 1994.

[8] W. Fan, B. Hu, J. Miller, and M. Li, "Comparative study between a new nonlinear model and common linear model for analysing laboratory simulated-forest hyperspectral data," *Int. J. Remote Sens.*, vol. 30, no. 11, pp. 2951–2962, Jun. 2009.

[9] A. Halimi, Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Unmixing hyperspectral images using the generalized bilinear model," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2011, pp. 1886–1889.

[10] Y. Altmann, A. Halimi, N. Dobigeon, and J.-Y. Tourneret, "Supervised nonlinear spectral unmixing using a postnonlinear mixing model for hyperspectral imagery," *IEEE Trans. Image Process.*, vol. 21, no. 6, pp. 3017–3025, Jun. 2012.

[11] B. Hapke, "Bidirectional reflectance spectroscopy: 1. Theory," *J. Geophys. Res.*, vol. 86, no. B4, pp. 3039–3054, 1981.

[12] R. Close, P. Gader, J. Wilson, and A. Zare, "Using physics-based macroscopic and microscopic mixture models for hyperspectral pixel unmixing," *Proc. SPIE*, vol. 8390, no. 1, May 2012, Art. no. 83901L.

[13] J. Chen, C. Richard, and P. Honeine, "Nonlinear unmixing of hyperspectral data based on a linear-mixture/nonlinear-fluctuation model," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 480–492, Jan. 2013.

[14] J. Chen, C. Richard, and P. Honeine, "Nonlinear estimation of material abundances in hyperspectral images with ℓ_1-norm spatial regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2654–2665, May 2014.

[15] R. Ammanouil, A. Ferrari, C. Richard, and S. Mathieu, "Nonlinear unmixing of hyperspectral data with vector-valued kernel functions," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 340–354, Jan. 2017.

[16] R. Heylen and P. Scheunders, "A multilinear mixing model for nonlinear spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 240–251, Jan. 2016.

[17] R. Heylen, V. Andrejchenko, Z. Zahiri, M. Parente, and P. Scheunders, "Nonlinear hyperspectral unmixing with graphical models," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4844–4856, Jul. 2019.

[18] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2094–2107, Jun. 2014.

[19] L. Mou, P. Ghamisi, and X. X. Zhu, "Deep recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3639–3655, Jul. 2017.

[20] Y. Chen, X. Zhao, and X. Jia, "Spectral–spatial classification of hyperspectral data based on deep belief network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2381–2392, Jun. 2015.

[21] W. S. Hu, H.-C. Li, L. Pan, W. Li, R. Tao, and Q. Du, "Spatial–spectral feature extraction via deep convlstm neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 4237–4250, Jun. 2020.

[22] G. A. Licciardi and F. D. Frate, "Pixel unmixing in hyperspectral data by means of neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4163–4172, Nov. 2011.

[23] X. Zhang, Y. Sun, J. 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.

[24] R. Guo, W. Wang, and H. Qi, "Hyperspectral image unmixing using autoencoder cascade," in *Proc. WHISPERS*, 2015, pp. 1–4.

[25] 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.

[26] Y. Qu, R. Guo, and H. Qi, "Spectral unmixing through part-based nonnegative constraint denoising autoencoder," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2017, pp. 209–212.

[27] Y. Qu and H. 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.

[28] Y. Su, A. Marinoni, J. Li, A. Plaza, and P. Gamba, "Nonnegative sparse autoencoder for robust endmember extraction from remotely sensed hyperspectral images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2017, pp. 205–208.

[29] Y. 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.

[30] S. Ozkan, B. Kaya, and G. B. Akar, "EndNet: Sparse AutoEncoder network for endmember extraction and hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 482–496, Jan. 2019.

[31] R. A. Borsoi, T. Imbiriba, and J. C. M. Bermudez, "Deep generative endmember modeling: An application to unsupervised spectral unmixing," *IEEE Trans. Comput. Imag.*, vol. 6, no. 1, pp. 374–384, Oct. 2020.

[32] 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.

[33] F. Khajehrayeni and H. Ghassemian, "Hyperspectral unmixing using deep convolutional autoencoders in a supervised scenario," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, no. 1, pp. 567–576, Feb. 2020.

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

[35] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sensing*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.

[36] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[37] M. E. Winter, "N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," *Proc. SPIE*, vol. 3753, pp. 266–276, Oct. 1999.

[38] A. Halimi, J. M. Bioucas-Dias, N. Dobigeon, G. S. Buller, and S. McLaughlin, "Fast hyperspectral unmixing in presence of nonlinearity or mismodeling effects," *IEEE Trans. Comput. Imag.*, vol. 3, no. 2, pp. 146–159, Jun. 2017.

[39] C. Févotte and N. Dobigeon, "Nonlinear hyperspectral unmixing with robust nonnegative matrix factorization," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4810–4819, Dec. 2015.

[40] N. Dobigeon, J.-Y. Tourneret, C. Richard, J. C. M. Bermudez, S. McLaughlin, and A. O. Hero, "Nonlinear unmixing of hyperspectral images: Models and algorithms," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 82–94, Jan. 2014.

[41] M. Zhao, J. Chen, and Z. He, "A laboratory-created dataset with ground truth for hyperspectral unmixing evaluation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 7, pp. 240–251, Jul. 2019.

Min Zhao (Graduate Student Member, IEEE) received the B.S. degree in electronic and information engineering and the M.S. degree in signal and information processing from Northwestern Polytechnical University, Xi'an, China, in 2017 and 2020, respectively, where she is pursuing the Ph.D. degree.

Her research interests include hyperspectral image analysis and object detection.

Dr. Zhao received the People's choice Award of Three Minute Thesis (3MT) Competition of IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2020 and the (with team) Champion of Grand Challenges on NIR Image Colorization of IEEE International Conference on Visual Communications and Image Processing (VCIP) 2020.

Mou Wang (Student Member, IEEE) received the B.S. degree in electronics and information engineering from Northwestern Polytechnical University, Xi'an, China, in 2016, where he is pursuing the Ph.D. degree.

His research interests include machine learning and speech signal processing.

Jie Chen (Senior Member, IEEE) received the B.S. degree from Xi'an Jiaotong University, Xi'an, China, in 2006, the Dipl.-Ing. degree in information and telecommunication engineering from the University of Technology of Troyes (UTT), Troyes, France, in 2009, and the M.S. degree in information and telecommunication engineering from Xi'an Jiaotong University, in 2009, and the Ph.D. degree in systems optimization and security from UTT in 2013.

From 2013 to 2014, he was with the Lagrange Laboratory, University of Nice Sophia Antipolis, Nice, France. From 2014 to 2015, he was with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA. He is a Professor with Northwestern Polytechnical University, Xi'an. His research interests include adaptive signal processing, distributed optimization, hyperspectral image analysis, and acoustic signal processing.

Dr. Chen was the Technical Co-Chair of the International Workshop on Acoustic Echo and Noise Control (IWAENC) 2016, the IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC) 2021 held in Xi'an. He serves as a Distinguished Lecturer of Asia-Pacific Signal and Information Processing Association from 2018 to 2019, and a Co-Chair of IEEE Signal Processing Society Summer School 2019 in Xi'an. He will be the General Co-Chair of IEEE Machine Learning for Signal Processing (MLSP) 2022.

Susanto Rahardja (Fellow, IEEE) received the B.Eng. degree from the National University of Singapore, Singapore, in 1991, and the M.Eng. and Ph.D. degrees in electronic engineering from Nanyang Technological University, Singapore, in 1993 and 1997, respectively.

He is a Chair Professor with the Northwestern Polytechnical University (NPU), Xi'an, China, under the Thousand Talent Plan of People's Republic of China. His research interests are in multimedia, signal processing, wireless communications, discrete transforms, machine learning and signal processing algorithms and implementation. He contributed to the development of a series of audio compression technologies such as Audio Video Standards AVS-L, AVS-2 and International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC) 14496-3:2005/Amd.2:2006, ISO/IEC 14496-3:2005/Amd.3:2006 in which some have been licensed to several companies. He has more than 15 years of experience in leading research team for media related research that cover areas in signal processing (audio coding, video/image processing), media analysis (text/speech, image, video), media security (biometrics, computer vision and surveillance) and sensor networks. He has published more than 300 articles and has been granted more than 70 patents worldwide out of which 15 are US patents.

Dr. Rahardja was a recipient of several honors including the IEE Hartree Premium Award, the Tan Kah Kee Young Inventors' Open Category Gold Award, the Singapore National Technology Award, the A*STAR Most Inspiring Mentor Award, the Finalist of the 2010 World Technology and Summit Award, the Nokia Foundation Visiting Professor Award, and the ACM Recognition of Service Award. He attended the Stanford Executive Program at the Graduate School of Business in Stanford University, Stanford, CA, USA. He was a past Associate Editors of IEEE TRANSACTIONS ON AUDIO, SPEECH AND LANGUAGE PROCESSING and the IEEE TRANSACTIONS ON MULTIMEDIA, a past Senior Editor of the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, and is serving as an Associate Editors for the Elsevier *Journal of Visual Communication and Image Representation* and the IEEE TRANSACTIONS ON MULTIMEDIA. He was the Conference Chair of the 5th the Association of Computing Machinery (ACM) Special Interest Group on Computer GRAPHics Asia (SIGGRAPHASIA) in 2012 and the Asia–Pacific Signal and Information Processing Association (APSIPA) 2nd Summit and Conference in 2010 and 2018 as well as other conferences in ACM, the International Society for Optical Engineering (SPIE), and IEEE.