

# Chapter 13

## Recent Advances in Hyperspectral Unmixing Using Sparse Techniques and Deep Learning



Shaoquan Zhang, Yuanchao Su, Xiang Xu, Jun Li, Chengzhi Deng and Antonio Plaza

**Abstract** Spectral unmixing is an important technique for remotely sensed hyperspectral image interpretation that expresses each (possibly mixed) pixel vector as a combination of pure spectral signatures (endmembers) and their fractional abundances. Recently, sparse unmixing and deep learning have emerged as two powerful approaches for spectral unmixing. In this chapter, we focus on two particularly innovative contributions. First, we provide an overview of recent advances in semi-supervised sparse unmixing algorithms, with particular emphasis on techniques that include spatial–contextual information for a better scene interpretation. These algorithms require a spectral library of signatures available a priori to conduct the unmixing. Then, we describe new developments in the use of deep learning for spec-

---

S. Zhang · C. Deng

Jiangxi Province Key Laboratory of Water Information Cooperative Sensing and Intelligent Processing, Nanchang Institute of Technology, Nanchang 330099, China

e-mail: [zhangshaoquan@nit.edu.cn](mailto:zhangshaoquan@nit.edu.cn)

C. Deng

e-mail: [dengchengzhi@126.com](mailto:dengchengzhi@126.com)

Y. Su · J. Li (✉)

Guangdong Provincial Key Laboratory of Urbanization and Geo-Simulation, School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China

e-mail: [lijun48@mail.sysu.edu.cn](mailto:lijun48@mail.sysu.edu.cn)

Y. Su

e-mail: [suych3@mail2.sysu.edu.cn](mailto:suych3@mail2.sysu.edu.cn)

X. Xu

Zhongshan Institute, University of Electronic Science and Technology of China, Zhongshan 528402, China

e-mail: [xuxiang8@mail2.sysu.edu.cn](mailto:xuxiang8@mail2.sysu.edu.cn)

A. Plaza

Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, 10003 Cáceres, Spain

e-mail: [aplaza@unex.es](mailto:aplaza@unex.es)

© Springer Nature Switzerland AG 2020

S. Prasad and J. Chanussot (eds.), *Hyperspectral Image Analysis*,

Advances in Computer Vision and Pattern Recognition,

[https://doi.org/10.1007/978-3-030-38617-7\\_13](https://doi.org/10.1007/978-3-030-38617-7_13)

tral unmixing purposes, focusing on a new fully unsupervised deep auto-encoder network (DAEN) method. Our experiments with simulated and real hyperspectral datasets demonstrate the competitive advantages of these innovative approaches over some well-established unmixing methods, revealing that these methods are currently at the forefront of hyperspectral unmixing.

## 13.1 Introduction

### 13.1.1 Spectral Unmixing

Hyperspectral remote sensing sensors collect spectral information from the Earth's surface using hundreds of narrow and contiguous wavelength bands [1]. It has been widely applied in various fields, such as target detection, material mapping, and material identification [2]. However, due to insufficient spatial resolution and spatial complexity, pixels in remotely sensed hyperspectral images are likely to be formed by a mixture of pure spectral constituents (*endmembers*) rather than a single substance [3]. The existence of mixed pixels complicates the exploitation of hyperspectral images [4]. Spectral unmixing, aimed at estimating the fractional abundance of the pure spectral signatures or endmembers, was proposed to deal with the problem of spectral mixing and effectively identifies the components of the mixed spectra in each pixel [5].

Unmixing algorithms rely on specific mixing models, which can be characterized as either linear or nonlinear [5, 6]. On the one hand, the linear model assumes that the spectral response of a pixel is given by a linear combination of the endmembers present in the pixel. On the other hand, the nonlinear mixture model assumes that the incident radiation interacts with more than one component and is affected by multiple scattering effects [3, 7]. As a result, nonlinear unmixing generally requires prior knowledge about object geometry and the physical properties of the observed objects [8]. The linear mixture model exhibits practical advantages, such as ease of implementation and flexibility in different applications. In this chapter, we will focus exclusively on the linear mixture model.

Under the linear mixture model, a group of unmixing approaches has been proposed [9–13]. Depending on whether a spectral library is available or not, we classify these methods into two categories, i.e., unsupervised and semi-supervised unmixing algorithms. With the wide availability of spectral libraries, sparse unmixing [8], as a semi-supervised approach in which mixed pixels are expressed in the form of combinations of a number of pure spectral signatures from a large spectral library, is able to handle the drawbacks introduced by such *virtual* endmembers and the unavailability of pure pixels.

### 13.1.2 Sparse Unmixing

The sparse unmixing approach exhibits significant advantages over unsupervised approaches, as it does not need to extract endmembers from the hyperspectral data or estimate the number of endmembers. Another advantage of sparse unmixing is that it provides great potential for accurate estimation of the fractional abundances, as all endmembers are normally represented in the library. However, these algorithms fully rely on the availability of a library *in advance*, and hence their semi-supervised nature.

The success of sparse unmixing relies on the fact that the unmixing solution is sparse, as the number of endmembers used to represent a mixed pixel is generally much smaller than the number of spectral signatures in the library [8]. As a result, new algorithms have been developed to enforce the sparsity on the solution. The sparse unmixing algorithm via variable splitting and augmented Lagrangian (SUnSAL) [8] adopts the  $\ell_1$  regularizer on the abundance matrix, which aims at introducing sparsity through the spectral domain that a pixel is unlikely to be mixed by a high number of components. The introduction of SUnSAL brought new insights into the concept of sparse unmixing. However, the real degree of sparsity is beyond the reach of the  $\ell_1$  regularizer due to the imbalance between the number of endmembers in the library and the number of components that generally participate in a mixed pixel. New algorithms have been developed in order to perform a better characterization of the degree of sparsity. Some techniques have focused on the introduction of new orders over the sparse regularizer such as the collaborative SUnSAL (CLSUnSAL) algorithm [14] and the graph-regularized  $\ell_{1/2}$ -NMF (GLNMF) method [15]. Other algorithms have introduced weighting factors to penalize the nonzero coefficients on the sparse solution [16], such as the reweighted sparse unmixing method [17] and the double reweighted sparse unmixing (DRSU) algorithm [18]. Although these methods obtained promising results, they consider pixels in a hyperspectral data as independent entities, and the spatial–contextual information in the hyperspectral image is generally disregarded. Since hyperspectral images generally follow specific spatial arrangements by nature, it is important to consider spatial information for their characterization [19].

Following this observation, several algorithms have focused on incorporating spatial correlation into the final solution. For instance, the sparse unmixing via variable splitting augmented Lagrangian and total variation (SUnSAL-TV) [20] represents one of the first attempts to include spatial information in sparse unmixing. It exploits the spatial information via a first-order pixel neighborhood system. Similar to SUnSAL, SUnSAL-TV opened new avenues and brought new insights into the concept of spatial sparse unmixing, which is able to promote piece-wise transitions in the estimated abundances. However, its performance strongly relies on the parameter settings [21]. At the same time, its model complexity results in a heavy computational cost, further limiting its practical application potential. New developments aimed at fully exploiting the spatial correlation among image features (and further imposing sparsity on the abundance matrix) have been mainly developed along two

directions. High-order neighborhood information over spatial regularizers has been introduced to reach this goal. For instance, the nonlocal sparse unmixing (NLSU) algorithm [22] can take advantage of high-order structural information. However, the neighborhood of the pixel changes randomly, thus limiting the continuity of spectral information. Another drawback of NLSU is that its model is more complex than that of SUnSAL-TV, which limits its practical application. Spatially weighted factors (aimed at characterizing spatial information through the inclusion of a weight on the sparse regularizer) have also been used to account for the spatial information in sparse unmixing. For example, the local collaborative sparse unmixing (LCSU) uses a spatial weight to impose local collaborativity, thus addressing some of the issues observed in SUnSAL-TV (including oversmoothed boundaries and blurred abundance maps) [23]. With similar complexity as the SUnSAL-TV, the LCSU exhibits similar unmixing performance as the SUnSAL-TV. This indicates that using spatial weights (as compared to spatial regularizers) has good potential in terms of improved unmixing performance and computational complexity. In [24], the spectral–spatial weighted sparse unmixing ( $S^2$ WSU) is proposed, which simultaneously exploits the spectral and spatial information contained in hyperspectral images via weighting factors, aiming at enhancing the sparsity of the solution. As a framework, the  $S^2$ WSU algorithm with its open structure, it is able to accept multiple types of spectral and spatial weighting factors, thus providing great flexibility for the exploration of different spatial scenarios, such as edge information, nonlocal similarity, homogeneous neighborhood information, etc.

### ***13.1.3 Deep Learning for Spectral Unmixing***

With advances in computer technology, learning-based approaches for unmixing have achieved a fast development in the past few years. Joint Bayesian unmixing is a typical example of learning-based approaches, which leads to good abundance estimates due to the incorporation of a full additivity (i.e., sum-to-one) and nonnegativity constraints [25–27]. Approaches based on artificial neural networks (ANNs) have also been developed for the learning of abundance fractions, assuming the prior knowledge of the endmember signatures [28–30]. These approaches exhibit better performance when compared with handcrafted methods, but they assume that endmembers are known in advance and, therefore, need to incorporate endmember extraction algorithms to perform unmixing. More recently, as a common tool for deep learning, auto-encoders have achieved a fast development in unmixing applications. Nonnegative sparse auto-encoder (NNSAE) and denoising auto-encoder were employed to obtain the endmember signatures and abundance fractions simultaneously for unmixing, with advanced denoising and intrinsic self-adaptation capabilities [31–33]. However, their strength lies in the aspect of noise reduction and they exhibit limitations when dealing with outliers. Due to the fact that outliers likely lead to initialization problems, their presence can bring strong interference to the unmixing solutions. In [34], a stacked nonnegative sparse auto-encoder (SNSA) is

proposed to address the issue of outliers. For linear mixing model (LSM)-based hyperspectral unmixing, the physical meaning of the model implies the sum-to-one on abundance fractions when every material in a pixel can be identified [3, 33, 35]. However, similar to the NMF-based approaches, SNSA adopts an additivity penalty on the abundance coefficients. The additivity penalty denotes that a penalty coefficient is used for controlling approximation of the sum-to-one. As this is not a hard constraint, the sum-to-one constraint is not necessarily ensured [34].

In [36], the fully unsupervised deep auto-encoder network (DAEN) unmixing method was recently proposed to address the presence of outliers in hyperspectral data. The DAEN has two main steps. In the first step, the spectral features are learned by the stacked auto-encoders (SAEs), aiming at generating good initializations for the network. In the second step, it employs a variational auto-encoder (VAE) to perform unmixing for the estimation of the endmembers and abundances. VAE combines variational inference to perform unsupervised learning and inherit auto-encoder architecture which can be trained with gradient descent [37]. Different from conventional auto-encoders, VAEs include a reparameterization which strictly ensures the abundance sum-to-one constraint during unmixing. Compared with other NMF-based algorithms, the DAEN has three main advantages: (1) with the use of SAEs, it can effectively tackle the problem of outliers and generate a good initialization of the unmixing network; (2) with the adoption of a VAE, it can ensure the nonnegativity and sum-to-one constraints, resulting in the good performance on abundance estimation; and (3) the endmember signatures and abundance fractions are obtained simultaneously. We emphasize the fully unsupervised nature of DAEN as one of its most powerful features.

### ***13.1.4 Contributions of This Chapter***

In this chapter, we focus on two types of techniques that are currently at the forefront of spectral unmixing. First, we provide an overview of advances in sparse unmixing algorithms, which can improve over traditional sparse unmixing algorithms by including spatial-contextual information that is crucial for a better scene interpretation. As these algorithms are semi-supervised and dependent on a library, we then describe new developments in the use of deep learning to perform spectral unmixing in fully unsupervised fashion, focusing on the DAEN method. Our experiments with simulated and real hyperspectral datasets demonstrate the competitive advantages of these innovative approaches over some well-established unmixing methods.

The remainder of this paper is organized as follows. The principles of sparse unmixing theory are presented in Sect. 13.2. The DAEN unmixing method is described in detail in Sect. 13.3. Section 13.4 describes several experiments to evaluate sparse unmixing algorithms. Section 13.5 describes several experiments to evaluate the DAEN algorithm. Finally, Sect. 13.6 concludes with some remarks and hints at plausible future research lines.

## 13.2 Sparse Unmixing Techniques

### 13.2.1 Sparse Versus Spectral Unmixing

The linear mixture model assumes that the spectral response of a pixel in any given spectral band is a linear combination of all of the endmembers present in the pixel at the respective spectral band. For each pixel, the linear model can be written as follows:

$$\begin{aligned} \mathbf{y} &= \mathbf{M}\boldsymbol{\alpha} + \mathbf{n} \\ \text{s.t.: } \boldsymbol{\alpha}_j &\geq 0, \quad \sum_{j=1}^q \boldsymbol{\alpha}_j = 1, \end{aligned} \tag{13.1}$$

where  $\mathbf{y}$  is a  $d \times 1$  column vector (the measured spectrum of the pixel),  $d$  denotes the number of bands.  $\mathbf{M}$  is a  $d \times q$  matrix containing  $q$  pure spectral signatures (endmembers),  $\boldsymbol{\alpha}$  is a  $q \times 1$  vector containing the fractional abundances of the endmembers, and  $\mathbf{n}$  is a  $d \times 1$  vector collecting the errors affecting the measurements at each spectral band. The so-called abundance nonnegativity constraint (ANC) ( $\boldsymbol{\alpha}_j \geq 0$  for ( $j = 1, 2, \dots, q$ )) and the abundance sum-to-one constraint (ASC) ( $\sum_{j=1}^q \boldsymbol{\alpha}_j = 1$ ).

Sparse unmixing reformulates (13.1) assuming the availability of a library of spectral signatures a priori as follows:

$$\mathbf{y} = \mathbf{A}\mathbf{h} + \mathbf{n}, \tag{13.2}$$

where  $\mathbf{h} \in \mathbb{R}^{m \times 1}$  is the fractional abundance vector compatible with spectral library  $\mathbf{A} \in \mathbb{R}^{d \times m}$  and  $m$  is the number of spectral signatures in  $\mathbf{A}$ .

Assuming that the dataset contains  $n$  pixels organized in the matrix  $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_n] \in \mathbb{R}^{d \times n}$  we may write then

$$\mathbf{Y} = \mathbf{A}\mathbf{H} + \mathbf{N} \quad \text{s.t.:} \quad \mathbf{H} \geq 0, \tag{13.3}$$

where  $\mathbf{N} = [\mathbf{n}_1, \dots, \mathbf{n}_n] \in \mathbb{R}^{d \times n}$  is the error.  $\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_n] \in \mathbb{R}^{m \times n}$  denotes the abundance maps corresponding to library  $\mathbf{A}$  for the observed data  $\mathbf{Y}$ , and  $\mathbf{H} \geq 0$  is the so-called abundance nonnegativity constraint (ANC). It should be noted that we explicitly enforce the ANC constraint without the abundance sum-to-one constraint (ASC), due to some criticisms about the ASC in the literature [8].

As the number of endmembers involved in a mixed pixel is usually very small when compared with the size of the spectral library, the abundance matrix  $\mathbf{H}$  is sparse. With these considerations in mind, the unmixing problem can be formulated as an  $\ell_2 - \ell_0$  optimization problem,

$$\min_{\mathbf{H}} \frac{1}{2} \|\mathbf{A}\mathbf{H} - \mathbf{Y}\|_F^2 + \lambda \|\mathbf{H}\|_0 \quad \text{s.t.:} \quad \mathbf{H} \geq 0, \tag{13.4}$$

where  $\|\cdot\|_F$  is the Frobenius norm and  $\lambda$  is a regularization parameter. Problem (13.4) is nonconvex and difficult to solve [38, 39]. The SUnSAL alternatively uses the  $\ell_2 - \ell_1$  norm to replace the  $\ell_2 - \ell_0$  norm and solves the unmixing problem as follows [40]:

$$\min_{\mathbf{H}} \frac{1}{2} \|\mathbf{A}\mathbf{H} - \mathbf{Y}\|_F^2 + \lambda \|\mathbf{H}\|_{1,1} \quad \text{s.t.:} \quad \mathbf{H} \geq 0, \quad (13.5)$$

where  $\|\mathbf{H}\|_{1,1} = \sum_{i=1}^n \|\mathbf{h}_i\|_1$  with  $\mathbf{h}_i$  ( $i = 1, \dots, n$ ) being the  $i$ th column of  $\mathbf{H}$ . SUnSAL solves the optimization problem in (13.5) efficiently using the ADMM [40]. However, as stated before, the real degree of sparsity is generally beyond the reach of the  $\ell_1$  regularizer.

### 13.2.2 Collaborative Regularization

Similar to (13.5), in [14], an  $\ell_{2,1}$  mixed norm (called collaborative regularization) was proposed, which globally imposes sparsity among the endmembers in collaborative fashion for all pixels. According to the collaborative sparse unmixing model described in [14], the objective function can be defined as follows:

$$\min_{\mathbf{H}} \frac{1}{2} \|\mathbf{A}\mathbf{H} - \mathbf{Y}\|_F^2 + \lambda \sum_{k=1}^m \|\mathbf{h}^k\|_2 \quad \text{s.t.} \quad \mathbf{h} \geq 0, \quad (13.6)$$

where  $\mathbf{h}^k$  denotes the  $k$ -th line of  $\mathbf{H}$  ( $k = 1, 2, \dots, m$ ) and  $\sum_{k=1}^m \|\mathbf{h}^k\|_2$  is the so-called  $\ell_{2,1}$  mixed norm. Note that the main difference between SUnSAL and CLSUnSAL is that the former employs pixel-wise independent regressions, while the latter enforces joint sparsity among all the pixels.

### 13.2.3 Total Variation Regularization

In order to take into account the spatial information of the image, a total variation (TV) regularizer can be integrated with SUnSAL (called SUnSAL-TV) to promote spatial homogeneity among neighboring pixels [20]:

$$\begin{aligned} \min_{\mathbf{H}} \quad & \frac{1}{2} \|\mathbf{A}\mathbf{H} - \mathbf{Y}\|_F^2 + \lambda \|\mathbf{H}\|_{1,1} + \lambda_{TV} TV(\mathbf{H}) \\ \text{s.t.:} \quad & \mathbf{H} \geq 0, \end{aligned} \quad (13.7)$$

where  $TV(\mathbf{H}) \equiv \sum_{\{k,i\} \in \mathcal{N}} \|\mathbf{h}_k - \mathbf{h}_i\|_1$ ,  $\mathcal{N}$  represents the set of (horizontal and vertical) pixel neighbors in the image, and  $\mathbf{h}_k$  denotes a series of the neighboring pixels of  $\mathbf{h}_i$  in abundance matrix  $\mathbf{H}$ . SUnSAL-TV shows great potential to exploit the spatial information for sparse unmixing. However, it may lead to oversmoothness and blurred boundaries.

### 13.2.4 Local Collaborative Regularization

In [23], the proposed LCSU assumes that endmembers tend to appear localized in spatially homogeneous areas instead of distributed over the full image. The proposed approach can also preserve global collaborativity (e.g., in the case that an endmember appears in the whole image), since it generalizes to global collaborativity through local searching:

$$\begin{aligned} \min_{\mathbf{h}} \quad & \frac{1}{2} \|\mathbf{A}\mathbf{H} - \mathbf{Y}\|_F^2 + \lambda \sum_{k=1}^m \sum_{i=1}^n \|\mathbf{h}_{x \in \mathcal{N}(i)}^k\|_2 \\ \text{s.t.} \quad & \mathbf{h} \geq 0, \end{aligned} \quad (13.8)$$

where  $\mathbf{h}^k$  denotes the  $k$ th line of matrix  $\mathbf{H}$  ( $k = 1, 2, \dots, m$ ),  $\sum_{k=1}^m \|\mathbf{h}^k\|_2$  is the so-called  $\ell_{2,1}$  mixed norm,  $\mathcal{N}(i)$  is the neighborhood of pixel  $i$  ( $i = 1, 2, \dots, n$ ), and  $\lambda$  is a regularization parameter controlling the degree of sparseness. The main difference between the proposed approach and SUnSAL-TV is that LCSU imposes collaborative sparsity among neighboring pixels, while SUnSAL-TV aims at promoting piece-wise smooth transitions in abundance estimations. In other words, SUnSAL-TV enforces that neighboring pixels share similar fractional abundances for the same endmember, while LCSU focuses on imposing local collaborativity among the full set of endmembers, thus addressing problems observed in SUnSAL-TV such as oversmoothed or blurred abundance maps. The main difference between problem (13.8) and problem (13.6) is that LCSU introduces spatial information to promote *local* collaborativity, while CLSUnSAL focuses on *global* collaborativity. In comparison with CLSUnSAL, the proposed LCSU assumes that neighboring pixels share the same support. This is more realistic, as a given endmember is likely to appear localized in a spatially homogeneous region rather than in the whole image.

### 13.2.5 Double Reweighted Regularization

Inspired by the success of weighted  $\ell_1$  minimization in sparse signal recovery, the double reweighted sparse unmixing and total variation (DRSU-TV) [41] was proposed to simultaneously exploit the spectral dual sparsity as well as the spatial smoothness of fractional abundances, as follows:

$$\min_{\mathbf{H}} \quad \frac{1}{2} \|\mathbf{A}\mathbf{H} - \mathbf{Y}\|_F^2 + \lambda \|(\mathbf{W}_{\text{spe2}} \mathbf{W}_{\text{spe1}}) \odot \mathbf{H}\|_{1,1} + \lambda_{\text{TV}} \text{TV}(\mathbf{H}), \quad \text{s.t.:} \quad \mathbf{H} \geq 0, \quad (13.9)$$

where the operator  $\odot$  denotes the element-wise multiplication of two variables. The first regularizer  $\lambda \|(\mathbf{W}_{\text{spe2}} \mathbf{W}_{\text{spe1}}) \odot \mathbf{X}\|_{1,1}$  introduces a prior with spectral sparsity, where  $\lambda$  is the regularization parameter,  $\mathbf{W}_{\text{spe1}} = \{w_{\text{spe1},ki} | k = 1, \dots, m, i = 1, \dots, n\} \in \mathbb{R}^{m \times n}$  and  $\mathbf{W}_{\text{spe2}} = \text{diag}(w_{\text{spe2},11}, \dots, w_{\text{spe2},kk}, \dots, w_{\text{spe2},mm}) \in \mathbb{R}^{m \times m}$ ,



for  $k = 1, \dots, m$ , are the dual weights, with  $\mathbf{W}_{\text{spe1}}$  being the original weight introduced in [16] aimed at penalizing the nonzero coefficients on the solution and  $\mathbf{W}_{\text{spe2}}$  promoting nonzero row vectors. The latter regularizer  $\lambda_{\text{TV}}\text{TV}(\mathbf{H})$  exploits the spatial prior with  $\lambda_{\text{TV}}$  being the parameter controlling the degree of smoothness. It can be seen that DRSU-TV incorporates a TV-based regularizer to enforce the spatial smoothness of abundances compared to DRSU.

In [41], Problem (13.9) is optimized via ADMM under an iterative scheme. The dual weights  $\mathbf{W}_{\text{spe1}}$  and  $\mathbf{W}_{\text{spe2}}$  are updated as follows, at iteration  $t + 1$ :

$$w_{\text{spe1},ki}^{(t+1)} = \frac{1}{h_{ki}^{(t)} + \varepsilon}, \quad (13.10)$$

where  $\varepsilon > 0$  is a small positive value and

$$w_{\text{spe2},kk}^{(t+1)} = \frac{1}{\|\mathbf{H}^{(t)}(k, :)\|_2 + \varepsilon}, \quad (13.11)$$

where  $\mathbf{H}^{(t)}(k, :)$  is the  $k$ th row in the estimated abundance of the  $t$ th iteration. Notice that, as shown in (13.10) and (13.11), it is suggested that large weights be used to discourage nonzero entries in the recovered signal, while small weights encourage nonzero entries. DRSU-TV, exploiting the spectral and spatial priors simultaneously under the sparse unmixing model, exhibits good potential in comparison with the  $\ell_1$ - or TV-based methods. However, as an adaptation of the  $\ell_1$ - and TV-based approach, the limitations of DRSU-TV are associated with the use of a regularizer-based spatial prior. That is, the computational complexity is similar to that of SUnSAL-TV. Such high computational complexity constrains the practical applications of DRSU-TV. Furthermore, the unmixing performance of the method is sensitive to the regularization parameter  $\lambda_{\text{TV}}$ .

### 13.2.6 Spectral–Spatial Weighted Regularization

In [24], the  $\text{S}^2\text{WSU}$  algorithm is developed, which aims at exploiting the spatial information more efficiently for sparse unmixing purposes. As opposed to the approaches that exploit a regularizer-based spatial prior (which have one additional parameter for the spatial regularizer and often exhibit high complexity), the  $\text{S}^2\text{WSU}$  algorithm includes the spatial correlation via a weighting factor, resulting in good computational efficiency and less regularization parameters. Let  $\mathbf{W}_{\text{spe}} \in \mathbb{R}^{m \times m}$  be the spectral weighting matrix and  $\mathbf{W}_{\text{spa}} \in \mathbb{R}^{m \times n}$  be the spatial one. Following [16], the objective function of the  $\text{S}^2\text{WSU}$  is given as follows:

$$\min_{\mathbf{H}} \frac{1}{2} \|\mathbf{A}\mathbf{H} - \mathbf{Y}\|_F^2 + \lambda \|\mathbf{W}_{\text{spe}}\mathbf{W}_{\text{spa}}\| \odot \mathbf{H}\|_{1,1}, \quad \text{s.t.}: \mathbf{H} \geq 0. \quad (13.12)$$

For the spectral weighting factor  $\mathbf{W}_{\text{spe}}$ , relying on the success of [14, 18], it adopts row collaborativity to enforce joint sparsity among all the pixels. Similar to the  $\mathbf{W}_{\text{spe}2}$  in DRSU-TV, the  $\mathbf{W}_{\text{spe}}$  aims at enhancing the sparsity of the endmembers in the spectral library. In detail, at iteration  $t + 1$ , it can be updated as

$$\mathbf{W}_{\text{spe}}^{(t+1)} = \text{diag} \left[ \frac{1}{\|\mathbf{H}^{(t)}(1, :)\|_2 + \varepsilon}, \dots, \frac{1}{\|\mathbf{H}^{(t)}(m, :)\|_2 + \varepsilon} \right]. \quad (13.13)$$

For the spatial weighting factor  $\mathbf{W}_{\text{spa}}$ , let  $w_{\text{spa},ki}^{(t+1)}$  be the element of the  $k$ th line and  $i$ th row in  $\mathbf{W}_{\text{spa}}$  at iteration  $t + 1$ , it incorporates the neighboring information as follows:

$$w_{\text{spa},ki}^{(t+1)} = \frac{1}{f_{x \in \mathcal{N}(i)}(h_{kx}^{(t)}) + \varepsilon}, \quad (13.14)$$

where  $\mathcal{N}(i)$  denotes the neighboring set for element  $h_{ki}$ , and  $f(\cdot)$  is a function explicitly exploiting the spatial correlations through the neighborhood system. It uses the neighboring coverage and importance to incorporate the spatial correlation as follows:

$$f(h_{ki}) = \frac{\sum_{x \in \mathcal{N}(i)} \theta_{kx} h_{kx}}{\sum_{x \in \mathcal{N}(i)} \theta_{kx}}, \quad (13.15)$$

where  $\mathcal{N}(i)$  corresponds to the neighboring coverage and  $\theta$  represents the neighborhood importance. It considers the 8-connected ( $3 \times 3$  window) for algorithm design and experiments. With respect to the neighboring importance, for any two entries  $k$  and  $i$ , we compute it as follows:

$$\theta_{ki} = \frac{1}{\text{im}(k, i)}, \quad (13.16)$$

where function  $\text{im}(\cdot)$  is the important measurement over the two elements  $h_k$  and  $h_i$ . Let  $(a, b)$  and  $(c, d)$  be the spatial coordinates of  $h_k$  and  $h_i$ . The European distance is specifically considered, that is,  $\theta_{ki} = 1/\sqrt{(a-c)^2 + (b-d)^2}$ .

It should be noted that the optimization problem of  $S^2\text{WSU}$  can be iteratively solved by an outer-inner looping scheme, where the inner loop updates the unmixing coefficients via ADMM and the outer loop updates the spectral and spatial weights [24].

### 13.3 Deep Learning for Hyperspectral Unmixing

As one of the very few unsupervised approaches available, the deep auto-encoder network (DAEN) unmixing method specifically addresses the presence of outliers in hyperspectral data [36]. In the following subsections, we describe the different

processing modules that compose this promising approach for deep hyperspectral unmixing.

### 13.3.1 NMF-Based Unmixing

Let  $\mathbf{Y} \equiv [\mathbf{y}_1, \dots, \mathbf{y}_n] \in \mathbb{R}^{d \times n}$  be matrix representation of a hyperspectral dataset with  $n$  spectral vectors and  $d$  spectral bands. Under the linear mixing model, we have [3, 42]

$$\begin{aligned} \mathbf{Y} &= \mathbf{W}\mathbf{H} + \mathbf{N} \\ \text{s.t.: } \mathbf{H} &\geq 0, \mathbf{1}_m^T \mathbf{H} = \mathbf{1}_n^T, \end{aligned} \quad (13.17)$$

where  $\mathbf{W} \equiv [\mathbf{w}_1, \dots, \mathbf{w}_m] \in \mathbb{R}^{d \times m}$  is the mixing matrix containing  $m$  endmembers,  $\mathbf{w}_i$  denotes the  $i$ th endmember,  $\mathbf{H} \geq 0$  and  $\mathbf{1}_m^T \mathbf{H} = \mathbf{1}_n^T$  are the so-called abundance nonnegativity and sum-to-one constraints, which stem from a physical interpretation of the abundance vectors, and  $\mathbf{1}_m = [1, 1, \dots, 1]^T$  is a column vector of size  $m$  (the notation  $[\cdot]^T$  stands for vector or matrix transpose). Finally,  $\mathbf{N} \in \mathbb{R}^{d \times n}$  is the error matrix that may affect the measurement process (e.g., noise). It should be noted that the symbol naming in this section is not the same as the naming in Sect. 13.2. In addition, we have a detailed description of each symbol.

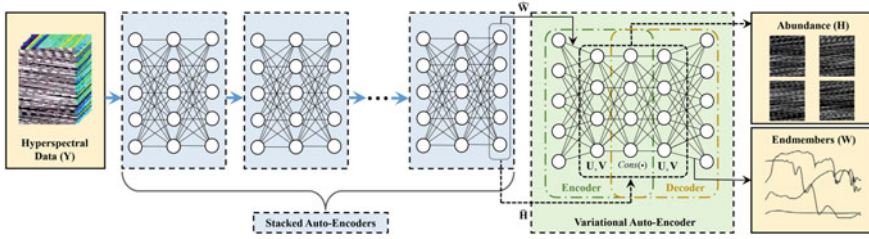
For a given observation  $\mathbf{Y}$ , unmixing aims at obtaining the mixing matrix  $\mathbf{W}$  and the abundance matrix  $\mathbf{H}$ . In this work, we tackle the simultaneous estimation of  $\mathbf{W}$  and  $\mathbf{H}$  by seeking a solution with the following NMF-based optimization:

$$(\mathbf{W}, \mathbf{H}) = \arg \min_{\mathbf{W}, \mathbf{H}} \frac{1}{2} \|\mathbf{Y} - \mathbf{W}\mathbf{H}\|_F^2 + \mu f_1(\mathbf{W}) + \lambda f_2(\mathbf{H}), \quad (13.18)$$

where  $\|\cdot\|_F^2$  denotes the Frobenius norm,  $f_1(\mathbf{W})$  and  $f_2(\mathbf{H})$  are two regularizers on the mixing matrix  $\mathbf{W}$  and the abundance fractions  $\mathbf{H}$ , respectively, with  $\mu$  and  $\lambda$  being the regularization parameters.

### 13.3.2 Deep Auto-Encoder Network

In this section, the DAEN unmixing method [36] is described (illustrated in Fig. 13.1), where  $\mathbf{U}$  and  $\mathbf{V}$  are the latent variables (LV) of the reparameterization of the VAE, respectively. As shown in Fig. 13.1, the endmember matrix  $\mathbf{W}$  corresponds to the last weight matrix of the decoder in VAE, and the abundance  $\mathbf{H}$  is estimated from the hidden layers of VAE, while  $\hat{\mathbf{W}}$  and  $\hat{\mathbf{H}}$  denote the initializations for VAE generated by SAEs, respectively.



**Fig. 13.1** The flowchart of the proposed DAEN, which includes two parts, i.e., stacked auto-encoders (SAEs) and a variational auto-encoder (VAE). The stacked auto-encoders (SAEs) generate the initializations  $\hat{\mathbf{W}}$  and  $\hat{\mathbf{H}}$  for the VAE, while the VAE performs the NMF-based unmixing aiming at obtaining the endmembers  $\mathbf{W}$  and abundances  $\mathbf{H}$ , respectively

### 13.3.3 Stacked Auto-Encoders for Initialization

Based on the geometry assumption that endmembers are generally located around the vertices of the data simplex, we use a pure pixel-based method to extract a set of candidate pixels as the training set for the SAEs. Specifically, we adopt VCA to obtain a set of  $k$  candidates, with  $k > m$ . As VCA considers random directions in the subspace projection [3, 43], we run it for  $p$  times, resulting in  $q$  candidates, with  $q = p \cdot k$ . These  $q$  candidates are then grouped into  $m$  training sets  $\{\mathbf{C}_i\}_{i=1}^m$  based on the spectral angle distance (SAD) and clustering, with  $\mathbf{C}_i = [\mathbf{c}_1, \dots, \mathbf{c}_{i_n}] \in \mathbb{R}^{d \times i_n}$  and  $i_n$  is the number of samples in  $\mathbf{C}_i$ . Let  $\mathbf{c}_{i_o}$  and  $\mathbf{c}_{j_o}$  be the cluster centers of  $\mathbf{C}_i$  and  $\mathbf{C}_j$ , respectively. For any candidate  $\mathbf{c}_{i_s}$  in  $\mathbf{C}_i$ , for  $i_s = 1, \dots, i_n$ , we have  $\text{SAD}(\mathbf{c}_{i_o}, \mathbf{c}_{i_s}) \leq \text{SAD}(\mathbf{c}_{j_o}, \mathbf{c}_{i_s})$ , for any  $j = 1, \dots, m$  and  $j \neq i$ , where

$$\text{SAD}(\mathbf{c}_{i_o}, \mathbf{c}_{i_s}) = \arccos \left( \frac{[\mathbf{c}_{i_o}, \mathbf{c}_{i_s}]}{\|\mathbf{c}_{i_o}\| \cdot \|\mathbf{c}_{i_s}\|} \right). \tag{13.19}$$

In this work, for  $p$  and  $k$ , we empirically set  $p = 30$  and  $k = 3m$ , respectively. By enforcing nonnegativity, the training of SAEs minimizes the reconstruction error as follows:

$$\min \sum_{s=1}^{i_n} \|\mathbf{c}_s - \hat{\mathbf{w}}_i\|_2^2, \tag{13.20}$$

where  $\hat{\mathbf{w}}_i$  is the reconstructed signature of the  $i$ th endmember and  $\hat{\mathbf{W}} = [\hat{\mathbf{w}}_1, \dots, \hat{\mathbf{w}}_m]$  are the reconstructed endmember matrix. Following [44], the reconstructed signature is denoted as

$$\hat{\mathbf{w}}_i = \mathbf{M}_i f(\mathbf{M}_i^T \mathbf{C}_i), \tag{13.21}$$

where  $\mathbf{M}_i$  is the matrix of weights between the input and hidden neurons or those from hidden to output neurons, and  $f(\cdot)$  is the activation function [44] given by

$$f(\mathbf{g}_i) = \frac{1}{1 + \exp(-\mathbf{a}_i \cdot * \mathbf{g}_i - \mathbf{b}_i)}, \quad (13.22)$$

where  $\mathbf{g}_i = \mathbf{M}_i^T \mathbf{C}_i$ ,  $\mathbf{a}_i$  and  $\mathbf{b}_i$  are parameters aimed at controlling the information transmission between neurons, and  $\cdot *$  is the dot product, i.e., element-wise operator. Notice that the number of input neurons and output neurons is the same as the hidden neurons, while the number of hidden neurons here is set as the number of bands. Then, we can use a gradient rule to update  $\mathbf{a}_i$  and  $\mathbf{b}_i$  as follows:

$$\begin{cases} \Delta \mathbf{a}_i = \gamma(1 - (2 + \frac{1}{\tau})f_i + \frac{1}{\tau}f_i^2), \\ \Delta \mathbf{b}_i = \gamma \frac{1}{\mathbf{b}_i} + \mathbf{g}_i \Delta \mathbf{a}_i, \end{cases} \quad (13.23)$$

where  $\gamma$  and  $\tau$  are hyper-parameters in the learning process controlling the mean activity level of the desired output distribution. Following the empirical settings in [44], we set  $\gamma = 0.0001$  and  $\tau = 0.2$ . With the aforementioned definition in hand, the learning reduces to the following update rule:

$$\Delta \mathbf{M}_i \leftarrow \eta \Delta \widehat{\mathbf{w}}_i f_i^T + |\mathbf{M}_i|, \quad (13.24)$$

where  $\Delta \widehat{\mathbf{w}}_i$  is the gradient of candidate  $i$  for update,  $|\mathbf{M}_i|$  enforces the weight matrix to be nonnegative, and  $\eta$  is an adaptive learning rate. In this work, following [44], we set  $\eta = \widehat{\eta}(\|f_i\|^2 + \epsilon)^{-1}$  with  $\widehat{\eta} = 0.002$ , where  $\epsilon = 0.001$  is a small parameter to ensure the positivity of  $\eta$ .

Finally, let  $\widehat{\mathbf{w}}_i^t, \widehat{\mathbf{w}}_i^{t+1}$  be the reconstructions from the  $t$ -th and  $(t+1)$ -th auto-encoders, respectively. The SAEs ends when  $\|\widehat{\mathbf{w}}_i^{t+1} - \widehat{\mathbf{w}}_i^t\|_2^2$  converges.

After the endmember matrix  $\widehat{\mathbf{W}}$  is reconstructed, based on the linear mixing model (13.17), the abundances  $\widehat{\mathbf{H}}$  can be obtained via the fully constrained least square (FCLS) [42]. In the learning of the VAE,  $\widehat{\mathbf{W}}$  and  $\widehat{\mathbf{H}}$  are used as initializations of  $\mathbf{W}$  and  $\mathbf{H}$ , respectively.

### 13.3.4 Variational Auto-Encoders for Unmixing

First, let us recall the NMF-based objective function in (13.18), which contains two regularizers on the mixing matrix and abundance matrix, respectively. For the first regularizer  $f_1(\mathbf{W})$  on the mixing matrix, following [11], we have

$$f_1(\mathbf{W}) = \text{MinVol}(\mathbf{W}), \quad (13.25)$$

where  $\text{MinVol}(\cdot)$  is a function aiming at enclosing all the pixels into the simplex constructed by the endmembers. Specifically, following [11], we set  $\text{MinVol}(\mathbf{W}) = \|\det(\mathbf{W})\|$ , with  $\|\det(\mathbf{W})\|$  being the volume defined by the origin and the columns of  $\mathbf{W}$ .

With respect to regularizer  $f_2(\mathbf{H})$  on the abundance matrix, in order to ensure the nonnegativity and sum-to-one constraints, we employ the variational auto-encoder (VAE) to penalize the solution of  $\mathbf{H}$ , denoted as

$$f_2(\mathbf{H}) = \text{VAE}(\mathbf{H}), \quad (13.26)$$

where the neurons of all hidden layers are set as the number of endmembers, while the number of inputs and outputs corresponds to the number of pixels.

With these definitions in mind, we obtain the following objective function:

$$(\mathbf{W}, \mathbf{H}) = \arg \min_{\mathbf{W}, \mathbf{H}} \frac{1}{2} \|\mathbf{Y} - \mathbf{W}\mathbf{H}\|_F^2 + \mu \text{MinVol}(\mathbf{W}) + \lambda \text{VAE}(\mathbf{H}). \quad (13.27)$$

In the following, we present the VAE-based regularizer in detail. Let  $\mathbf{U}$  and  $\mathbf{V}$  be the LV, we define  $f_2(\mathbf{H})$  as

$$f_2(\mathbf{H}(\mathbf{U}, \mathbf{V})) = \left\| \frac{1}{2n} (\mathbf{1}_{m \times n} + \ln \mathbf{V}^2 - \mathbf{U}^2 - \mathbf{V}^2) \mathbf{1}_n \right\|_2^2, \quad (13.28)$$

where  $\mathbf{1}_{m \times n} \in \mathbb{R}^{m \times n}$  with all elements being 1, and vector  $\mathbf{1}_n = [1, \dots, 1]^T \in \mathbb{R}^n$ ,  $\mathbf{U} = \{\mathbf{u}_1, \dots, \mathbf{u}_n\} \in \mathbb{R}^{m \times n}$ ,  $\mathbf{V} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\} \in \mathbb{R}^{m \times n}$ . The derivation of (13.28) is shown in [36]. Following [37], let  $\mathbf{u}_j = [u_{1,j}, \dots, u_{m,j}]^T \in \mathbb{R}^m$  and  $\mathbf{v}_j = [v_{1,j}, \dots, v_{m,j}]^T \in \mathbb{R}^m$  be the reparameters of LV, we define  $h_{i,j} = \text{Cons}(u_{i,j}, v_{i,j})$ , where  $\text{Cons}(\cdot)$  represents a decay function as follows:

$$\text{Cons}(u_{i,j}, v_{i,j}) = \begin{cases} u_{i,j} + \sigma v_{i,j}, & 0 < (u_{i,j} + \sigma v_{i,j}) < 1 \\ 0, & \text{otherwise,} \end{cases} \quad (13.29)$$

where  $\sigma$  is a parameter that, as indicated in [25], can be obtained via Monte Carlo (MC) sampling. In order to meet the abundance sum-to-one constraint, we have

$$h_{m,j} = 1 - \sum_{i=1}^{m-1} h_{i,j}. \quad (13.30)$$

The objective function in (13.27) is a combinational problem, which is nonconvex, and therefore it is difficult to solve. In [36], it proposes an iterative scheme to optimize  $\mathbf{W}$  and  $\mathbf{H}$ , respectively, both of which are solved by a gradient descent method. The first-order derivatives of the objective function are computed as follows:

$$\begin{cases} \nabla_{\mathbf{U}}(\mathbf{W}, \mathbf{H}) = d(\mathbf{U}) - \frac{2\lambda}{n} \mathbf{z}(\mathbf{1}_n)^T \cdot * \mathbf{U}, \\ \nabla_{\mathbf{V}}(\mathbf{W}, \mathbf{H}) = d(\mathbf{V}) + \frac{2\lambda}{n} \mathbf{z}(\mathbf{1}_n)^T \cdot * (\ln \mathbf{V} ./ \mathbf{V} - \mathbf{V}), \end{cases} \quad (13.31)$$

where  $./$  is the dot division,  $\mathbf{z} = \frac{1}{2n}(\mathbf{1}_{m \times n} + \ln \mathbf{V}^2 - \mathbf{U}^2 - \mathbf{V}^2)\mathbf{1}_n$ .  $d(\mathbf{U})$  and  $d(\mathbf{V})$  are gradients of reconstructed errors, which are

$$\begin{cases} d(\mathbf{U}) = \mathbf{W}^T (\mathbf{W}\mathbf{H} - \mathbf{Y}) \cdot * \mathbb{C}_{cons}, \\ d(\mathbf{V}) = \sigma \mathbf{W}^T (\mathbf{W}\mathbf{H} - \mathbf{Y}) \cdot * \mathbb{C}_{cons}, \end{cases} \quad (13.32)$$

where  $\mathbb{C}_{cons}$  is an indicative function,  $\mathbb{C}_{cons} = \mathbf{1}_{m \times n} \{0 < (\mathbf{U} + \sigma \mathbf{V}) < 1\}$ . For more details, the derivation of (13.31) is given in [36].

---

### Algorithm 1 DAEN for hyperspectral unmixing

---

**Input:** dataset  $\mathbf{Y}$ .

**Output:** endmembers  $\mathbf{W}$ , abundances  $\mathbf{H}$ .

**Step 1.** */\* SAE for initialization \*/*

1. **Initialization:**  $\mathbf{M}_i$ .
2. Set hyper-parameters following [44].
3. Obtain  $p \times k$  candidates via VCA[43].

**repeat**

4. Update  $\{\widehat{\mathbf{w}}_i\}_{i=1}^m$  in (13.21).
5. Update  $\{\mathbf{M}_i\}_{i=1}^m$  in (13.24).

**until** convergence

6. Compute  $\widehat{\mathbf{H}}$  via FCLS[42].

**Step 2.** */\* VAE for unmixing \*/*

7. **Initialization:**  $\mathbf{U}$  and  $\mathbf{V}$ .

**repeat**

8. Update  $\Delta \mathbf{H}$  in (13.33).
9. Update  $\Delta \mathbf{W}$  in (13.34).

**until** convergence

---

With respect to the updates of  $\mathbf{H}$  and  $\mathbf{W}$ , we employ the gradient descent method for the solutions as follows:

$$\mathbf{H} \leftarrow \mathbf{H} + \Delta \mathbf{H},$$

and

$$\mathbf{W} \leftarrow \mathbf{W} + \Delta \mathbf{W},$$

where  $\Delta \mathbf{H}$  and  $\Delta \mathbf{W}$  are the gradients for  $\mathbf{H}$  and  $\mathbf{W}$ , respectively. Specifically,

- For  $\mathbf{H}$ , we have

$$\Delta \mathbf{H} = -\varphi(\nabla_{\mathbf{U}}(\mathbf{W}, \mathbf{H}) + \sigma \nabla_{\mathbf{V}}(\mathbf{W}, \mathbf{H})), \quad (13.33)$$

where  $\varphi$  is the learning rates that can be estimated by the Armijo rule [45].

- For  $\mathbf{W}$ , we obtain  $\Delta\mathbf{W}$  via Adadelta [46] as follows:

$$\Delta\mathbf{W} = -\frac{\text{RMS}[\Delta\mathbf{W}]}{\text{RMS}[\nabla\mathbf{W}(\mathbf{W}, \mathbf{H})]} \nabla\mathbf{W}(\mathbf{W}, \mathbf{H}), \quad (13.34)$$

where  $\text{RMS}[\cdot]$  is the root-mean-square [46]. The first-order derivatives of the objective function (13.27) are calculated as follows:

$$\nabla_{\mathbf{W}}(\mathbf{W}, \mathbf{H}) = (\mathbf{W}\mathbf{H} - \mathbf{Y})\mathbf{H}^T + \mu d(\text{MinVol}(\mathbf{W})). \quad (13.35)$$

where  $d(\text{MinVol}(\mathbf{W}))$  is the gradient for the volume function, which can be computed as the one in [47].

Finally, a pseudocode of the proposed DAEN is given in Algorithm 1. As shown in Algorithm 1, DAEN consists of two main parts, a set of SAEs for initialization and one VAE for unmixing. Specifically, in Line 1,  $\mathbf{M}_i$  is randomly initialized. In Line 2, the hyper-parameters are set following [44], while in Line 3, the candidate samples used for training are generated via VCA. In Lines 4 and 5,  $\{\widehat{\mathbf{w}}_i\}$  and  $\{\mathbf{M}_i\}$  are iteratively updated until SAE terminates. In Line 6, it computes the abundance estimation  $\widehat{\mathbf{H}}$  via FCLS. In Line 7, the LV variables,  $\mathbf{U}$  and  $\mathbf{V}$ , are randomly initialized. Finally, in Lines 8 and 9, the endmember matrix  $\mathbf{W}$  and the abundance matrix  $\mathbf{H}$  are iteratively updated, respectively.

### 13.4 Experiments and Analysis: Sparse Unmixing

In this section, we illustrate the unmixing performance of these sparse unmixing methods using simulated hyperspectral datasets. For quantitative analysis, the signal-to-reconstruction error (SRE, measured in dB) is used to evaluate the unmixing accuracy. For comparative purposes, the results obtained by SUnSAL [8], SUnSAL-TV [20], LCSU [23], DRSU [18], DRSU-TV [41], and S<sup>2</sup>WSU [24] algorithms are reported. Let  $\widehat{\mathbf{h}}$  be the estimated abundance, and  $\mathbf{h}$  be the true abundance. The SRE(dB) can be computed as follows:

$$\text{SRE(dB)} = 10 \cdot \log_{10}(E(\|\mathbf{h}\|_2^2)/E(\|\mathbf{h} - \widehat{\mathbf{h}}\|_2^2)), \quad (13.36)$$

where  $E(\cdot)$  denotes the expectation function. Furthermore, we use another indicator, i.e., the probability of success  $p_s$ , which is an estimate of the probability that the relative error power be smaller than a certain threshold. It is formally defined as follows:  $p_s \equiv P(\|\widehat{\mathbf{h}} - \mathbf{h}\|^2/\|\mathbf{h}\|^2 \leq \text{threshold})$ . In our case, the estimation result is considered successfully when  $\|\widehat{\mathbf{h}} - \mathbf{h}\|^2/\|\mathbf{h}\|^2 \leq 3.16$  (5 dB). This threshold was demonstrated to be appropriate in [8]. The larger the SRE (dB) or the  $p_s$ , the more accurate the unmixing results.



### 13.4.1 Simulated Datasets

The spectral library that we use in our synthetic image experiments is a dictionary of minerals extracted from the United States Geological Survey (USGS) library.<sup>1</sup> Such library, denoted by  $\mathbf{A}$ , contains  $m = 240$  materials (different mineral types), with spectral signatures with reflectance values consisting of  $L = 224$  spectral bands and distributed uniformly in the interval 0.4–2.5  $\mu\text{m}$ . Following the work in [20], simulated data cube is generated with  $100 \times 100$  pixels and nine spectral signatures (Adularia GDS57 Orthoclase, Jarosite GDS99 K Sy 200C, Jarosite GDS101 Na Sy 200, Anorthite HS349.3B, Calcite WS272, Alunite GDS83 Na63, Howlite GDS155, Corrensite CorWa-1, Fassaite HS118.3B.), which are randomly chosen from the spectral library  $\mathbf{A}$ . The fractional abundances are piece-wise smooth, i.e., they are smooth with sharp transitions; moreover, they are subject to the ANC and ASC. These data can reveal the spatial features quite well for the different unmixing algorithms. For illustrative purposes, Fig. 13.2 shows the true abundance maps of the endmembers. After generating the data cube, it was contaminated with i.i.d. Gaussian noise, for three levels of the signal-to-noise (SNR) ratio: 30, 40, and 50 dB.

Table 13.1 shows the SRE (dB) and  $p_s$  results achieved by the different tested algorithms under different SNR levels. For all the tested algorithms, the input parameters have been carefully tuned for optimal performance. From Table 13.1, we can see that the methods of using double weights (DRSU, DRSU-TV, and  $\text{S}^2\text{WSU}$ ) have obtained better SRE (dB) results than other algorithms in all cases. Furthermore, the  $\text{S}^2\text{WSU}$  achieved better SRE (dB) results than the competitors in all cases, which indicates that the inclusion of a spatial factor in the sparse regularizer can further promote the spatial correlation on the solution and improve the unmixing performance. The  $p_s$  obtained by the  $\text{S}^2\text{WSU}$  is also much better than those obtained by other algorithms in the case of low SNR values, which reveals that the inclusion of spatial information leads to high robustness. Based on the aforementioned results, we can conclude that the spatial weighted strategy offers the potential to improve sparse unmixing performance.

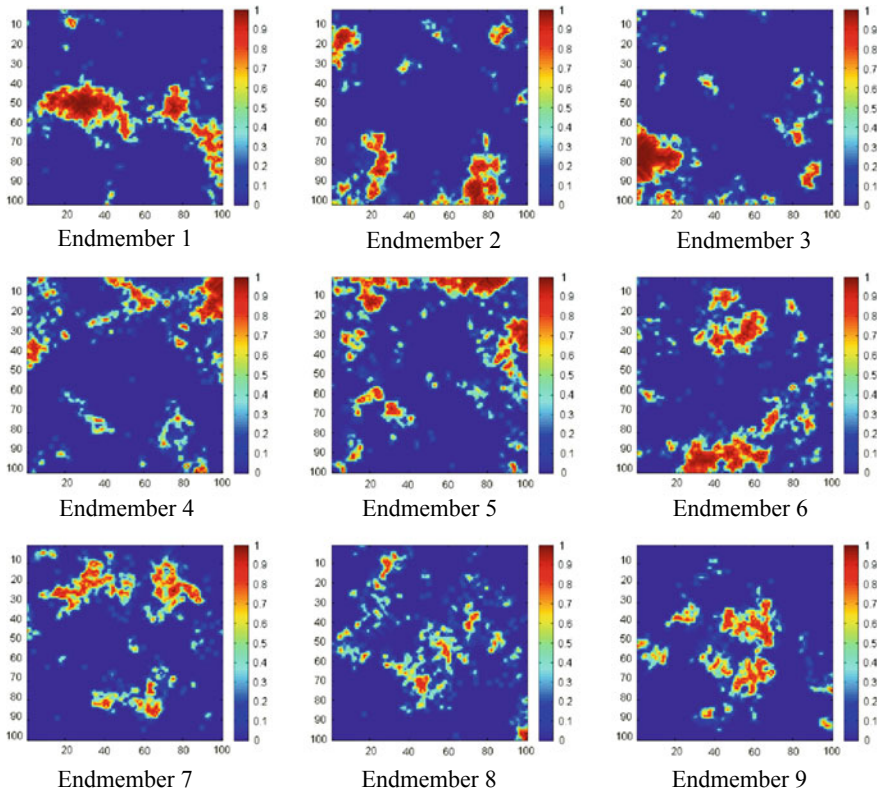
### 13.4.2 Real Hyperspectral Data

In this section, we resort to the well-known Airborne Visible Infrared Imaging Spectrometer (AVIRIS) Cuprite dataset for evaluation of the proposed approach, which is a common benchmark for validation of spectral unmixing algorithms. The data are available online in reflectance units.<sup>2</sup> The portion used in experiments corresponds to a  $350 \times 350$ -pixel subset of the scene, with 224 spectral bands in the range 0.4–2.5  $\mu\text{m}$  and nominal spectral resolution of 10 nm. Prior to the analysis, bands 1–2,

---

<sup>1</sup> Available online at <http://speclab.cr.usgs.gov/spectral.lib06>.

<sup>2</sup> <http://aviris.jpl.nasa.gov/html/aviris.freedata.html>.



**Fig. 13.2** True fractional abundances of the endmembers in the simulated data cube

105–115, 150–170, and 223–224 were removed due to water absorption and low SNR, leaving a total of 188 spectral bands. The spectral library used in this experiment is the same library **A** used in our simulated experiments and the noisy bands are also removed from **A**. The classification maps of these materials produced by Tricorder software<sup>3</sup> are also displayed. Figure 13.3 shows a mineral map produced in 1995 by USGS, in which the Tricorder 3.3 software product [48] was used to map different minerals present in the Cuprite mining district. The USGS map serves as a good indicator for qualitative assessment of the fractional abundance maps produced by the different unmixing algorithms. Note that the publicly available AVIRIS Cuprite data were collected in 1997 but the Tricorder map was produced in 1995. In addition, the true abundances of the real hyperspectral data are unavailable. Thus, we can only make a qualitative analysis of the performances of different sparse unmixing algorithms by comparing their estimated abundances with the mineral maps.

Figure 13.4 conducts a qualitative comparison between the classification maps produced by the USGS Tricorder algorithm and the fractional abundances estimated

<sup>3</sup><http://speclab.cr.usgs.gov/PAPER/tetracorder>.

**Table 13.1** SRE(dB) and  $p_s$  scores achieved after applying different unmixing methods to the simulated data Cube 1 (the optimal parameters for which the reported values were achieved are indicated in the parentheses)

Algorithm	SNR = 30 dB		SNR = 40 dB		SNR = 50 dB	
	SRE (dB)	$p_s$	SRE (dB)	$p_s$	SRE (dB)	$p_s$
SUnSAL	8.4373	0.7946	15.1721	0.9886	23.0894	1
	$(\lambda = 2e-2)$		$(\lambda = 5e-3)$		$(\lambda = 1e-3)$	
SUnSAL-TV	11.4304	0.9470	17.7695	0.9998	26.1655	1
	$(\lambda = 1e-2; \lambda_{TV} = 4e-3)$		$(\lambda = 5e-3; \lambda_{TV} = 1e-3)$		$(\lambda = 2e-3; \lambda_{TV} = 2e-4)$	
LCSU	11.4317	0.9463	18.1793	0.9999	26.2194	1
	$(\lambda = 3e-2)$		$(\lambda = 7e-3)$		$(\lambda = 1e-3)$	
DRSU	14.9876	0.9745	29.6861	1	41.1967	1
	$(\lambda = 3e-3)$		$(\lambda = 1e-3)$		$(\lambda = 6e-4)$	
DRSU-TV	18.8630	0.9994	30.9403	1	41.1967	1
	$(\lambda = 2e-3; \lambda_{TV} = 2e-3)$		$(\lambda = 2e-3; \lambda_{TV} = 4e-4)$		$(\lambda = 6e-4; \lambda_{TV} = 0)$	
S <sup>2</sup> WSU	20.5709	0.9995	31.9461	1	41.4053	1
	$(\lambda = 5e-3)$		$(\lambda = 3e-3)$		$(\lambda = 6e-4)$	

by SUnSAL, SUnSAL-TV, LCSU, DRSU, DRSU-TV, and S<sup>2</sup>WSU algorithms for three highly representative minerals in the Cuprite mining district (Alunite, Buddingtonite, and Chalcidony). In this experiment, the regularization parameters used for SUnSAL, LCSU, DRSU, and S<sup>2</sup>WSU were empirically set to  $\lambda = 0.001$ ,  $\lambda = 0.001$ ,  $\lambda = 0.0001$ , and  $\lambda = 0.002$ , respectively, while the parameters for SUnSAL-TV and DRSU-TV were set to  $\lambda = 0.001$ ,  $\lambda_{TV} = 0.001$  and  $\lambda = 0.002$ ,  $\lambda_{TV} = 0.0001$ , respectively. As shown in Fig. 13.4, all the algorithms obtained reasonable unmixing results, with high abundances for the pixels showing the presence of the considered minerals. This indicates that the sparse unmixing algorithms can lead to good interpretation of the considered hyperspectral dataset. However, it can be seen that some of the abundance maps (e.g., Buddingtonite mineral) estimated by SUnSAL and SUnSAL-TV look noisy and the results obtained by SUnSAL-TV are oversmoothed. In addition, DRSU yields abundance maps without good spatial consistency of the minerals of interest (e.g., Chalcidony mineral), and we can also find that the abundances estimated by S<sup>2</sup>WSU algorithms are generally comparable or higher in the regions classified as respective minerals in comparison to DRSU. Finally, the *sparsity* obtained by SUnSAL, SUnSAL-TV, LCSU, DRSU, DRSU-TV, and S<sup>2</sup>WSU are 0.0682, 0.0743, 0.0734, 0.0430, 0.0423, and 0.0420, respectively. These small differences lead to the conclusion that the proposed approaches use a smaller number of elements to explain the data, thus obtaining higher sparsity. Therefore, from a qualitative viewpoint, we can conclude that the S<sup>2</sup>WSU method exhibits good potential to improve the results obtained by other algorithms in real analysis scenarios.

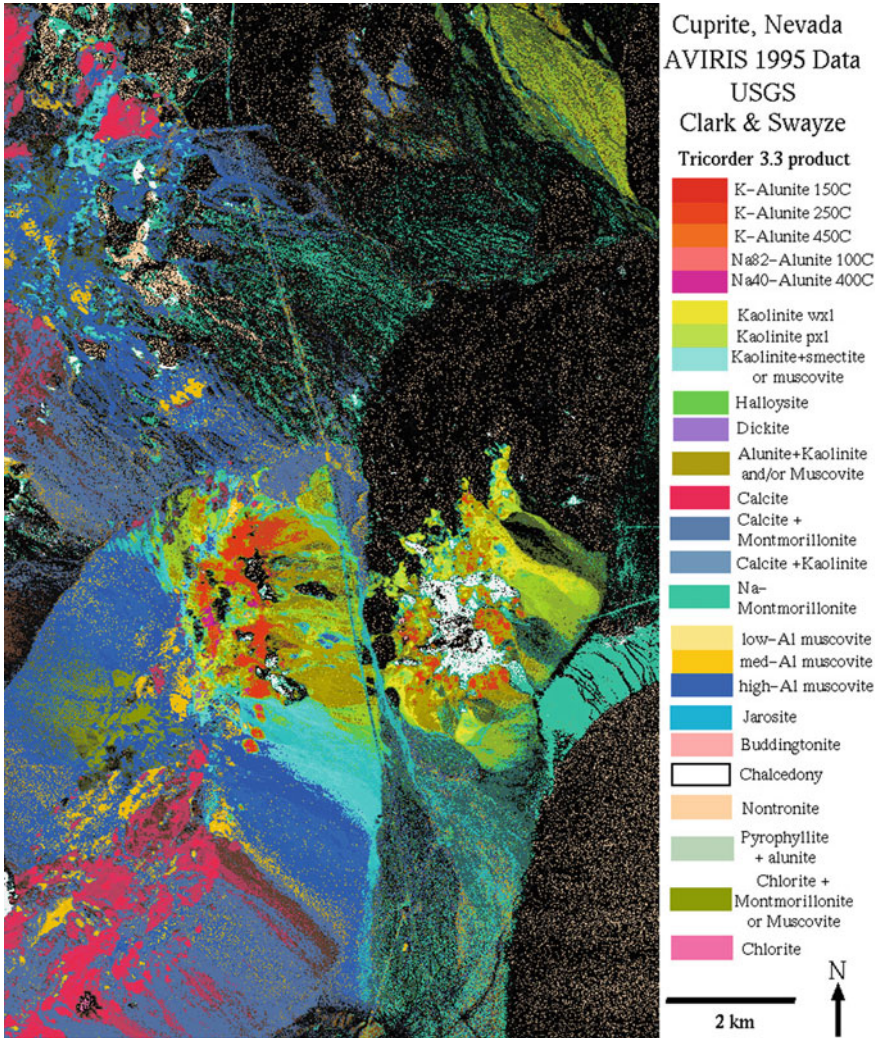
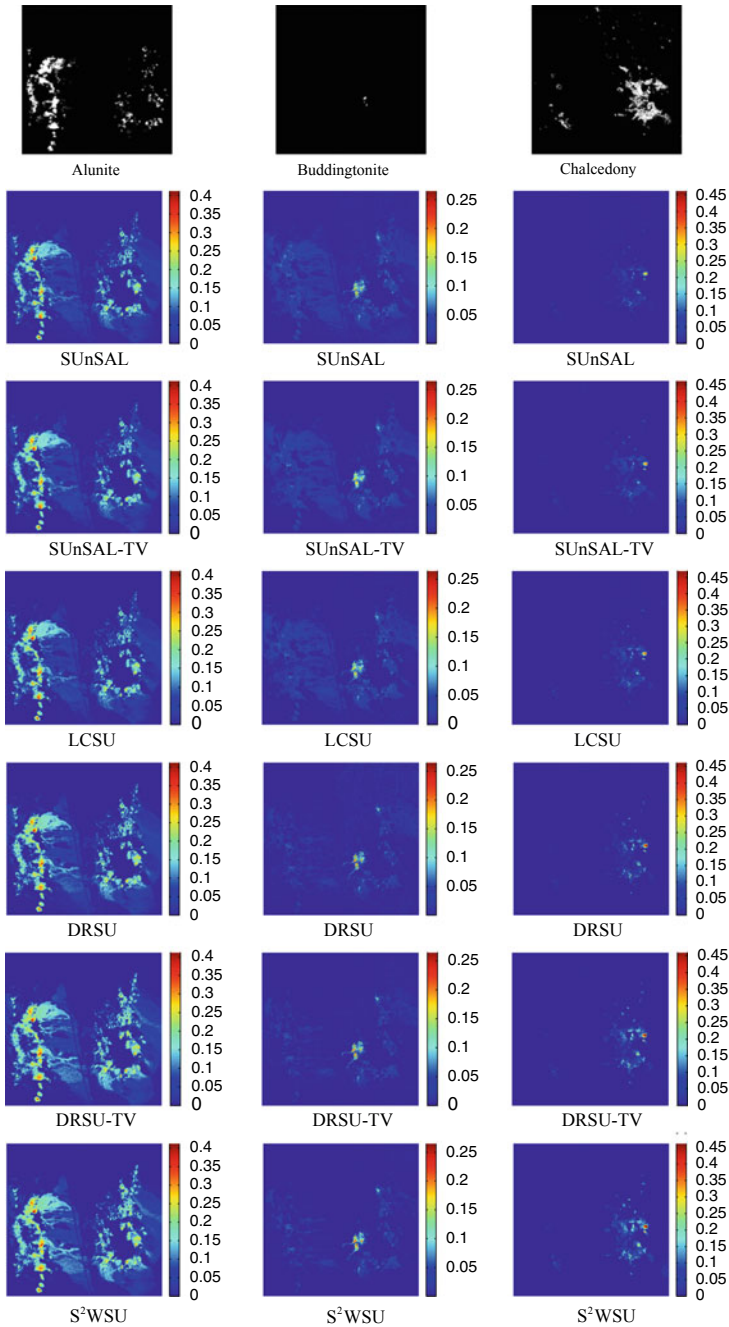


Fig. 13.3 USGS map showing the location of different minerals in the Cuprite mining district in Nevada

### 13.5 Experiments and Analysis: Deep Learning

In this section, the DAEN approach is applied to two real hyperspectral images: Mangrove [49] and Samson [50] datasets for evaluation. In these experiments, the parameters involved in the considered algorithms follow the settings in the simulated experiments, i.e., we use  $\mu = 0.1$  and  $\lambda = 0.1$ , respectively.



**Fig. 13.4** Fractional abundance maps estimated by SUnSAL, SUnSAL-TV, LCSU, DRSU, DRSU-TV, and S<sup>2</sup>WSU as compared to the classification maps produced by USGS Tricorder software for the considered 350 × 350-pixel subset of the AVIRIS Cuprite scene

We compare the DAEN approach presented in this work with other advanced unmixing algorithms, specifically with the N-FINDR [51], VCA [43], MVC-NMF [47], Bayesian [25], PCOMMEND [52], and SNSA [34] methods.

Three indicators, i.e., SAD, reconstruction error (RE), and root-mean-square error (RMSE) are used to measure the accuracy of the unmixing results, which are defined as follows:

$$\begin{cases} \text{SAD}(\mathbf{w}_i, \widehat{\mathbf{w}}_i) = \arccos\left(\frac{|\mathbf{w}_i \cdot \widehat{\mathbf{w}}_i|}{\|\mathbf{w}_i\| \cdot \|\widehat{\mathbf{w}}_i\|}\right), \\ \text{RE}(\{\mathbf{y}_j\}_{j=1}^n, \{\widehat{\mathbf{y}}_j\}_{j=1}^n) = \frac{1}{n} \sum_{j=1}^n \sqrt{\|\mathbf{y}_j - \widehat{\mathbf{y}}_j\|_2^2}, \\ \text{RMSE}(\widehat{\mathbf{h}}_j, \mathbf{h}_j) = \frac{1}{n} \sum_{i=1}^n \sqrt{\|\mathbf{h}_j - \widehat{\mathbf{h}}_j\|_2^2}, \end{cases} \quad (13.37)$$

where  $\widehat{\mathbf{w}}_i$  and  $\mathbf{w}_i$  denote the extracted endmember and the library spectrum,  $\widehat{\mathbf{y}}_j$  and  $\mathbf{y}_j$  are the reconstruction and original signature of pixel  $j$ , and  $\widehat{\mathbf{h}}_j$  and  $\mathbf{h}_j$  are the corresponding estimated and actual abundance fractions, respectively.

### 13.5.1 Mangrove Dataset

The Mangrove data is an EO-1 Hyperion (hyperspectral) image which has been obtained from the USGS Earth Resources Observation and Science (EROS) Center through a data acquisition request to the satellite data provider [49], and collected over the Henry Island of the Sunderban Biosphere Reserve of West Bengal, India. After applying atmospheric correction, we have converted the radiance data to reflectance units by using FLAASH model in ENVI software, and the endmembers (pure signatures of mangrove species) have been identified by a ground survey of the study area, including *Avicennia*, *Bruguiera*, *Excoecaria*, *Phoenix*. The Mangrove data, as shown in Fig. 13.5, includes  $137 \times 187$  pixels and 155 bands, with a spatial resolution of 30 m. For detailed information of the Mangrove data, we refer to [49].

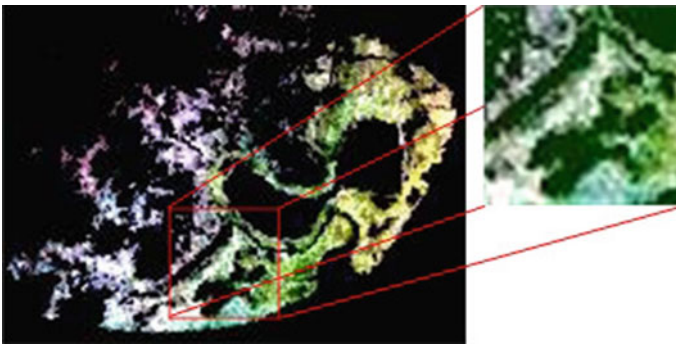


Fig. 13.5 The  $45 \times 45$  pixel subscene of the Mangrove data used in our experiment

**Table 13.2** SADs (in radians) and REs along with their standard deviations obtained by different methods for the Mangrove data from 10 Monte Carlo runs, where the best results are in bold

Mineral	N-FINDR	VCA	MVC-NMF	Bayesian	PCOMMEND	SNSA	DAEN
Avicennia	0.1495 ± 0%	0.9602 ± 4.65%	0.1973 ± 2.18%	0.1001 ± 3.05%	0.0992 ± 2.81%	0.0995 ± 4.26%	<b>0.0968 ±</b> 2.58%
Bruguiera	0.8235 ± 0%	0.8824 ± 7.79%	0.9825 ± 7.53%	1.5957 ± 4.16%	0.9564 ± 3.95%	0.1103 ± 3.84%	<b>0.1025 ±</b> 4.03%
Excoecaria	0.7361 ± 0%	0.7377 ± 1.50%	0.0904 ± 4.08%	0.0963 ± 2.96%	0.0876 ± 5.73%	0.0880 ± 3.27%	<b>0.0863 ±</b> 1.96%
Phoenix	0.1306 ± 0%	1.3063 ± 0.12%	0.9782 ± 3.93%	0.9624 ± 3.93%	1.7065 ± 8.23%	0.0711 ± 4.41%	<b>0.0706 ±</b> 3.72%
Mean SAD	0.4599	0.9717	0.5621	0.6886	0.7124	0.0922	<b>0.0890</b>
RE	0.0822 ± 0%	0.0980 ± 0.40%	0.0392 ± 2.36%	0.0129 ± 4.57%	0.0162 ± 5.19%	0.0057 ± 0.25%	<b>0.0050 ±</b> 0.13%

In our experiment, a subscene with  $45 \times 45$  pixels of the Mangrove data has been used to further evaluate the proposed DAEN. Following [49], the considered subscene contains four endmembers, i.e.,  $m = 4$ .

Table 13.2 presents the obtained quantitative results from the Mangrove data. It can be seen that the DAEN achieved very promising results for the four considered mangrove species. However, the other competitors ended up with errors when detecting or estimating the endmembers. This is due to the fact that, according to our observation, the Mangrove scene contains many outliers across the whole image, which brings a lot of difficulties for general unmixing methods. This point was verified by our experiment, in which we detected a total of 17 outliers. For illustrative purposes, Fig. 13.6 scatterplots the unmixing results obtained by the considered methods, in which the detected outliers are also illustrated. From Fig. 13.6, we can observe that the DAEN produced good unmixing results for this dataset, while all the other methods resulted in problems.

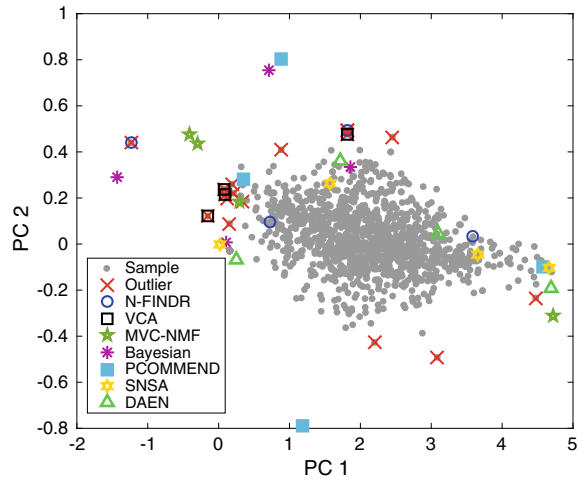
Finally, for illustrative purposes, the estimated endmember signatures, along with their ground references, and the corresponding abundance maps obtained by the DAEN are shown in Fig. 13.7. Effective results can be observed from these figures.

In summary, our experiments with this challenging Mangrove dataset demonstrate the effectiveness of the DAEN for real scenarios with outliers, which is a general situation in real problems.

### 13.5.2 Samson Dataset

In this experiment, we use the Samson dataset which includes 156 bands covering the wavelengths from 0.401 to 0.889  $\mu\text{m}$ , and  $95 \times 95$  pixels, as shown Fig. 13.8,

**Fig. 13.6** Unmixing results for the subscene of the Mangrove data, where the data are projected onto the first two principal components (PCs)



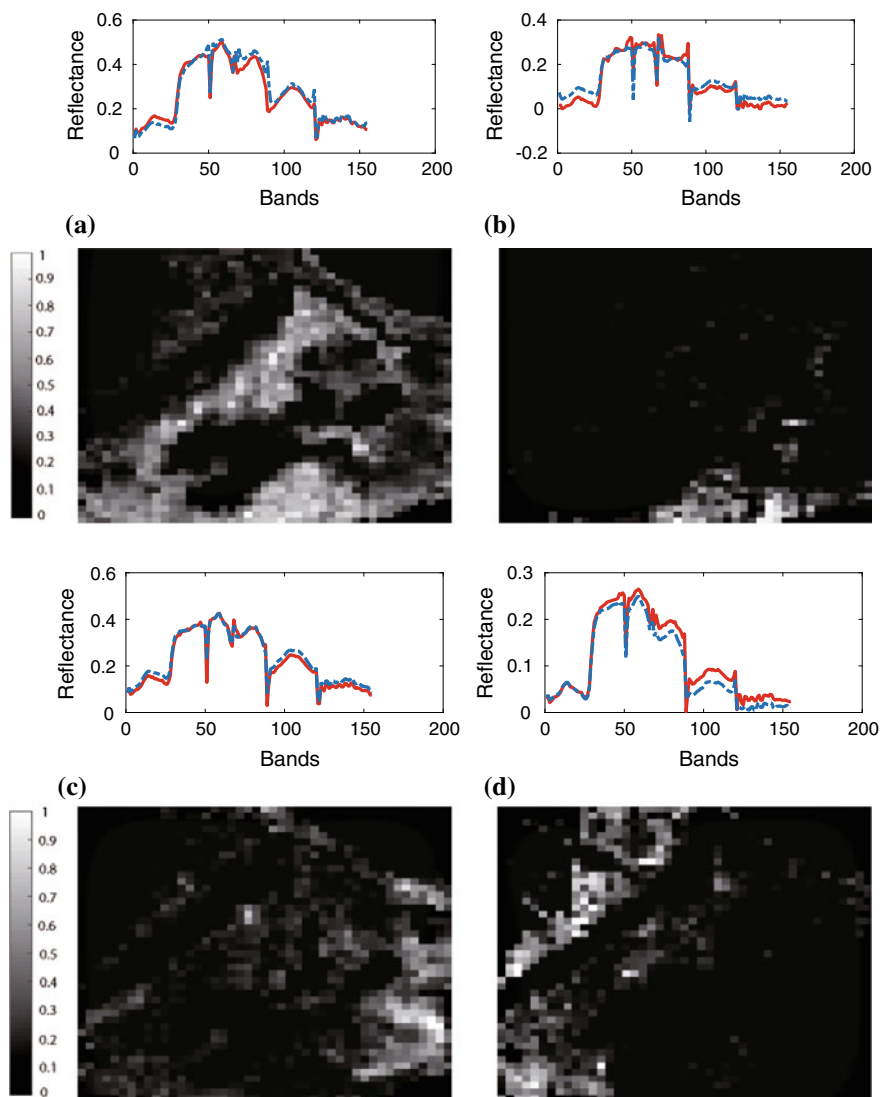
for validation [50]. There are three endmembers including Soil, Tree, and Water in the ground truth image.

Table 13.3 demonstrates the obtained quantitative results for the considered methods. It can be observed that the proposed DAEN obtained the best mean SAD and RMSE. For illustrative purposes, the endmember signatures and the estimated abundances are shown in Fig. 13.9. These figures reveal that the endmembers and abundances, estimated from DAEN, have good matches with regard to the corresponding ones in the ground truth.

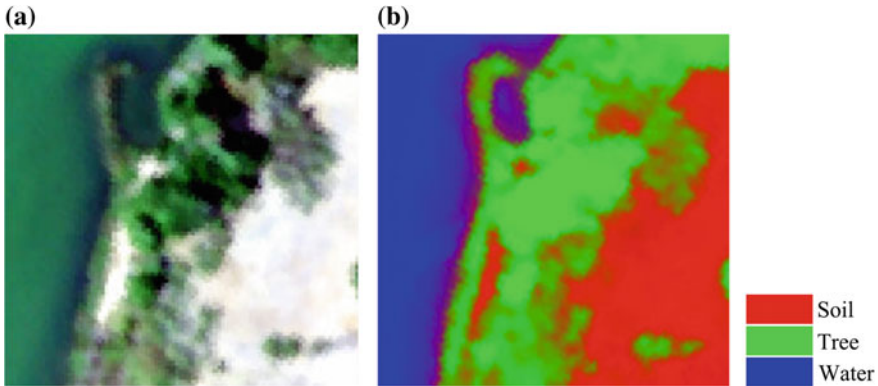
## 13.6 Conclusions and Future Work

Spectral unmixing provides a way to quantitatively analyze sub-pixel components in remotely sensed hyperspectral images [19]. Sparse unmixing has been widely used as a semi-supervised approach that requires the presence of a library of spectral signatures. In this context, spectral–spatial sparse unmixing methods, which aim at collaboratively exploiting spectral and spatial–contextual information, offer a powerful unmixing strategy in case a complete spectral library is available a priori. If no spectral library is available *in advance*, we suggest the fully unsupervised deep auto-encoder network (DAEN) unmixing as a powerful approach that can effectively deal with the presence of outliers in hyperspectral data. Our experimental results reveal that the two aforementioned techniques are currently at the forefront of spectral unmixing. Specifically, we empirically found that the  $S^2$ WSU algorithm consistently achieves better unmixing performance than other advanced spectral unmixing algorithms in case a spectral library is available. This implies that the integration of spectral and spatial–contextual information via the considered spectral–spatial weighted strat-





**Fig. 13.7** The estimated endmember signatures (in red), along with the ground reference (in blue) and their corresponding abundance maps by the proposed DAEN. **a** Avicennia, **b** Bruguiera, **c** Excoecaria, **d** Phoenix

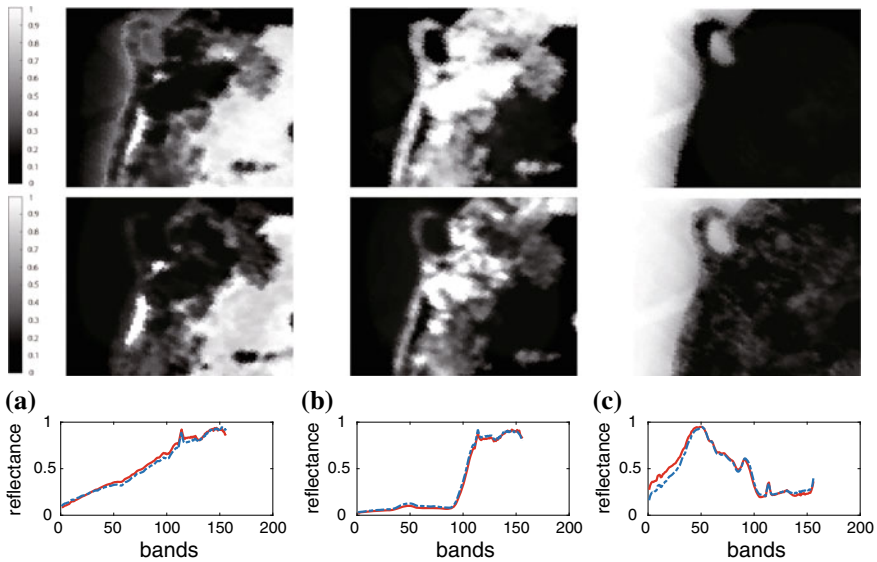


**Fig. 13.8** The Samson image (a) and its corresponding ground truth (b)

**Table 13.3** SADs (in radians) and REs along with their standard deviations obtained by different methods for the Samson data from 10 Monte Carlo runs, where the best results are in bold

Mineral	N-FINDR	VCA	MVC-NMF	Bayesian	PCOMMEND	SNSA	DAEN
Soil	0.0713 ± 0%	0.0627 ± 1.85%	<b>0.0402</b> ± 3.87%	0.1062 ± 7.25%	0.2849 ± 4.35%	0.0410 ± 5.02%	0.0405 ± 2.76%
Tree	0.0495 ± 0%	0.0501 ± 7.82%	0.0261 ± 3.62%	0.0610 ± 8.34%	0.0505 ± 6.14%	0.0205 ± 2.89%	<b>0.0196</b> ± 3.52%
Water	0.0408 ± 0%	0.0273 ± 3.74%	0.0304 ± 5.29%	0.0364 ± 2.48%	0.0716 ± 4.34%	0.0291 ± 2.59%	<b>0.0279</b> ± 3.83%
Mean SAD	0.0539	0.0467	0.0322	0.0679	0.1357	0.0302	<b>0.0293</b>
RMSE	0.9572 ± 0%	0.8926 ± 1.35%	0.6430 ± 0.98%	0.7501 ± 1.63%	0.9439 ± 2.35%	0.6143 ± 3.37%	<b>0.6097</b> ± 3.62%
RE	0.0129 ± 0%	0.0116 ± 0.19%	0.0075 ± 0.86%	0.0103 ± 0.42%	<b>0.0057</b> ± 1.15%	0.0066 ± 0.25%	0.0062 ± 0.85%

egy has great potential in improving unmixing performance. Our experiments also indicate that the fully unsupervised DAEN approach can handle problems with significant outliers more effectively than other popular spectral unmixing approaches. This is an important observation, since the presence of outliers is common in real problems and traditional unmixing algorithms are often misguided by outliers (that can be also understood as endmembers due to their singularity). Our future work will focus on exploring the combination of sparse unmixing and deep learning algorithms to further improve the unmixing performance.



**Fig. 13.9** Results obtained by the proposed DAEN on the Samson dataset. Top: Ground truth abundance maps on Samson data. Middle: Estimated abundance maps from the proposed DAEN. Bottom: Estimated endmember signatures (in red) along with their corresponding reference signatures (in blue). **a** Soil, **b** tree, **c** water

## References

1. Landgrebe D (2002) Hyperspectral image data analysis. *IEEE Signal Process Mag* 19(1):17–28
2. Bioucas-Dias JM, Plaza A, Camps-Valls G, Scheunders P, Nasrabadi N, Chanussot J (2013) Hyperspectral remote sensing data analysis and future challenges. *IEEE Geosci Remote Sens Mag* 1(2):6–36
3. Bioucas-Dias JM, Plaza A, Dobigeon N, Parente M, Du Q, Gader P, Chanussot J (2012) Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches. *IEEE J Sel Topics Appl Earth Observ Remote Sens* 5(2):354–379
4. Plaza A, Du Q, Bioucas-Dias JM, Jia X, Kruse FA (2011) Foreword to the special issue on spectral unmixing of remotely sensed data. *IEEE Trans Geosci Remote Sens* 49(11):4103–4110
5. Keshava N, Mustard JF (2002) Spectral unmixing. *IEEE Signal Process Mag* 19(1):44–57
6. Dobigeon N, Tourneret JY, Richard C, Bermudez JCM, McLaughlin S, Hero AO (2014) Non-linear unmixing of hyperspectral images: models and algorithms. *IEEE Signal Process Mag* 31(1):82–94
7. Ma WK, Bioucas-Dias JM, Chan TH, Gillis N, Gader P, Plaza AJ, Ambikapathi A, Chi CY (2014) A signal processing perspective on hyperspectral unmixing: Insights from remote sensing. *IEEE Signal Process Mag* 31(1):67–81
8. Iordache M-D, Bioucas-Dias JM, Plaza A (2011) Sparse unmixing of hyperspectral data. *IEEE Trans Geosci Remote Sens* 49(6):2014–2039 Jun
9. Nascimento JM, Bioucas-Dias JM (2005) Vertex component analysis: a fast algorithm to unmix hyperspectral data. *IEEE Trans Geosci Remote Sens* 43(4):898–910 Apr
10. Boardman J, Kruse F, Green R (1995) Mapping target signatures via partial unmixing of AVIRIS data. In: *Proceedings of the JPL airborne earth science workshop*, pp 23–26

11. Li J, Agathos A, Zaharie D, Bioucas-Dias JM, Plaza A, Li X (2015) Minimum volume simplex analysis: a fast algorithm for linear hyperspectral unmixing. *IEEE Trans Geosci Remote Sens* 53(9):5067–5082
12. Zhang S, Agathos A, Li J (2017) Robust minimum volume simplex analysis for hyperspectral unmixing. *IEEE Trans Geosci Remote Sens* 55(11):6431–6439
13. Berman M, Kiiveri H, Lagerstrom R, Ernst A, Dunne R, Huntington JF (2004) ICE: a statistical approach to identifying endmembers in hyperspectral images. *IEEE Trans Geosci Remote Sens* 42(10):2085–2095
14. Iordache M-D, Bioucas-Dias JM, Plaza A (2014) Collaborative sparse regression for hyperspectral unmixing. *IEEE Trans Geosci Remote Sens* 52(1):341–354
15. Lu X, Wu H, Yuan Y, Yan P, Li X (2013) Manifold regularized sparse NMF for hyperspectral unmixing. *IEEE Trans Geosci Remote Sens* 51(5):2815–2826
16. Candès EJ, Wakin MB, Boyd SP (2008) Enhancing sparsity by reweighted  $l_1$  minimization. *J Fourier Anal Appl* 14(5):877–905
17. Zheng CY, Li H, Wang Q, Philip Chen CL (2016) Reweighted sparse regression for hyperspectral unmixing. *IEEE Trans Geosci Remote Sens* 54(1):479–488
18. Wang R, Li HC, Liao W, Piturica A (2016) Double reweighted sparse regression for hyperspectral unmixing. In: *Proceedings of the IEEE international geoscience remote sensing symposium*, pp 6986–6989
19. Shi C, Wang L (2014) Incorporating spatial information in spectral unmixing: a review. *Remote Sens Environ* 149:70–87
20. Iordache M-D, Bioucas-Dias JM, Plaza A (2012) Total variation spatial regularization for sparse hyperspectral unmixing. *IEEE Trans Geosci Remote Sens* 50(11):4484–4502
21. Zhao X, Wang F, Huang T, Ng MK, Plemmons RJ (2013) Deblurring and sparse unmixing for hyperspectral images. *IEEE Trans Geosci Remote Sens* 51(7):4045–4058
22. Zhong Y, Feng R, Zhang L (2014) Non-local sparse unmixing for hyperspectral remote sensing imagery. *IEEE J Sel Topics Appl Earth Observ Remote Sens* 7(6):1889–1909
23. Zhang S, Li J, Liu K, Deng C, Liu L, Plaza A (2016) Hyperspectral unmixing based on local collaborative sparse regression. *IEEE Geosci Remote Sens Lett* 13(5):631–635
24. Zhang S, Li J, Li HC, Deng C, Plaza A (2018) Spectral-spatial weighted sparse regression for hyperspectral image unmixing. *IEEE Trans Geosci Remote Sens* 56(6):3265–3276
25. Dobigeon N, Moussaoui S, Coulon M, Tourneret JY, Hero AO (2009) Joint Bayesian end-member extraction and linear unmixing for hyperspectral imagery. *IEEE Trans Signal Process* 57(11):4355–4368
26. Dobigeon N, Moussaoui S, Tourneret J-S, Carteret C (2009) Bayesian separation of spectral sources under non-negativity and full additivity constraints. *Signal Process* 89(12):2657–2669
27. Schmidt F, Schmidt A, Treguier E, Guiheneuf M, Moussaoui S, Dobigeon N (2010) Implementation strategies for hyperspectral unmixing using Bayesian source separation. *IEEE Trans Geosci Remote Sens* 48(11):4003–4013
28. Plaza J, Plaza A, Perez R, Martinez P (2009) On the use of small training sets for neural network-based characterization of mixed pixels in remotely sensed hyperspectral images. *Pattern Recogn* 42(11):3032–3045
29. Plaza J, Plaza A (2010) Spectral mixture analysis of hyperspectral scenes using intelligently selected training samples. *IEEE Geosci Remote Sens Lett* 7(2):371–375
30. Giorgio LA, Frate FD (2011) Pixel unmixing in hyperspectral data by means of neural networks. *IEEE Trans Geosci Remote Sens* 49(11):4163–4172
31. Guo R, Wang W, Qi H (2015) Hyperspectral image unmixing using autoencoder cascade. In: *Proceedings of the 7th workshop hyperspectral image signal processing: evolution in Remote Sensing*, pp 1–4
32. Su Y, Marinoni A, Li J, Plaza A, Gamba P (2017) Nonnegative sparse autoencoder for robust endmember extraction from remotely sensed hyperspectral images. In: *IEEE international geoscience and remote sensing symposium (IGARSS)*, pp 205–208
33. Qu Y, Qi H, uDAS: an untied denoising autoencoder with sparsity for spectral unmixing. *IEEE Trans Geosci Remote Sens* to be published. <https://doi.org/10.1109/TGRS.2018.2868690>

34. Su Y, Marinoni A, Li J, Plaza J, Gamba P (2018) Stacked nonnegative sparse autoencoders for robust hyperspectral unmixing. *IEEE Geosci Remote Sens Lett* 15(9):1427–1431
35. Hapke B (1993) *Theory of reflectance and emittance spectroscopy*. Cambridge Univ. Press, Cambridge, UK
36. Su Y, Li J, Plaza A, Marinoni A, Gamba P, Chakravorty S, DAEN: deep autoencoder networks for hyperspectral unmixing. *IEEE Trans Geosci Remote Sens*, to be published. <https://doi.org/10.1109/TGRS.2018.2890633>
37. Kingmaand DP, Welling M (2014) Auto-encoding variational Bayes. In: 2014 the international conference on learning representations (ICLR), pp 1–14
38. Candès EJ, Tao T (2005) Decoding by linear programming. *IEEE Trans Inf Theory* 51(12):4203–4215
39. Candès EJ, Tao T (2006) Near-optimal signal recovery from random projections: universal encoding strategies. *IEEE Trans Inf Theory* 52(12):5406–5425
40. Bioucas-Dias JM, Figueiredo M (2010) Alternating direction algorithms for constrained sparse regression: application to hyperspectral unmixing. In: *Proceedings of the 2nd workshop hyperspectral image signal processing: evolution in Remote Sensing*, pp 1–4
41. Wang R, Li HC, Pizurica A, Li J, Plaza A, Emery WJ (2017) Hyperspectral unmixing using double reweighted sparse regression and total variation. *IEEE Geosci Remote Sens Lett* 14(7):1146–1150
42. Heinz DC, Chang C-I (2001) Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE Trans Geosci Remote Sens* 39(3):529–545
43. Nascimento JMP, Dias-Bioucas JM (2005) Vertex component analysis: a fast algorithm to unmix hyperspectral data. *IEEE Trans Geosci Remote Sens* 43(4):898–910
44. Lemme A, Reinhart RF, Steil JJ (2012) Online learning and generalization of parts-based image representations by non-negative sparse autoencoders. *Neural Netw* 33:194–203
45. Nocedal J, Wright SJ (2006) *Numerical Optimization*. Springer Science, New York, NY, Springer, USA, pp 33–36
46. Zeiler MD, ADADELTA: an adaptive learning rate method. *Comput Sci*. [arXiv:1212.5701](https://arxiv.org/abs/1212.5701)
47. Miao L, Qi H (2007) Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization. *IEEE Trans Geosci Remote Sens* 45(3):765–777
48. Clark R, Swayze G, Livo K, Kokaly R, Sutley S, Dalton J, McDougal R, Gent C (2003) *Imaging spectroscopy: earth and planetary remote sensing with the USGS tetracorder and expert systems*. *J Geophys Res* 108(E12):5131–5135 Dec
49. Chakravorty S, Li J, Plaza A (2017) A technique for subpixel analysis of dynamic mangrove ecosystems with time-series hyperspectral image data. *IEEE J Sel Topics Appl Earth Observ Remote Sens* 11(4):1244–1252
50. Zhu F, Wang Y, Fan B, Xiang S, Meng G, Pan C (2014) Spectral unmixing via data-guided sparsity. *IEEE Trans Signal Process* 23(12):5412–5427 Dec
51. Winter ME (1999) N-FINDR: an algorithm for fast autonomous spectral endmember determination in hyperspectral data. In: *Proceedings of the SPIE image symposium V*, vol 3753, pp 1–10
52. Zare A, Gader P, Bchir O, Frigui H (2013) Piecewise convex multiple-model endmember detection and spectral unmixing. *IEEE Trans Geosci Remote Sens* 51(5):2853–2862