ISPRS Journal of Photogrammetry and Remote Sensing 183 (2022) 307–320

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Deep-learning-based latent space encoding for spectral unmixing of geological materials

Arun Pattathal V. [a], Maitreya Mohan Sahoo [b], Alok Porwal [b], Arnon Karnieli [a,*]

[a] The Remote Sensing Laboratory, French Associates Institute for Agriculture and Biotechnology of Drylands, The Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, 8499000, Israel
[b] Centre of Studies in Resources Engineering, Indian Institute of Technology, Bombay, Mumbai 400076, India

ARTICLE INFO

Keywords:
Spectral unmixing
Airborne hyperspectral remote sensing
Deep learning
Denoising
Convolutional neural network
Feature learning
Autoencoder
Geological application

ABSTRACT

The relatively coarse spatial resolution of hyperspectral images causes the mixing of disparate materials' spectral responses in the sensor's instantaneous field of view (IFOV), resulting in mixed pixels. The current study proposes a capsule-based generative encoding model, called a denoising unmixing encoder network (DUENet), to formulate an end-to-end trainable spectral unmixing model. The reconstruction and cross-entropy losses and input prior-based constraints achieve joint optimization of denoising, data imputation, and spectral unmixing. Unlike earlier approaches, interpolation-based convolution and dynamic time wrapping (DTW)-based convolutional units facilitate DUENet to even unmix noisy spectra. In addition to embedding label information to improve the physical significance of the latent space, DUENet dynamically learns the parameters of the interpolation kernels. Benchmark airborne hyperspectral datasets (the Nabesna and Cuprite datasets) and simulated datasets were employed to evaluate the performance of the proposed approach. It was observed that the proposed joint optimization of spectral unmixing and denoising significantly improves the results. The adopted feature characterization using capsules improves the generalizability and gives good results, even with a limited number of training samples. This study shows the need for interpretability-based evaluation measures to analyze the unmixing frameworks based on the concepts learned for each endmember. The experiments confirm that the proposed strategy significantly reduces the models' sensitivity to network parameters.

1. Introduction

Surface geological material (regolith or fresh rock), which is remotely sensed by Earth Observation (EO) sensors, comprises an intricate mixture of minerals at the microscopic level. Each mineral is characterized by distinctive physical, chemical, and spectral properties, called endmembers. The surface geological material's spectral response in a hyperspectral EO-image pixel is a nonlinear function of the individual mineral components (or geological endmembers) weighted by their abundance. Given the relatively lower spatial resolutions of hyperspectral EO images, there is an extremely low probability of finding pure pixels comprising a single endmember. Hence most pixels are mixed pixels (Li et al., 2019; Yuan et al., 2020). Inverse modeling of the mixed pixels' spectral response to identify the individual endmembers in the pixel, called spectral unmixing, estimates endmember abundances and maps their distributions in the pixel. A variety of spectral unmixing approaches have been explored over the last three decades (Qian et al., 2020; Bullock et al., 2020). Most approaches involve the unsupervised extraction of pure pixels followed by abundance estimation based on the minimization of some cost function.

The mixing model is learned using either geometrical, statistical, or sparse regression-based approaches subjected to a subsequent inversion (Bullock et al., 2020; Shah et al., 2020). Spectral variability among endmembers and spatial-spectral contextual information have been further used to enhance unsupervised unmixing accuracy (Mei et al., 2020; Ayres et al., 2021). On the other hand, in endmember estimation, supervised unmixing approaches use available prior information, such as pure pixels, ground-truth information, or other ancillary data interactions (Heylen et al., 2014; Han et al., 2020). In recent years, deep learning (DL) approaches have yielded better results than conventional unmixing approaches owing to their ability to model complex nonlinearities (Qian et al., 2020; Heylen et al., 2014; Borsoi et al., 2020; Dou

* Corresponding author at: The Remote Sensing Laboratory, French Associates Institute for Agriculture and Biotechnology of Drylands, The Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, 8499000, Israel
E-mail address: karnieli@bgu.ac.il (A. Karnieli).

https://doi.org/10.1016/j.isprsjprs.2021.11.008
Received 21 July 2021; Received in revised form 15 October 2021; Accepted 12 November 2021
Available online 30 November 2021

et al., 2020; Ozkan et al., 2019; Palsson et al., 2019; Qi et al., 2020; Signoroni et al., 2019; Su et al., 2019). Most of the existing unmixing approaches do not consider the spectral noise and rely on uniformly and densely sampled unaliased input data (Borsoi et al., 2020). In addition, the serial nature of the spectra and the characteristics of the spectral features, such as the depth, width, and position of the absorption and reflective features, are usually ignored.

Although EO data are regularly sampled, atmospheric and sensor malfunctions affect the data, resulting in missing or erroneous data at irregular intervals. The unmixing models need to consider these malfunctions and degradations, as well as atmospheric and other effects, which incorporate systematic and non-systematic noises in the image spectra. Different techniques and algorithms have been employed to remove noise, particularly spikes in image spectra (Jiang et al., 2019; Rasti et al., 2018; Zhang et al., 2021). Most transform-domain-based denoising methods are sensitive to the selection of the transformation function and do not consider spatial variations in hyperspectral images (Jiang et al., 2021). The Savitzky-Golay filter has been widely employed to remove low amplitude spikes using a derivative transform that emphasizes the data's spectral features (Zhao et al., 2014; Julien and Sobrino, 2019; Koziol et al., 2018; Shekhar, 2016). However, these approaches require the appropriate selection of polynomial order and filter length to the spectra's mineral-specific characteristics and avoid extraneous oscillations. In addition, most approaches consider denoising and spectral unmixing as separate optimization problems.

In this study, a DL-based framework is proposed for jointly optimizing denoising and unmixing, taking into account the specific nature and features of the spectra. It is hypothesized that the joint optimization, considering the specific nature of the pixel and reference spectra, can significantly improve the results, even with minimal training samples. The data imputation approach, proposed for denoising the spectra, considers the local features and piecewise characteristics rather than using static interpolations such as bilinear, nearest neighbor, bicubic, etc. Moreover, using data augmentations to simulate atmospheric and sensor-specific noises makes the proposed approach easily trainable and effective for a wide range of noises, including random and systematic noises. It is also hypothesized that interpretability- and explainability-based evaluation mechanisms can effectively evaluate the DL-based unmixing models in terms of the physical significance of the concepts and features learned. It should be noted that the proposed approach uses the labeled samples for training the network and does not require the samples to be pure.

The main contributions of the current study are (1) joint optimization of the denoising, data imputation, and spectral unmixing; (2) neural architecture for considering the specific nature of the spectra; (3) loss functions and regularizations for learning a meaningful feature manifold; (4) illustration of the proposed approaches in comparison with the main existing unmixing approaches; and (5) interpretability-based evaluation measures for DL-based unmixing models.

2. Related works

2.1. Deep-learning-based unmixing

Among different DL approaches, deep generative models have been successfully explored to learn the low-dimensional representations of the pure pixels for addressing the spectral variability of endmembers (Borsoi et al., 2020; Borsoi et al., 2020). The spatial correlation and spectral information were estimated using deep feature layers to formulate a constrained unmixing problem (Palsson et al., 2019). Recently, the iterative shrinkage-thresholding algorithm was employed to unfold the linear mixing model to incorporate the prior knowledge combining the advantages of model-based and learning-based methods (Qian et al., 2020). Most of these approaches ignore feature-based modeling of the spectra and lack generalizability. Although an autoencoder-based unmixing strategy with a hyper-Laplacian loss function was adopted in (Dou et al., 2020), the approach assumed linear mixing, thereby affecting the modeling of the nonlinearly mixed images. The Endnet (Ozkan et al., 2019) used a double-staged autoencoder pipeline and a spectral-angle-based projection to learn the latent manifold that incorporated the correlations with the ground-truth abundances. Recently, deep convolutional autoencoders with tied decoder weights were employed to estimate abundances utilizing both spectral and spatial information (Khajehrayeni and Ghassemian, 2020). Similarly, a gated autoencoder network was also adopted to adaptively control the contribution of spectral and spatial features in the unmixing process (Hua et al., 2021). An interpretable sparse encoding scheme was proposed in (Xiong et al., 2021), while a perception-based cycle consistency encoding model was adopted in (Gao et al., 2021). The autoencoder-based approaches generally yielded limited results (Hong et al., 2021) owing to their inability to model the spectral and spatial features in a generalizable manner. The use of a two-stream Siamese network to generalize various spectral variabilities and extract physically meaningful endmembers improved the results as compared to the existing autoencoder frameworks (Hong et al., 2021). Using gradient descent and proximity operator, an alternating network was also explored to solve the unmixing problem as a separable denoising problem (Xiong et al., 2021). Similarly, a novel dual-alternating direction method of multipliers for the sparse unmixing problem yielded better results as compared to the conventional alternating direction approaches (Ren et al., 2021). The difficulty in obtaining the ground-truth abundances and the non-consideration of spectral noises affected the practical implementation of these approaches for real-world data (Karami et al., 2014; Zhao et al., 2014). A 3D-convolution-based feature transformation, followed by sparsity estimation, was adopted to achieve acceptable unmixing results (Qi et al., 2020). The model-inspired neural networks (Qian et al., 2020) integrated physics-based and data-based priors to improve interpretability and reduce the vast training sample requirement. Although a sparse unmixing framework, where the Soft-Max activation enforced sum to one constraint, was explored in (Rasti et al., 2021) to obtain sparse optimum abundances, the approach lacked generalizability. A convolutional framework consisting of a linear mixture part and an additive nonlinear mixture part was found to be optimal for unsupervised nonlinear unmixing (Zhao et al., 2021). Similarly, autoencoders, along with kernelization and cross-product layers to account for nonlinear mixing mechanisms also give good results (Shahid and Schizas, 2021). However, most of the above discussed spectral unmixing approaches ignore the geological-class-specific characteristics of the spectra, and either adopt an unsupervised formulation or an abundance ground-truth. Additionally, these approaches ignore the noise effects in the spectra further affecting the performance for real world datasets (Karami et al., 2014).

2.2. Deep-learning-based denoising

DL-based denoising approaches learn nonlinear feature spaces to avoid the traditional interpolation methods' varying assumptions and have reported state-of-the-art results (Rasti et al., 2018; Liu et al., 2021; Sun et al., 2021). The analysis of image denoising approaches indicated that the DL-based approaches obtain high-quality restorations in terms of image quality despite the higher computational complexity (Salamat et al., 2021). In addition, most of the DL-based approaches attempt to model the spatial-spectral characteristics and incorporate reasonable assumptions, such as total variation, sparsity, and low-rank models (Rasti et al., 2018; Koziol et al., 2018; Salamat et al., 2021; Song et al., 2021; Wang et al., 2021). The DL-based denoising approach for the line spectral estimation (Jiang et al., 2019) used a convolutional residual network that first estimated noise by removing the signal component and then eliminated the noise component to obtain a denoised signal. A nonlinear end-to-end mapping between the noisy and clean hypercube using a multi scale spatial-spectral deep convolutional neural network was discussed in (Yuan et al., 2019). Similarly, an encoding framework

was explored in (Xu and Peng, 2021. (2021)) to model the inter-band correlation and spatial-spectral features to denoise hyperspectral images. For hyperspectral images with inconsistent and mixed noises, the traditional data-driven denoising algorithms limit the generalization ability. In this regard, multi-patch collaborative learning coupled with variational low-rank matrix factorization has successfully suppressed the mixed noise in hyperspectral images based on the spatial consistency and nonlocal self-similarities (Liu et al., 2021). Self-supervised hyperspectral image denoising networks consisting of a noise estimator and a CNN denoiser were explored in (Zhao et al., 2021; Wang et al., 2021). A multiscale denoising autoencoder that exploited the invariant information for target detection and spectral reconstruction has given optimal results (Sun et al., 2021). Zhang et al. (Zhang et al., 2021) extended the hyperspectral observation model and proposed a double low-rank matrix decomposition method for HSI denoising and destriping. Most of these existing methods failed to consider the global dependence and correlation between spatial-spectral information. A deep spatial-spectral global reasoning network, discussed in (Cao et al., 2021), estimated the global spatial relations between pixels and across the channels to consider both the local and global information for hyperspectral noise removal. Although DL approaches adopted complex network architectures to combine the multiband information, most of the frameworks lacked generalizability. Besides, most of the above-discussed approaches exploited local similarity in hyperspectral images ignoring the inherent spectral-spatial feature modeling. In this regard, a partial densenet model, discussed in (Yuan et al., 2019), was proposed to extract multiscale spatial-spectral features and band-wise noise intensity to generate the feature maps through the channel attention mechanism. Similarly, a 3D-CNN along with attention modules was explored in (Shi et al., 2021) to model the spatial-spectral features and correlations A 3D kernel regression with L1 constraint was also adopted to denoise the hypercube while preserving the local geometry (Zhang et al., 2021). Although the 3D CNN-based approaches yielded good results, they generally ignored the spectral mixing and the endmember specific features that are essential priors for spectral recovery (Karami et al., 2014; Zhao et al., 2014). An adaptive denoising approach by Jiang et al. (Jiang et al., 2021) exploited the intrinsic characteristics of hypercube and modeled the observation noise as a mixture of Gaussian densities to estimate the mixture parameters. The graph-convolutional neural network (CNN)-based approaches that successfully processed irregularly distributed data exploit local structures using a permutation-invariant feed-forward network to aggregate local features (Mao et al., 2019). Although a few of these approaches are applicable to denoising spectra, the difficulty in obtaining graph-based formulation makes the joint optimization of unmixing and denoising challenging (Jiang et al., 2021; Borsoi et al., 2020).

Most of the state-of-the-art DL-based hyperspectral denoising approaches, including those discussed above, adopt complex constraints, regularizations, and alternating minimization, which are difficult to optimize. Additionally, most of them do not consider the mineral-specific characteristics and serial nature of the spectra. Besides, spectral denoising and unmixing are considered separately in the conventional approaches, thereby ignoring the prior information through reconstruction and label information. It may also be noted that an optimal spectral denoising model should also consider the different types of noise prevalent in the hyperspectral images.

2.3. Evaluation measures

Evaluation of DL-based unmixing models is usually achieved through error computation on validation sets disjointed from the training data. However, the validation error is only a proxy for the true error as the validation set might differ statistically from the actual distribution (Samek et al., 2020). Therefore, the evaluation measure should consider the physical significance of the learned features and the concepts/prototypes learned by the networks for each class. In addition, the analysis of the relevance scores assigned to input features, indicating the extent to which each feature contributes, can also be employed to evaluate the DL-based unmixing models. A recent survey of the different approaches to analyze the interpretability and explainability of DL models can be referred to (Barredo Arrieta et al., 2020; Fan et al., 2020; Tjoa and Guan, 2021). Among the various approaches, the layer-wise relevance propagation (LRP) (Barredo Arrieta et al., 2020; Fan et al., 2020; Simonyan et al., 2014) and the activation-maximization-based (Shankaranarayana and Runje, 2019) explanation and interpretation strategies, respectively, are most suitable for unmixing models. The rich feedback provided by explanation allows, in principle, the space of DL models to be explored in a more guided manner than a validation procedure based only on classification or mean squared errors. This study explores the modeling of activation maximization and LRP for analyzing and comparing the different DL-based spectral unmixing models.

Table 1
Details of the number of training samples for different datasets.

Dataset	Mineral class	No. samples
Nabesna Dataset	Carbonates	64,714
	Epidote	51,009
	Chlorite	95,417
	Muscovite	84,036
	Illite	13,253
	Kaolinite	47,069
	Montmorillonite	80,631
	Nontronite	1833
	Alunite	400
	Jarosite	1997
	Serpentine	471
	Amphibole	39,888
	Talc or Tremolite	568
	Hydrated silica and Chalcedony	464
	Gypsum	479
	Vegetation	96,514
	Snow Melt	62,970
	Water	20,800
Cuprite Dataset	K-alunite 1	15,687
	K-alunite 2	1167
	Na-K + Ca alunite	434
	Alunite + Kaolinite	23,634
	Buddingtonite + Na-Montmorillonite	491
	Well-ordered Kaolinite	27,031
	Dickite	1067
	Kaolinite + White mica	50,551
	Na-Montmorillonite	50,205
	Nontronite	2847
	White mica 2	56,481
	White mica 3	2050
	White mica 4	9586
	Kaolinite + White mica 2	5322
	Chlorite + White mica intimate mixture	12,538
	Calcite	31,092
	Calcite + White mica	97,359
	Calcite + Na-Montmorillonite	18,978
	Calcite + Ca-Montmorillonite	423
	Hydrated Silica	3496
Simulated Dataset	Alunite	449
	Buddingtonite	288
	Chalcedony	499

3. Materials and methods

3.1. Datasets

In the current study, the proposed approach was applied to the spectral unmixing of the real-world benchmark hyperspectral images obtained from the Nabesna region of Alaska and the Cuprite region of Nevada in the United States. In addition, a simulated dataset generated using the nonlinear mixing of some geological endmembers—alunite, buddingtonite, and chalcedony—was also used.

The Nabesna image was acquired by the HyMap imaging

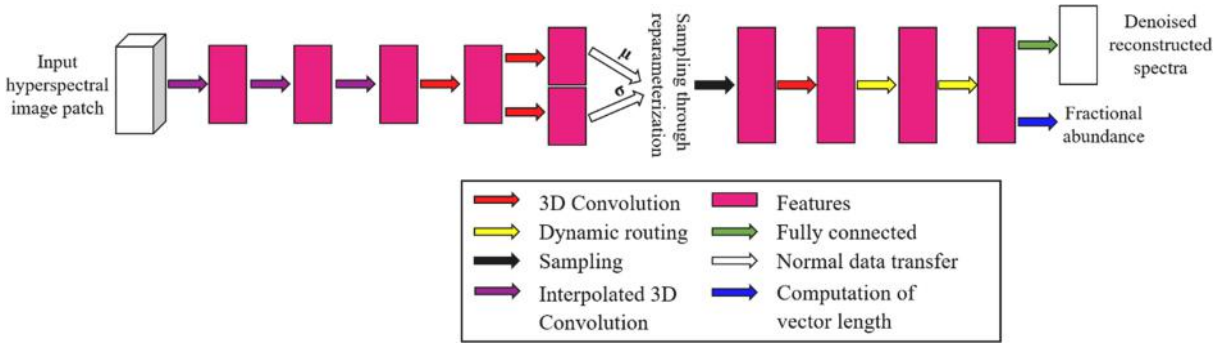

Fig. 1. Proposed denoising unmixing encoder network (DUENet) (μ and σ denote mean and standard deviation, respectively).

spectrometer (Cocks et al., 1998) on July 14 and 21, 2014. The calibrated reflectance image consists of 126 spectral bands covering the visible-near infrared (VNIR) and shortwave infrared (SWIR) regions, with a spatial resolution of 6.7 m. The imaged area extends from the Nabesna Glacier to Wrangell-St. Elias National Park in the Eastern Alaskan range (Graham et al., 2018). It contains good bedrock exposure of early cretaceous Nabesna pluton. The primary mineral constituents of the plutons are quartz, hornblende, biotite, diorite, and monzonite (Richter and Jones, 1973). Several zones of secondary alteration minerals, such as biotite, K-feldspar, chlorite, and sericite have been mapped. A 2352 × 1470-pixel subset of the Nabesna image, featuring the Orange Hill-Bond Creek (142°45′50′′W, 62°15′40′′N), was used in this study. The sub-image contains prominent zones of alteration minerals that have distinctive spectral features in the VNIR-SWIR region. The reflectance image of the study area was processed using the Material Identification and Characterization Algorithm (MICA) to generate a mineral predominance map. For obtaining the predominance map, similar to the one discussed in (Kokaly et al., 2018), the continuum-removed reflectance spectra of each pixel were compared with the continuum-removed absorption features of the endmember spectral library. The reference spectra having the highest fit value (in terms of coefficient of determination) with the pixel spectra were identified as the mineral class for that pixel. The total number of samples of the Nabesna dataset used for training and testing the proposed unmixing models is summarized in Table 1.

The Cuprite dataset was acquired by Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). The image comprises 224 spectral bands covering the VNIR-SWIR region with a spatial resolution of 10.6 m. The bands affected by water vapor absorption and low signal-to-noise ratio were not used, namely, bands 1–2, 105–115, 150–170, and 223–224 (Chang, 2006). A subset of the image (117°13′18.85″ W, 37°39′44.66″ N) consisting of 1172 × 942 pixels was selected for the current study. The study area contains well-exposed alteration zones, including advanced argillic and silicic alteration zones. Some of the key alteration minerals are quartz, chalcedony, alunite, kaolinite, and calcite (Ashley et al., 1980). The reflectance image of the study area was processed using the Tetracorder software (Livo and Clark, 2014) to generate the ground-truth classification map as discussed by (Swayze et al., 2014). In this regard, a modified linear least-squares correlation between the spectral absorption features of the image and the library spectra was employed to identify the best spectral match to the pixel spectra. Further, the abundance maps of endmembers were generated as a product of their fit and spectral absorption depth values. Table 1 summarizes the number of samples of the Cuprite dataset employed for different experiments in this study.

The diagnostic features of the alteration minerals exposed in both areas are absorption features caused by crystal field splitting of the d-orbital of Fe^{2+} and molecular vibrations of metal-OH bonds. The former features are observed in the VNIR spectra of minerals such as chlorite, goethite, hematite, and jarosite. At the same time, the latter appears in the SWIR spectra of minerals such as alunite, buddingtonite, kaolinite, and muscovite (Rossman and Ehlmann, 2019).

The simulated dataset used in this study was generated by the nonlinear mixing of alunite, buddingtonite, and chalcedony using a polynomial post-nonlinear mixing model (Altmann et al., 2011) with an additive Gaussian noise with a signal to noise ratio of 20. The total number of samples generated for each of the endmember classes is presented in Table 1. The spectra of the endmembers comprising 48 bands in the SWIR region and random abundances derived from a spherical Gaussian distribution were used for the nonlinear image simulation.

3.2. Proposed approach

This subsection presents the proposed DL-based unmixing framework, called the denoising unmixing encoder network (DUENet), presented in Fig. 1. The DUENet adopts variationally encoded capsules to model the characteristic features, such as depth, width, and relative location, of the reflective and absorption features specific to each of the geological classes. The fractional abundance estimation is implemented as a capsule-based soft-classification objective constrained with the reconstruction loss.

In DUENet, the hyperspectral input patch is fed to 3D convolutional units that attempt interpolation to address the issue of denoising. The resulting features are convolved, and they are used to sample the latent vector through the reparameterization trick. The sampled latent features are then routed to the feature capsules that are connected to the class capsules. The capsules (Sabour et al., 2017) are a group of neurons that model feature characteristics as the orientation of their output vectors. The length of each capsule's output vector is computed to estimate the fractional abundance of the corresponding class. A fully connected network is employed to reconstruct the pixel spectra, and the reconstruction and cross-entropy losses are minimized to learn the network weights.

The use of interpolated convolution along with the reconstruction of the denoised spectra facilitates denoising in conjunction with unmixing. The interpolated convolution centered at the location p of the pixel spectra (x) is formulated as:

$$\mathbf{x}^{*}\kappa(\mathbf{p}) = \sum_{\mathbf{p'}} \frac{1}{N_{\mathbf{p'}}} \sum_{\mathbf{p}_\delta} \varphi(\mathbf{p}_\delta, \mathbf{p'})\mathbf{x}(\mathbf{p}+\mathbf{p}_\delta).\omega(\mathbf{p'}) \tag{1}$$

where κ is the kernel composed of kernel weights ω. Each kernel weight vector $\omega(\mathbf{p'})$ has a coordinate location $\mathbf{p'}$ relative to the kernel center, and its weight is initialized and updated during training. The interpolation function $\varphi(\mathbf{p}_\delta, \mathbf{p'}) = e^{\frac{\|\mathbf{p}_\delta - \mathbf{p'}\|^2}{\sigma^2}}$ takes the coordinate $\mathbf{p'}$ of a kernel weight vector $\boldsymbol{\omega}(\mathbf{p'})$ and the coordinate $\mathbf{p}_\delta$ of a neighboring input point and computes the weight using the Gaussian interpolation algorithm (Mao et al., 2019). The hyperparameter σ of the interpolation function controls the decay rate. In order to make the convolutions sparsity

invariant, a density normalization term Np′, which sums the interpolation weights or the number of input points in the neighborhood of p′, is employed for each kernel weight vector **ω** (**p'**). The vector coordinate **p'** can either be fixed or updated during training. It may be noted that the interpolated 3D convolution in Eq. (1) dynamically implements the denoising.

To resolve the effects of shifts and noise and to consider the serial nature of the spectra, the Dynamic Time Wrapping (DTW)-based (Iwana et al., 2020) nonlinear units are employed instead of the conventional neurons. The DTW units match features similar to the input and skip elements with a considerable distance to the weights, and they perform small translations. The activation of a given DTW node n is computed as:

$$\mathbf{a_n} = \phi\left(\sum_{(i,j)\in S}\left\|\mathbf{w_{n,i}} - \mathbf{x_j}\right\|\right) \tag{2}$$

where **x** is a vector of the input values $\mathbf{x_1},\ldots,\mathbf{x_m}$ and $\mathbf{w_n}$ is a vector of the respective weights $\mathbf{w_{n,1}},\ldots,\mathbf{w_{n,m}}$ to the node n. The function $\phi(\cdot)$ is a nonlinear activation function applied to the result, and **S** is a set of all the matched pairs between **x** and $\mathbf{w_n}$ computed using the dynamic wrapping approach discussed in (Iwana et al., 2020). In addition to modifying the neural units, the convolution operation is also modified to consider the specific characteristics and noisy nature of the spectra.

The capsule layers, employed to transform the latent representations to fractional abundances, use dynamic routing to refine the transformations in conjunction with the back propagations. The class capsule's output vector length denotes the fractional abundance of each endmember (class) in each pixel. On this subject, a nonlinear squashing function is employed to ensure that the short vectors get shrunk to almost zero length and long vectors get shrunk to a length slightly below 1. Hence, the output of the j[th] spectral class capsule ($\mathbf{v_j}$) is computed as:

$$\mathbf{v_j} = \frac{\|\mathbf{s_j}\|^2}{1+\|\mathbf{s_j}\|^2}\frac{\mathbf{s_j}}{\|\mathbf{s_j}\|} \tag{3}$$

where $\mathbf{s_j}$ is the total input that is computed as the weighted sum of the prediction vectors ($\hat{\mathbf{u}}_{\mathbf{j/i}}$) of the capsules that are connected to the j[th] class capsule. The coupling coefficients ($\mathbf{C_{ij}}$) between the i[th] and j[th] capsules are determined by a SoftMax routing procedure as:

$$\mathbf{C_{ij}} = \frac{\exp(\mathbf{b_{ij}})}{\sum_k \exp(\mathbf{b_{ik}})} \tag{4}$$

The log priors ($\mathbf{b_{ij}}$) are learned together with the network weights and are iteratively refined by measuring the agreement between the output of the j[th] spectral class capsule ($\mathbf{v_j}$) and the prediction vectors $\hat{\mathbf{u}}_{\mathbf{j/i}}$. However, unlike in the conventional implementations, a DTW-based similarity measure is employed to consider the shape of the index curve features, i.e.,

$$\mathbf{b_{ij}}^{\mathbf{k}} = \mathbf{b_{ij}}^{\mathbf{k-1}} + \sum_{m,n\in S}\|\mathbf{v_j}^{\mathbf{k-1}}{}_{\mathbf{m}} - \hat{\mathbf{u}}_{\mathbf{j/i}}{}^{\mathbf{k-1}}{}_{\mathbf{n}}\|^2 \tag{5}$$

where $\mathbf{b_{ij}}^{\mathbf{k}}$ denotes the logits at each iteration k, and **S** is the set containing the indices of elements along the warping path between the $\mathbf{v_j}^{\mathbf{k-1}}$ and $\hat{\mathbf{u}}_{\mathbf{j/i}}{}^{\mathbf{k-1}}$. The warping path between two vectors is computed according to the discussions in (Iwana et al., 2020; Cuturi and Blondel, 2017). The current study adopts the cross-entropy loss to fine-tune the latent representations for the classification task as:

$$L_C = \sum_{i=1}^{M}\alpha_1\left(\sum_{j=1}^{N} -\left[\mathbf{y_{ij}}\log\left(\mathbf{y}_{ij}\right) + \left(1-\mathbf{y}_{ij}\right)\log\left(1-\mathbf{y}_{ij}\right)\right]\right) + \alpha_2|z_i| \tag{6}$$

where $\mathbf{y_i}$ and $\boldsymbol{y}_i$ are the one-hot encoded label and predicted label vectors, respectively, for the i[th] sample, $\mathbf{z_i}$ is the latent representation of the i[th] sample, α_1 and α_2 are the scaling factors and are, respectively, set to 0.92 and 0.08 for all the datasets, M is the number of samples, and N is the number of classes.

The DUENet framework implements denoised spectra reconstruction in conjunction with the incorporation of prior geological class information as:

$$L_G = E_{\mathbf{x}}\left[E_{q_r(\mathbf{z}|\mathbf{x})}[-\log(p_{\tilde{I}}(\mathbf{x}|\mathbf{z}))]\right] - E_{\mathbf{x}}\left[D_{KL}(q_r(\mathbf{z}|\mathbf{x})||N(\mathbf{z}|\mu_y,I))\right] - \beta D_{KL}\left(q_r(\mathbf{z})||\prod_{j=1}^{d}q_r(\mathbf{z_j})\right) \tag{7}$$

where **x** is the input, **x** is the reconstructed pixel spectra, $\mu_{\mathbf{y}}$ is the posterior mean of class **y**, **z** is the latent representation, $q_r(\mathbf{z}|\mathbf{x})$ is the encoding distribution, $p_{\tilde{I}}(\mathbf{x}|\mathbf{z})$ is the decoding distribution, d is the dimension of the latent space modeled by the encoder, β is the entanglement penalty factor, and $D_{KL}(.||.)$ is the Kullback–Leibler divergence. Although penalizing $D_{KL}\left(q_r(\mathbf{z}|\mathbf{x})||N(\mathbf{z}|\mu_y,I)\right)$ facilitates disentanglement, it amounts to the loss of the information about × stored in **z**, resulting in a poor reconstruction. The reformulation adopted in Eq. (7) resolves the effect of the stochastic sampling in the latent code, where the last term measures the dependence for multiple variables. In addition, $p_{\tilde{I}}(\mathbf{z})$ is considered as a mixture of unit Gaussians and is conditioned on the label information to refine the latent space based on the class label information. The embedding of label information obviates the need to marginalize all classes to compute the K-L divergence. In order to facilitate the denoising and data imputation of the spectra, in addition to the variational loss in Eq. (7), a spectral-dissimilarity loss (L_S) is employed as:

$$L_S = \frac{1}{m\pi}\sum_{i=1}^{m}\arccos\left(\frac{\mathbf{x}.\mathbf{x}}{|\mathbf{x}||\mathbf{x}|}\right) \tag{8}$$

where m is the length of the ground truth pixel spectra **x**, and **x** is the reconstructed pixel spectra. In addition to the spectral dissimilarity loss, a multiscale version of the structural dissimilarity loss is also employed to ensure piecewise similarity as:

$$L_{SD} = \sum_{p\in P} 1 - \left(\frac{2\mu_p\mu_p^{'}+C_1}{\mu_p^2+\mu_p^{'2}+C_1}\cdot\frac{2\sigma_p\sigma_p^{'}+C_2}{\sigma_p^2+\sigma_p^{'2}+C_2}\right) \tag{9}$$

where $P \subseteq R$ denotes the set of all relative locations on the spectra, C_1 and C_2 are constants, and μ_p and $\mu_{p'}$, respectively, represent the means of the reconstructed and ground-truth patches. It may be noted that σ_p and $\sigma_p^{'}$, respectively, denote the corresponding standard deviations. The means and standard deviations are computed in the neighborhoods (context) of varied extents to implement multiscale measurements of structural dissimilarity.

The reconstruction and cross-entropy losses, along with the proposed constraints, discussed above, are minimized to train the network weights in accordance with the classification objective. The final loss objective of the DUENet is formulated as:

$$L_{DUENet} = \gamma_1(L_G + L_S + + L_{SD}) + \gamma_2 L_C \tag{10}$$

where L_G denotes the variational reconstruction loss, Ls is the spectral dissimilarity loss, L_{SD} is the structural dissimilarity loss, L_C is the misclassification loss, and γ_1, and γ_2 are the scale factors. Empirically, the setting of γ_1 and γ_2 respectively to 0.6 and 0.4 gives the best results for all the datasets considered in this study. The normalized length of the class capsules' output serves as a fuzzified measure for the presence of a class in a given pixel and hence corresponds to its abundance.

3.3. Implementation details

The implementation of DUENet, adopted in this study, for spectral unmixing used 64 multi-size kernels of sizes 3 × 3 and 5 × 5 in the

Table 2
Analysis of the effect of proposed architectural variations and constraints on different datasets for 70% of training samples.

Dataset	Architectural variations/ losses	Overall Accuracy	Kappa	NRMSE
Nabesna Dataset	Implementation without variational encoding (Eq. (3))	95.61	0.92	–
	Implementation without capsules (Eqs. (7)–(11))	95.56	0.92	–
	Implementation without an adversarial loss for incorporating input prior to the Gaussian space (Eq. (4))	96.45	0.93	–
	Implementation without interpolated convolution (Eq. (6))	96.64	0.94	–
	Implementation without piecewise loss (Eq. (14))	97.60	0.95	–
	Implementation without DTW-based dissimilarity loss (Eq. (13))	97.19	0.95	–
	Implementation without cosine dissimilarity loss (Eq. (12))	97.05	0.96	–
	Proposed DUENet implementation	**98.09**	**0.96**	**–**
Cuprite Dataset	Implementation without capsules (Eqs. (7)–(11))	93.92	0.91	
	Implementation without variational encoding (Eq. (3))	94.12	0.92	–
	Implementation without piecewise loss (Eq. (14))	95.91	0.93	–
	Implementation without DTW-based dissimilarity loss (Eq. (13))	94.23	0.93	–
	Implementation without interpolated convolution (Eq. (6))	95.06	0.93	–
	Implementation without an adversarial loss for incorporating input prior to the Gaussian space (Eq. (4))	96.24	0.94	–
	Implementation without cosine dissimilarity loss (Eq. (12))	96.07	0.94	–
	Proposed DUENet implementation	**97.56**	**0.95**	**–**
Simulated Dataset	Implementation without capsules (Eqs. (7)–(11))	95.46	0.93	0.39
	Implementation without interpolated convolution (Eq. (6))	96.18	0.94	0.31
	Implementation without variational encoding (Eq. (3))	92.08	0.94	0.36
	Implementation without piecewise loss (Eq. (14))	96.03	0.94	0.28
	Implementation without an adversarial loss for incorporating input prior to the Gaussian space (Eq. (4))	97.09	0.95	0.23
	Implementation without cosine dissimilarity loss (Eq. (12))	97.15	0.95	0.32
	Implementation without DTW-based dissimilarity loss (Eq. (13))	96.83	0.95	0.29
	Proposed DUENet implementation	**98.69**	**0.97**	**0.09**

encoder and decoder streams. The integrated spectral-spatial features of the central pixel spectrum, with a dimension of 126 × 128, were used to sample the unit normal distribution to yield the latent representations. The latent representations, thus obtained, were routed to the capsules. The primary capsule layer comprises 16 capsules of vector length 64. The class capsule layer had 15 class capsules, each with a dimension of 16. The stride of pooling and unpooling layers were kept to two, respectively, for halving and doubling the resolution of the resulting feature maps. Convolutions following the down-sampling steps doubled the number of feature maps, while they were halved by the convolutions following upsampling. The interpolated convolution kernels used a Gaussian interpolation as the interpolation function, and the Gaussian bandwidth (3σ) was fixed to 0.1. As the length of the class capsule's output vector was used to estimate the fractional abundance, no further network modification was required to estimate the fractional area covered by each endmember in a given pixel's IFOV. The MSE loss and cosine dissimilarity loss (Arun et al., 2020) were employed in conjunction with the proposed piecewise dissimilarity loss to learn the network weights.

The pixel spectra used to train the network were labeled into different classes based on the corresponding endmembers. The pure pixel spectra, used for training the DUENet, were selected by employing a spectral feature fitting (SFF)-based similarity measure (Pan et al., 2013). Different downscaling strategies (bilinear, bi-cubic, and nearest neighbor), GAN-based augmentation techniques (Zhang et al., 2019) and noise simulations (Rasti et al., 2018; Ghamisi et al., 2017; Chang et al., 2016) were employed to generate training and testing samples. Multiple downscaling strategies were adopted to avoid the bias of the trained network towards a particular approach. Along with other systematic noise simulations, a random approach was employed to remove values or add Gaussian noise at irregular intervals to train the models to facilitate denoising. The network was trained for 300 epochs with an initial learning rate of 0.01 and a decay rate of 0.5 every 100 epochs with a batch size of 50. Hyper-parameter optimization, proposed in (Bochinski et al., 2018), was employed to optimize the proposed network's parameters.

It should be noted that for implementing the multiscale structural dissimilarity measurements (Eq. (9)), the context extents were varied, including 1, 3, 5, 7, and 9. Also, the constants C_1 and C_2 (Eq. (9)) were respectively set empirically to 0.9 and 1.5 for the Nabesna, Cuprite, and simulated datasets. The entanglement penalty (Eq. (7)) was empirically set to 4 and yielded effective disentanglement for the datasets considered in the current study. The optimal values of scaling factors γ_1 and γ_2 (Eq. (10)) were empirically found to be 0.6, and 0.40, respectively for the Nabesna and Cuprite datasets. A Bayesian approach, proposed in (Falkner et al., 2018), was adopted to optimize the parameter values of the loss functions. As a general trend, the weightage to the spectral dissimilarity- and DTW- based reconstruction losses were found to increase with the increase in the number of noisy samples.

4. Results

To verify the effectiveness of DUENet, extensive experiments were conducted using the Hymap image of the Nabesna region (Nabesna dataset) and the AVIRIS image of Cuprite region (Cuprite dataset) for spectral unmixing. As the standard datasets do not have actual ground-truth fractional abundances, a simulated nonlinearly mixed image with known abundance fractions was also used for evaluating the DUENet. Moreover, simulated datasets obtained by downscaling the standard datasets were also employed to analyze the proposed approaches. Hyper-parameter optimization, proposed in (Bochinski et al., 2018), was employed to optimize the parameters of the different models used in this study. It should be noted that an early stopping framework using a k-fold validation formed the basis of the parameter selection. The ablation analysis of DUENet is discussed in subsection 3.1, and a comparative analysis of DUENet with the benchmark approaches is presented in Subsection 3.2. Augmentations and noise simulations (discussed in subsection 3.3) were employed to evaluate the effect of denoising and data imputation. The confusion matrix-based statistics of predominance maps (overall accuracy and Kappa coefficient) derived from the unmixing results were used as accuracy measures. In addition, the normalized root mean squared error (NRMSE) measured between the predicted and ground-truth abundances was also adopted. Bicubic downscaling was adopted to simulate the low-resolution data for standard datasets where the ground-truth abundance was not available. High values of confusion matrix-based statistics (Kappa and overall accuracy) and low values of NRMSE indicated high accuracy. It should be noted that Z-score-based test statistics (discussed in (Herrmann et al.,

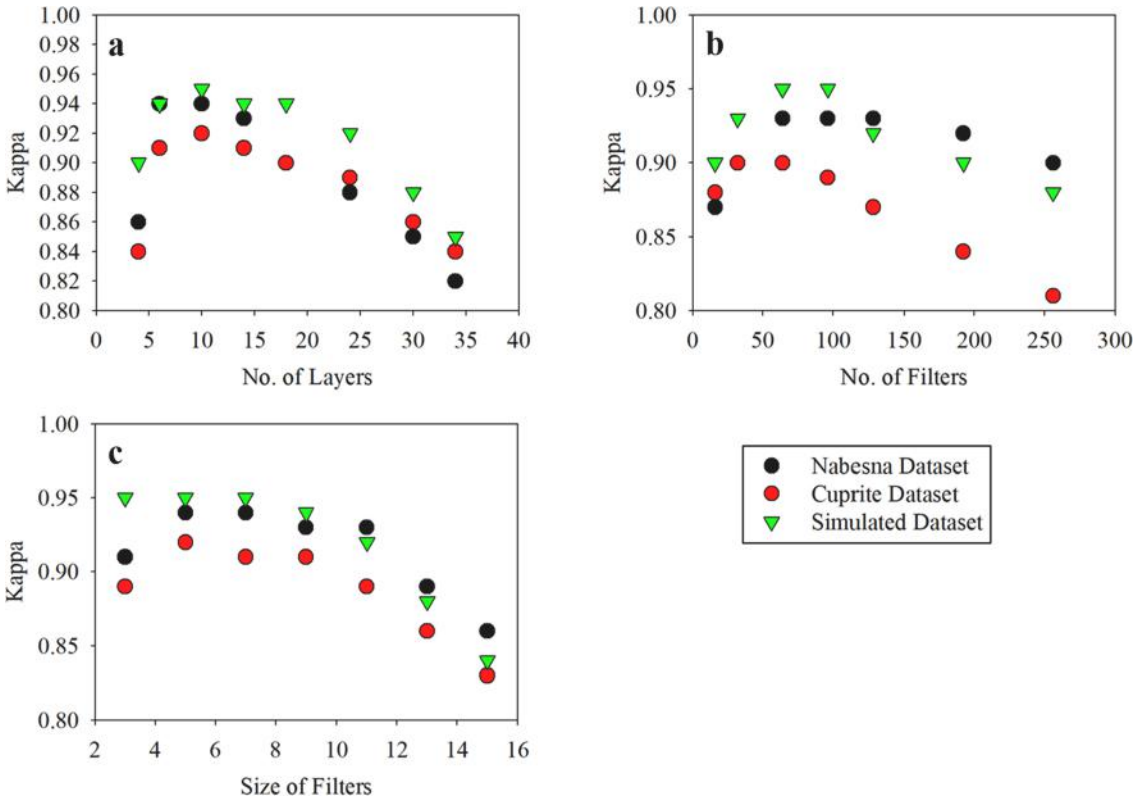

Fig. 2. Analysis of the sensitivity of denoising unmixing encoder network (DUENet) towards (a) depth of the network layers; (b) size of filters; and (c) number of filters for spectral unmixing of different datasets for 70% of training samples when using uniformly sized filters.

2013) was employed to analyze the significance of the results presented in this study. Together with the statistical measures, interpretability, and explainability techniques were used to evaluate the physical significance of the learned features and interpretability of the models. For all the experiments performed in this study, stratified k-fold validation was adopted with k set following the percentage of training samples to be employed.

4.1. Ablation analysis of DUENet implementation for spectral unmixing

The analysis of the proposed architectural variations and constraints (Subsection 2.2) for spectral unmixing of different datasets is summarized in Table 2. Note that the results presented are for the DUENet model trained with 70% of the samples through stratified k-fold cross-validation. As is evident, the proposed architecture is shallower and has fewer parameters as compared to the existing DL-based spectral unmixing approaches. The use of piecewise reconstruction loss, interpolation-based convolution, and DTW-based neural units significantly reduced the training sample requirement and improved the unmixing accuracy. The joint optimization of denoising and unmixing stages, along with the use of constraints to learn disentangled and interpretable representations, further improved the performance (in terms of statistical measures and interpretability). Additionally, fine-tuning of the entanglement penalty facilitated disentangling the latent codes concerning different endmembers. The entanglement penalty values, ranging from 2 to 4, were empirically found to be optimal for the datasets considered in this study. The analysis of latent features and relevance scores indicated that DUENet properly models the characteristic spectral features of different endmembers.

As is evident from Table 2, the implementation of DUENet without capsule stream resulted in lower kappa and overall accuracy values. These results illustrate that capsules effectively model the characteristic features of the spectra so that different classes are distinguished. Moreover, the use of capsules significantly reduced the network depth. It may be seen in Table 2 that the implementation of DUENet without variational encoding strategy resulted in lower kappa and overall accuracy values and higher NRMSE values. The improvement in accuracy due to variational encoding strategy can be attributed to the consideration of the endmember's spectral variability and to the resilience to outliers. The use of DTW-based neural units also improved the results by considering spectral features' shapes. In addition, as observed from the kappa, overall accuracy, and NRMSE values in Table 2, the use of DTW- and spectral dissimilarity-based reconstruction losses resolved the vanishing gradient and convergence issues. The use of reconstruction loss to regulate the classification loss facilitated the denoising and data imputation of the hyperspectral spectra. Training of the proposed network for denoising and imputation made the classification resilient to noises and other irregularities.

The sensitivity analysis of the proposed DUENet model toward network parameters for the different datasets is presented in Fig. 2. It should be noted that the results presented are for the model trained on 70% of the samples through a stratified k-fold cross-validation. Also, different from the results presented in other analyses, uniformly sized kernels were employed instead of multi-sized kernels. Additionally, the parameters other than the one under analysis were fixed to the best value, i.e., for the sensitivity analysis of the network toward the size of the filters for the Nabesna dataset, the depth of the layer and number of filters were, respectively, set to 12 and 128. As is evident from Fig. 2, an increase in the network depth and the number of filters beyond a limit without increasing the training data deteriorated the accuracy. The increase in the size of filters improved the accuracy to a limit, after which it saturated. Empirically, for the Nabesna dataset, with input patches having spatial and spectral dimensions of $28 \times 28 \times 126$, a 6–10 layered network, with a filter size of 5–7 and 64–128 filters, yielded the best results. For the Cuprite sub-image, considered in this study, a filter size of 5×5, network depth of 8–10, and several filters ranging from 32 to 64

Table 3
Comparison of denoising unmixing encoder network (DUENet) with benchmark deep learning (DL) classifiers for 70% of training samples*.

Dataset	Benchmark approaches	NRMSE	Overall Accuracy	Kappa
Nabesna Dataset	(Borsoi et al., 2020)	0.49	91.17	0.88
	(Su et al., 2019) (Su et al., 2019)	0.58	92.09	0.89
	(Qi et al., 2020)	0.30	93.18	0.90
	(Palsson et al., 2019)	0.31	94.45	0.91
	Qian et al., 2020a)	0.41	93.08	0.91
	(Ozkan et al., 2019)	0.36	94.76	0.92
	(Dou et al., 2020)	0.28	96.54	0.93
	(Gao et al., 2021)	0.24	96.21	0.93
	(Jiang et al., 2021) (Jiang et al., 2021)	0.26	97.14	0.94
	(Yuan et al., 2021)	0.18	97.02	0.94
	DUENet	**0.09**	**98.09**	**0.96**
Cuprite Dataset	(Su et al., 2019) (Su et al., 2019)	0.63	90.13	0.87
	(Ozkan et al., 2019)	0.40	92.80	0.88
	(Borsoi et al., 2020)	0.51	92.92	0.89
	(Palsson et al., 2019)	0.47	91.08	0.89
	(Qi et al., 2020)	0.35	93.08	0.89
	(Qian et al., 2020)	0.42	93.19	0.90
	(Dou et al., 2020)	0.33	94.15	0.91
	(Jiang et al., 2021)	0.30	95.32	0.92
	(Gao et al., 2021)	0.28	96.14	0.93
	(Yuan et al., 2021)	0.22	96.07	0.93
	DUENet	**0.12**	**97.56**	**0.95**
Simulated Dataset	Su et al., 2019)	0.43	91.15	0.88
	(Qian et al., 2020)	0.27	92.53	0.89
	(Borsoi et al., 2020)	0.39	93.06	0.91
	(Qi et al., 2020)	0.18	94.51	0.92
	(Yuan et al., 2021)	0.15	96.05	0.93
	(Ozkan et al., 2019)	0.23	96.96	0.93
	(Palsson et al., 2019)	0.30	97.18	0.94
	(Dou et al., 2020)	0.20	97.08	0.94
	(Gao et al., 2021)	0.17	96.89	0.94
	(Jiang et al., 2021)	0.13	97.69	0.94
	DUENet	**0.06**	**98.69**	**0.97**

*Benchmark methods are implemented based on the related publications. Overall accuracy and Kappa coefficient are computed based on the results of approaches for simulated downscaled images.

Table 4
Z-score-based significance analysis of denoising unmixing encoder network (DUENet) in comparison with the benchmark unmixing approaches*.

Dataset	Benchmark smoothing approaches	Z-score as compared to DUENet
Nabesna Dataset	(Su et al., 2019)	2.15
	(Borsoi et al., 2020)	2.76
	(Palsson et al., 2019)	2.10
	(Qian et al., 2020a)	2.62
	(Ozkan et al., 2019)	2.18
	(Qi et al., 2020) (Qi et al., 2020)	2.49
	(Dou et al., 2020) (Dou et al., 2020)	1.98
	(Gao et al., 2021) (Gao et al., 2021)	2.12
	(Jiang et al., 2021) (Jiang et al., 2021)	1.99
	(Yuan et al., 2021)	2.03
Cuprite Dataset	(Su et al., 2019) (Su et al., 2019)	1.97
	(Borsoi et al., 2020)	2.06
	(Palsson et al., 2019)	2.45
	(Qian et al., 2020a)	2.30
	(Ozkan et al., 2019)	2.04
	(Qi et al., 2020) (Qi et al., 2020)	1.93
	(Dou et al., 2020) (Dou et al., 2020)	2.15
	(Gao et al., 2021) (Gao et al., 2021)	2.08
	(Jiang et al., 2021) (Jiang et al., 2021)	1.99
	(Yuan et al., 2021)	2.05
Simulated Dataset	(Su et al., 2019) (Su et al., 2019)	2.08
	(Borsoi et al., 2020)	1.97
	(Palsson et al., 2019)	2.54
	(Qian et al., 2020)	2.26
	(Ozkan et al., 2019)	2.17
	(Qi et al., 2020) (Qi et al., 2020)	2.06
	(Dou et al., 2020) (Dou et al., 2020)	1.98
	(Gao et al., 2021) (Gao et al., 2021)	2.01
	(Jiang et al., 2021) (Jiang et al., 2021)	2.13
	(Yuan et al., 2021)	1.98

* Z-score > 1.96 shows a significant (>95%) difference between the confusion matrices of existing approaches and the denoising unmixing encoder network (DUENet).

were found to be optimal. A network depth of 7–12, a filter size of 3–5, and the number of filters ranging from 64 to 96 were found to be optimal for the unmixing of simulated datasets using DUENet. For all three datasets, the performance of DUENet was found to be less sensitive to slight changes in the depth of the reparameterization stream. The use of capsule layers in the decoder stream significantly reduces the number of parameters and hence the sensitivity of the network and the requirement of training samples.

This study employed multi-sized kernels in each convolutional unit, unlike the conventional use of uniformly sized kernels. However, the model employed for generating the results in Fig. 2 used uniformly sized kernels to compare the filter size effect. The use of multi-sized kernels yielded better results than the use of uniformly sized ones. As the length of spectral features can vary, for modeling relevant features, multi-sized kernels are recommended. It should also be noted that the proposed use of 3D convolutional units effectively learns the characteristic spectral-spatial features that are relevant in distinguishing the different endmembers.

4.2. Comparison of DUENet with the recent spectral unmixing approaches

The recent unmixing approaches, which have reported state-of-the-art accuracies, were compared with the proposed DUENet-based approach. The results on the Nabesna, Cuprite, and the simulated datasets are summarized in Table 3. The significance of the results (at a confidence level of 95%) is analyzed in Table 4. The illustration of the predominance classification maps obtained using the proposed approach for the standard datasets, compared with the corresponding ground-truth, is presented in Figs. 3 and 4. The result of the proposed approach for the simulated dataset is presented in Fig. 5. The variation in accuracy of selected benchmark approaches with the change in training samples for the standard and simulated datasets is summarized in Figs. 6 and 7. As is evident from the results, the joint optimization of denoising and unmixing stages significantly improved the abundance estimation even with a limited number of training samples. In addition, the use of DTW-based neural units and DTW-based loss functions resolved the issues of shifts, distortions, and missing data prevalent in the spectral comparisons. The interpolation-based convolution considered the serial and point nature of the data, and the abrupt changes in the spectral curves facilitated effective missing data imputation. The use of the entanglement penalty helped to identify the relevant features of the endmembers under consideration. The spectral dissimilarity-, DTW-, and structural dissimilarity-based losses, adopted in DUENet, addressed data imputation issues and denoising jointly with the unmixing. The learning of physically significant features also resolved the domain bias and inter-field variability of endmember spectra. The proposed reconstruction-regularized classification facilitated the effective

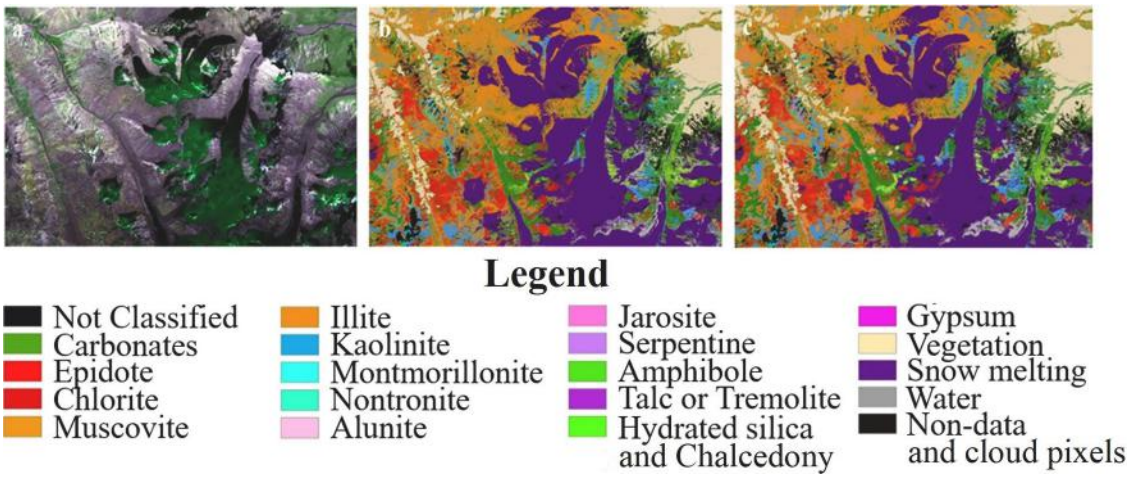

Fig. 3. Results of the proposed spectral unmixing approach on the Nabesna dataset: (a) true color HyMap image of the study area; (b) ground-truth predominance map; and (c) results of the proposed unmixing approach based on the denoising unmixing encoder network (DUENet) model.

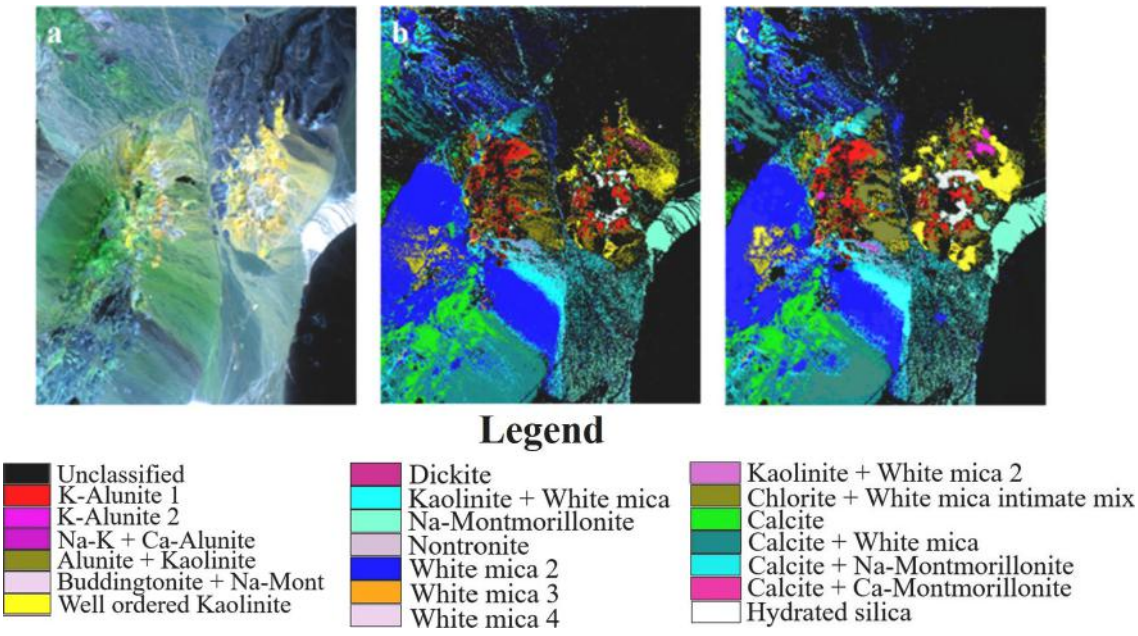

Fig. 4. Results of the proposed spectral unmixing approach on the Cuprite dataset: (a) true color AVIRIS image of the study area; (b) ground-truth Tetracorder map; and (c) results of the proposed unmixing approach – the denoising unmixing encoder network (DUENet) model.

transformation of spectral curves to a latent space that was more discriminative than the original space. The use of reconstruction loss and classification loss facilitated the denoising and data imputation in conjunction with unmixing through effective transformation to a disentangled latent space.

In addition to the classification-based accuracy assessment, the models were also evaluated based on the prototypes learned for each endmember. The prototypes learned by DUENet for the selected endmembers of the Nabesna and Cuprite datasets are presented in Figs. 8–9. The characteristic spectral features for each endmembers were verified by analyzing the learned concepts, and the same process was employed to refine the training sets. The comparison of the learned concepts for each endmember with the corresponding library spectra for different datasets is presented in Table 5. For the Nabesna dataset, the analysis indicated that for a few endmembers, such as kaolinite, montmorillonite, and hydrated silica, the training data needed to be further refined. However, the DUENet can give acceptable results owing to variational encoding and capsule layer-based strategies. In addition, for the Nabesna dataset, the layer-wise propagation of relevance indicated that DUENet gives importance to the spectral bands sensitive to the endmembers as discussed in (Rossman and Ehlmann, 2019; van der Meer, 2004; Shankar, 2015; Fang et al., 2018; Amthauer and Rossman, 1984). The spectral similarity comparison (Arun et al., 2019) of the learned concepts for the Cuprite dataset indicated that DUENet gives importance to the spectral bands sensitive to the endmembers, as discussed in (Rossman and Ehlmann, 2019). The practical learning of concepts for different classes can be attributed to properly modeling spectral characteristics through generative and capsule-based feature learning. The analysis of concepts learned for the endmembers of the simulated dataset also illustrated that DUENet effectively models the features of each endmember, thereby improving the spectral unmixing results. It may be further noted, from the learned concepts, that the proposed approaches give accurate results even when the simulated datasets are noisy. An analysis of the learned features from the bottom layers indicated that the physically significant spectral features are combined linearly and nonlinearly to guide the decision.

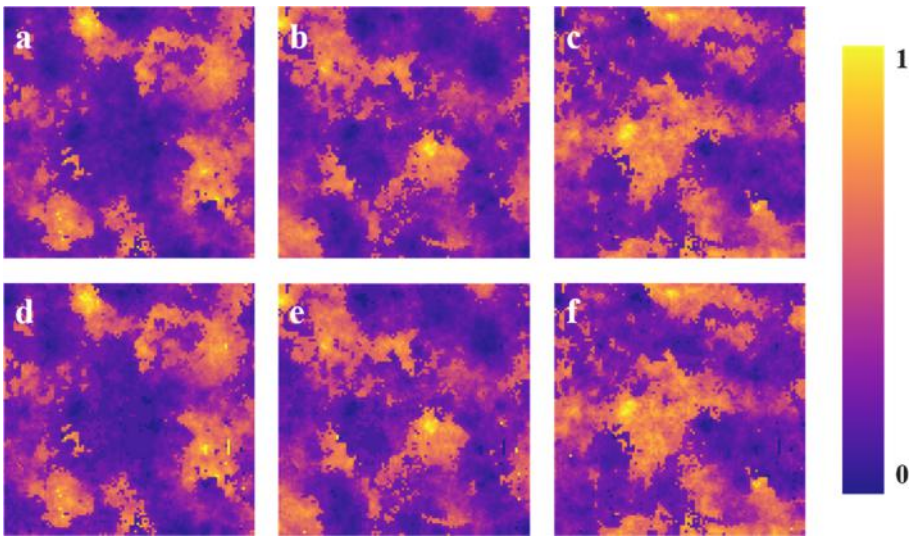

Fig. 5. Results of the proposed spectral unmixing approach on simulated dataset: Ground-truth abundance of (a) alunite (b) buddingtonite (c) chalcedony; abundances of (d) alunite (e) buddingtonite (f) chalcedony estimated by the denoising unmixing encoder network (DUENet) model.

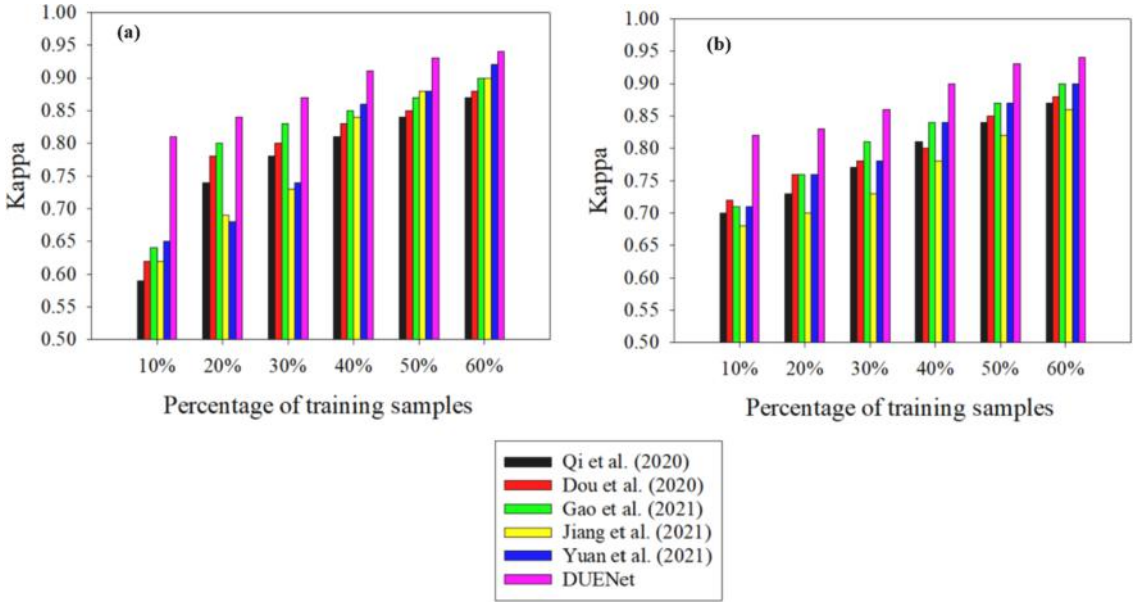

Fig. 6. Accuracy analysis of denoising unmixing encoder network (DUENet) and benchmark unmixing approaches for the: (a) Nabesna dataset; and (b) Cuprite dataset.

5. Discussion

Spectral features and their characteristics, such as absorption depth, width, and position, are crucial in mineral spectral classification. Although normal convolutional architectures and variants have proved effective in classification and abundance estimation, the proposed DUENet employed a generative encoding scheme coupled with capsule-based classification to effectively unmix the spectra in conjunction with denoising and data imputation. As is evident from the results presented in Tables 3 and 4, the joint optimization of spectral denoising and unmixing improves the results of DUENET significantly as the priors derived from both the stages are mutually complementing. The multiple kernels and multi-layer abstractions, constrained with losses to embed discriminative priors, facilitate the modeling of disentangled manifolds, resulting in learning improved meaningful features. The use of multi-sized kernels instead of uniformly sized ones enables the modeling of varied length features. The capsule-based classification module effectively learns spectral features and their characteristics accurately, reducing the training sample requirement. Hence, as is evident from the results, the DUENet performs significantly well even at a reduced percentage of training samples. The analysis of the activation maps indicated that the capsules learn physically significant features compared to those learned using normal convolutional units. However, the conventional convolutional units, aggregated as capsules, did not consider the curved nature of the spectrum while computing the similarity measures. The proposed DTW-based capsule units, used in DUENet, which employ time wrapping similarity measures to compute weight vectors instead of scalars, resolved the distortions and shifts prevalent in spectral curves. In addition, the DTW capsules improved the generalizability of the proposed approaches. The use of DTW-based capsule units in DUENet significantly improved the unmixing results (presented in Table 3) as compared to the conventional approaches. The normalized length of the class capsules' output vectors gave an accurate estimate of the fractional abundances of the corresponding classes, thereby significantly reducing

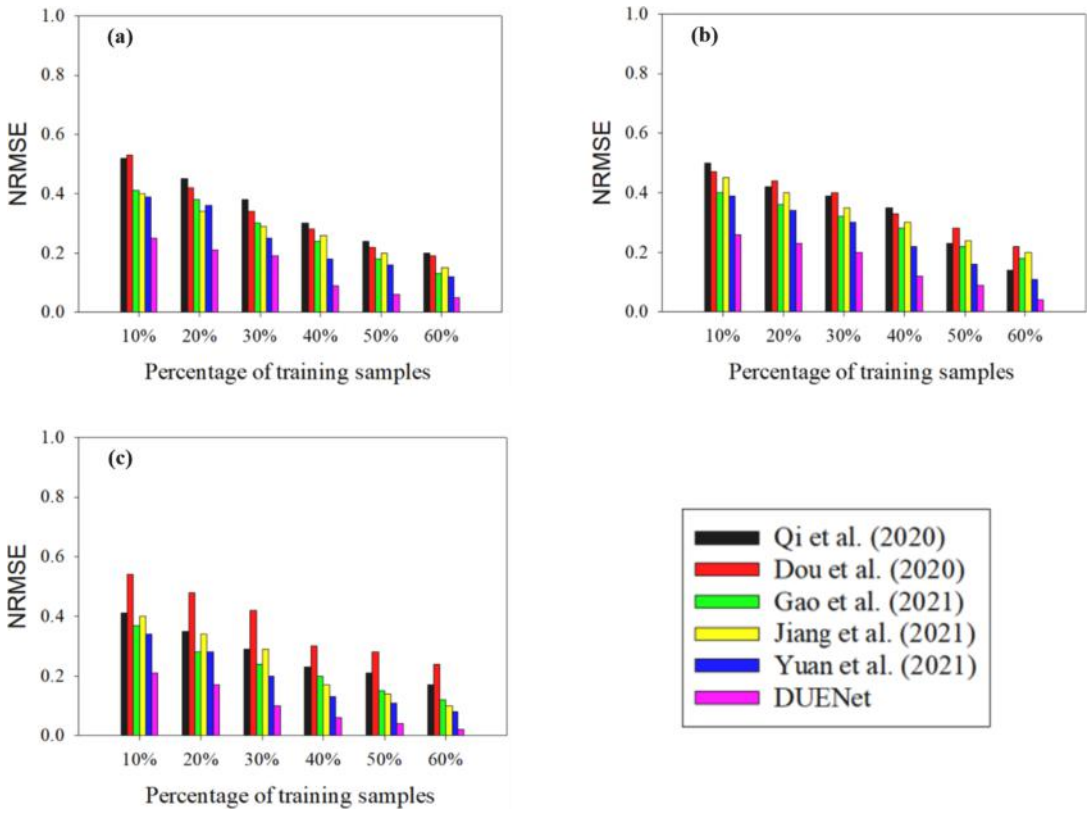

Fig. 7. Accuracy analysis of the denoising unmixing encoder network (DUENet) and benchmark unmixing approaches for the: (a) Nabesna dataset; (b) Cuprite dataset; and (c) simulated dataset.

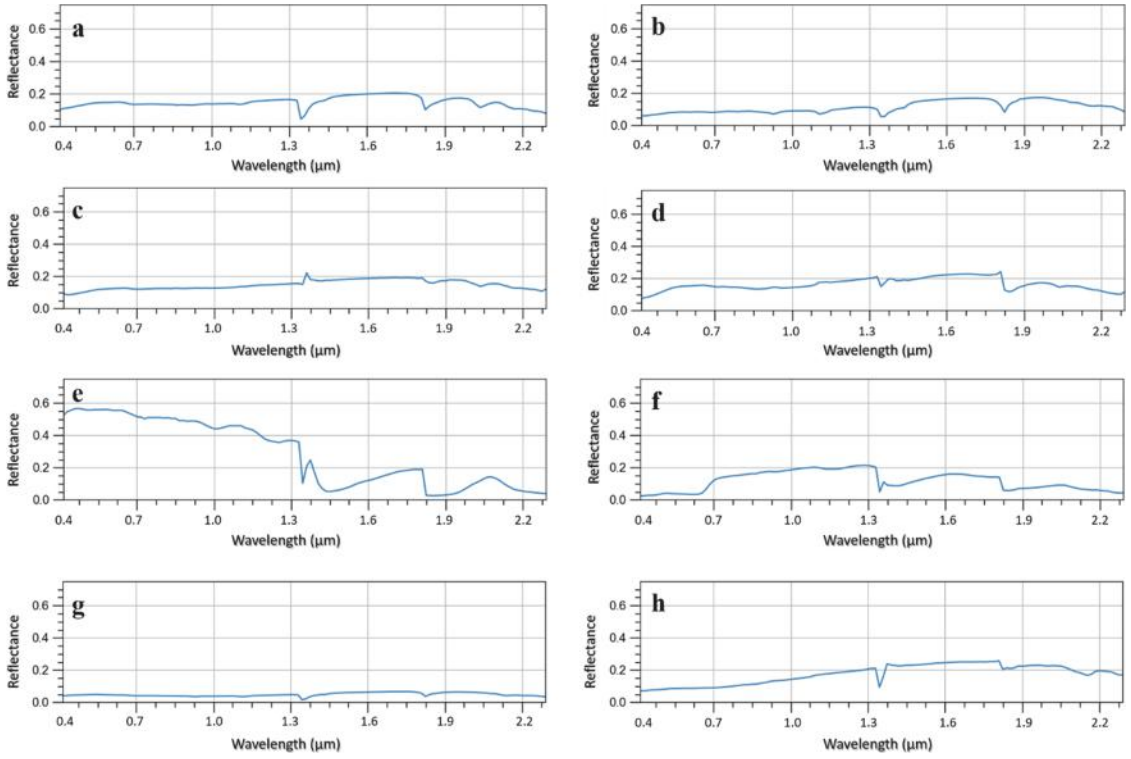

Fig. 8. Concepts learned by the proposed DUENet for different endmembers of the Nabesna dataset: (a) muscovite; (b) epidote; (c) kaolinite; (d) montmorillonite; (e) snow; (f) vegetation; (g) hydrated silica; and (h) carbonate.

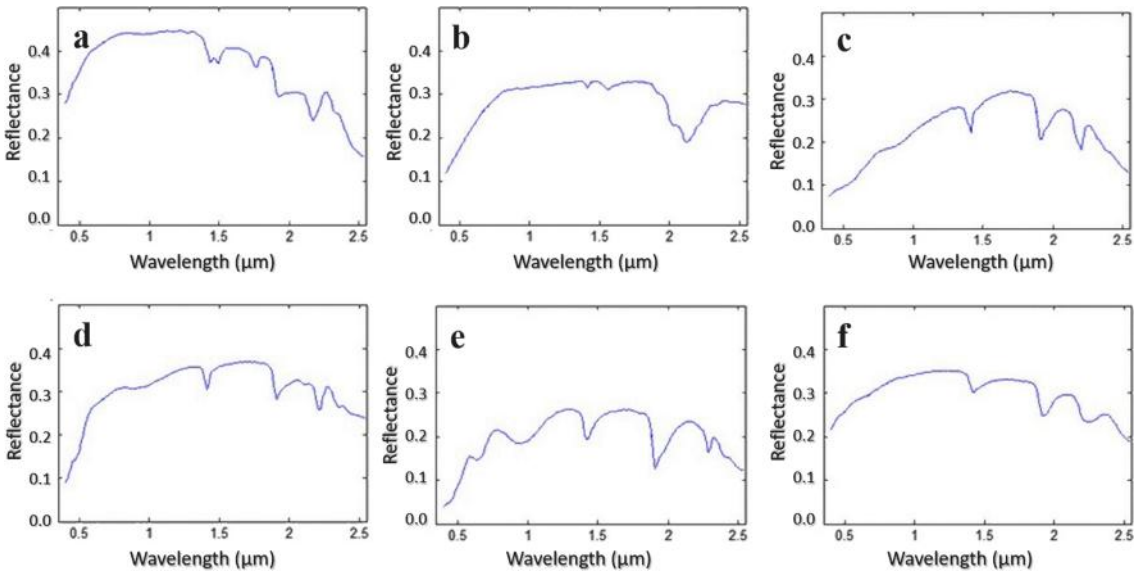

Fig. 9. Concepts learned by the proposed DUENet for selected endmembers of the Cuprite dataset: (a) alunite; (b) buddingtonite; (c) kaolinite; (d) montmorillonite; (e) nontronite; and (f) hydrated silica.

Table 5
Similarity of the concepts learned by DUENet with corresponding endmembers of the Nabesna dataset.

Datasets	Benchmark approaches	Spectral similarity
Nabesna Dataset	Muscovite	0.9994
	Epidote	0.9912
	Kaolinite	0.9841
	Montmorillonite	0.9873
	Snow	0.9979
	Vegetation	0.9938
	Hydrated Silica	0.9924
	Carbonate	0.9991
	Amphibole	0.9949
	Water	0.9999
Cuprite Dataset	Alunite	0.9996
	Calcite	0.9965
	Kaolinite	0.9978
	Dickite	0.9992
	Montmorillonite	0.9909
	Buddingtonite	0.9894
	Hydrated silica	0.9992
	Nontronite	0.9973
Simulated Dataset	Alunite	0.9996
	Buddingtonite	0.9988
	Chalcedony	0.9991

the NRMSE and improving the Kappa values. The similarity of the concepts learned by the DUENet, as presented in Table 5, and the prototypes learned for each class presented in Fig. 9 can be attributed to the proposed architectural modifications that have improved the learning capability of the network. The capsule layers facilitated proper modeling of features for each class. Also, the use of shallow capsule layers for spectral unmixing overcame the vanishing gradient issue. The proposed DUENet gave better results than the conventional approaches, even with a limited number of training samples. The DUENet provided an optimal framework for joint optimization of spectral unmixing and denoising, thereby improving the unmixing performance, even for the noisy samples.

The MSE-based reconstruction loss, along with cross-entropy loss, affected the classification accuracy. However, the proposed constraints and losses addressed this issue and successfully optimized the denoising, data imputation, and unmixing stages. Additionally, DUENet gave good results even when trained with spectra having different resolutions and domain biases.

6. Conclusion

Most of the existing spectral unmixing approaches are unsupervised and do not consider the classification priors. The joint optimization of denoising, data imputation, and spectral unmixing, proposed in this study, significantly improve the results. The proposed approach of coupling variational encoding with capsule layers can resolve the issues of noisy spectra prevalent in spectral unmixing. Unlike the conventional convolution approaches, point-based convolution and DTW-based convolutional units are proposed to process the noisy spectra and to avoid resolution biases. The projection of latent representations to a normally distributed space, constrained by an entanglement penalty and classification prior, improved the endmember discriminability in the learned manifold. The entanglement penalty improved the independence of the latent dimensions and enhanced sparsity. DUENet employs classification and reconstruction losses in conjunction with additional constraints to implement the unmixing. The normalized length of the class capsule's output is employed to estimate the fractional abundance, eliminating the need for additional SoftMax layers and reducing the number of parameters. The enhanced interpretability and physical significance of the learned representations and the transparency of the proposed pipeline significantly improved the results. The current study highlights the need to adopt interpretability-based evaluation measures in conjunction with NRMSE- and confusion matrix-based (overall accuracy and Kappa coefficient) accuracy measures for comparing different unmixing models. The predictor-conditioned input distribution was modeled to understand the most likely model input for a given output. Additionally, the relevance of each of the input spectral bands towards the class-assignment decision was estimated and analyzed. The concepts learned for each endmember indicate that DUENet learns the meaningful spectral features that are crucial to distinguish them. In addition, the network assigned more relevance to the spectral bands whose responses were sensitive to the endmembers. As proposed in this study, the unique use of generative models for unmixing can significantly reduce the network's sensitivity toward hyperparameters.

Declaration of Competing Interest

The authors declare that they have no known competing financial

interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors thank the United States Geological Survey (USGS) and the National Aeronautics and Space Administration – Jet Propulsion Laboratory (NASA-JPL) for the open-access to the datasets, ancillary information, and the Tetracorder software. The authors also thank Dr. K. Eric Livo, Dr. Gregg A. Swayze, and Dr. Raymond Kokaly of the USGS for the discussions related to the Tetracorder software and the development of ground-truth maps.

References

Y. Altmann, A. Halimi, N. Dobigeon, J.Y. Tourneret, Supervised nonlinear spectral unmixing using a polynomial post nonlinear model for hyperspectral imagery, in: ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc., 2011: pp. 1009–1012. https://doi.org/10.1109/ICASSP.2011.5946577.

Amthauer, Georg, Rossman, George R., 1984. Mixed valence of iron in minerals with cation clusters. Phys. Chem. Miner. 11 (1), 37–51. https://doi.org/10.1007/BF00309374.

Arun, P.V., Buddhiraju, Krishna Mohan, Porwal, Alok, 2019. Capsulenet-Based Spatial-Spectral Classifier for Hyperspectral Images. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12 (6), 1849–1865. https://doi.org/10.1109/JSTARS.460944310.1109/JSTARS.2019.2913097.

Arun, P.V., Buddhiraju, Krishna Mohan, Porwal, Alok, Chanussot, Jocelyn, 2020. CNN-Based Super-Resolution of Hyperspectral Images. IEEE Trans. Geosci. Remote Sens. 58 (9), 6106–6121. https://doi.org/10.1109/TGRS.3610.1109/TGRS.2020.2973370.

R.P. Ashley, M.J. Abrams, Alteration Mapping Using Multispectral Image - Cuprite Mining District, Esmeralda County, Nevada, U. S. Geol. Surv. Open-File Rep. (1980) 80–367.

L.C. Ayres, S.J.M. de Almeida, J.C.M. Bermudez, R.A. Borsoi, A Homogeneity-Based Multiscale Hyperspectral Image Representation for Sparse Spectral Unmixing, in: ICASSP 2021 - 2021 IEEE Int. Conf. Acoust. Speech Signal Process., IEEE, 2021: pp. 1460–1464. https://doi.org/10.1109/ICASSP39728.2021.9415024.

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., Herrera, F., 2020. Explainable Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf. Fusion. 58, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012.

E. Bochinski, T. Senst, T. Sikora, Hyper-parameter optimization for convolutional neural network committees based on evolutionary algorithms, in: Proc. - Int. Conf. Image Process. ICIP, IEEE Computer Society, 2018: pp. 3924–3928. https://doi.org/10.1109/ICIP.2017.8297018.

Borsoi, R.A., Imbiriba, T., Closas, P., Bermudez, J.C.M., Richard, C., 2020. Kalman Filtering and Expectation Maximization for Multitemporal Spectral Unmixing. IEEE Geosci. Remote Sens. Lett. 1–5. https://doi.org/10.1109/LGRS.2020.3025781.

Borsoi, R.A., Imbiriba, T., Bermudez, J.C.M., 2020. Deep Generative Endmember Modeling: An Application to Unsupervised Spectral Unmixing. IEEE Trans. Comput. Imaging. 6, 374–384. https://doi.org/10.1109/TCI.674585210.1109/TCI.2019.2948726.

Borsoi, R.A., Imbiriba, T., Bermudez, J.C.M., Richard, C., 2020. A Blind Multiscale Spatial Regularization Framework for Kernel-Based Spectral Unmixing. IEEE Trans. Image Process. 29, 4965–4979. https://doi.org/10.1109/TIP.8310.1109/TIP.2020.2978342.

Bullock, E.L., Woodcock, C.E., Olofsson, P., 2020. Monitoring tropical forest degradation using spectral unmixing and Landsat time series analysis. Remote Sens. Environ. 238, 110968. https://doi.org/10.1016/j.rse.2018.11.011.

Cao, X., Fu, X., Xu, C., Meng, D., 2021. Deep Spatial-Spectral Global Reasoning Network for Hyperspectral Image Denoising. IEEE Trans. Geosci. Remote Sens. https://doi.org/10.1109/TGRS.2021.3069241.

Chang, W.S., 2006. Chein-I, Constrained Band Selection for Hyperspectral Imagery. IEEE Trans. Geosci. Remote Sens. 44, 1575–1585. https://doi.org/10.1109/LGRS.2018.2830795.

Chang, Yi, Yan, Luxin, Wu, Tao, Zhong, Sheng, 2016. Remote Sensing Image Stripe Noise Removal: From Image Decomposition Perspective. IEEE Trans. Geosci. Remote Sens. 54 (12), 7018–7031. https://doi.org/10.1109/TGRS.2016.2594080.

T. Cocks, R. Jenssen, A. Stewart, I. Wilson, T. Shields, THE HYMAP TM AIRBORNE HYPERSPECTRAL SENSOR: THE SYSTEM, CALIBRATION AND PERFORMANCE, 1998.

M. Cuturi, M. Blondel, Soft-DTW: a Differentiable Loss Function for Time-Series, in: 34th Int. Conf. Mach. Learn. ICML 2017, International Machine Learning Society (IMLS), 2017: pp. 1483–1505. http://arxiv.org/abs/1703.01541 (accessed November 24, 2020).

Dou, Z., Gao, K., Zhang, X., Wang, H., Wang, J., 2020. Hyperspectral unmixing using orthogonal sparse prior-based autoencoder with hyper-laplacian loss and data-driven outlier detection. IEEE Trans. Geosci. Remote Sens. 58 (9), 6550–6564. https://doi.org/10.1109/TGRS.3610.1109/TGRS.2020.2977819.

S. Falkner, A. Klein, F. Hutter, BOHB: Robust and Efficient Hyperparameter Optimization at Scale, PMLR, 2018. http://proceedings.mlr.press/v80/falkner18a.html (accessed April 12, 2021).

Fan, F., Xiong, J., Wang, G., 2020. On Interpretability of Artificial Neural Networks. ArXiv. http://arxiv.org/abs/2001.02522.

Fang, Qian, Hong, Hanlie, Zhao, Lulu, Kukolich, Stephanie, Yin, Ke, Wang, Chaowen, 2018. Visible and Near-Infrared Reflectance Spectroscopy for Investigating Soil Mineralogy: A Review. J. Spectrosc. 2018, 1–14. https://doi.org/10.1155/2018/3168974.

Gao, L., Han, Z., Hong, D., Zhang, B., Chanussot, J., 2021. CyCU-Net: Cycle-Consistency Unmixing Network by Learning Cascaded Autoencoders. IEEE Trans. Geosci. Remote Sens. https://doi.org/10.1109/TGRS.2021.3064958.

Ghamisi, Pedram, Yokoya, Naoto, Li, Jun, Liao, Wenzhi, Liu, Sicong, Plaza, Javier, Rasti, Behnood, Plaza, Antonio, 2017. Advances in Hyperspectral Image and Signal Processing: A Comprehensive Overview of the State of the Art. IEEE Geosci. Remote Sens. Mag. 5 (4), 37–78. https://doi.org/10.1109/MGRS.2017.2762087.

Graham, G.E., Kokaly, R.F., Kelley, K.D., Hoefen, T.M., Johnson, M.R., Hubbard, B.E., 2018. Application of imaging spectroscopy for mineral exploration in Alaska: A study over porphyry Cu deposits in the eastern Alaska Range. Econ. Geol. 113, 489–510. https://doi.org/10.5382/econgeo.2018.4559.

Han, H., Wang, G., Wang, M., Miao, J., Guo, S.i., Chen, L., Zhang, M., Guo, K.e., 2020. Hyperspectral Unmixing Via Nonconvex Sparse and Low-Rank Constraint. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 5704–5718. https://doi.org/10.1109/JSTARS.460944310.1109/JSTARS.2020.3021520.

Herrmann, I., Shapira, U., Kinast, S., Karnieli, A., Bonfil, D.J., 2013. Ground-level hyperspectral imagery for detecting weeds in wheat fields. Precis. Agric. 14 (6), 637–659. https://doi.org/10.1007/s11119-013-9321-x.

Heylen, R., Parente, M., Gader, P., 2014. A review of nonlinear hyperspectral unmixing methods. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7 (6), 1844–1868. https://doi.org/10.1109/JSTARS.460944310.1109/JSTARS.2014.2320576.

Hong, D., Gao, L., Yao, J., Yokoya, N., Chanussot, J., Heiden, U., Zhang, B., 2021. Endmember-Guided Unmixing Network (EGU-Net): A General Deep Learning Framework for Self-Supervised Hyperspectral Unmixing. IEEE Trans. Neural Networks Learn. Syst. https://doi.org/10.1109/TNNLS.2021.3082289.

Hua, Z., Li, X., Feng, Y., Zhao, L., 2021. Dual Branch Autoencoder Network for Spectral-Spatial Hyperspectral Unmixing. IEEE Geosci. Remote Sens. Lett. 1–5. https://doi.org/10.1109/LGRS.2021.3091858.

Iwana, Brian Kenji, Frinken, Volkmar, Uchida, Seiichi, 2020. DTW-NN: A novel neural network for time series recognition using dynamic alignment between inputs and weights. Knowledge-Based Syst. 188, 104971. https://doi.org/10.1016/j.knosys.2019.104971.

Jiang, Y., Li, H., Rangaswamy, M., 2019. Deep Learning Denoising Based Line Spectral Estimation. IEEE Signal Process. Lett. 26 (11), 1573–1577. https://doi.org/10.1109/LSP.9710.1109/LSP.2019.2939049.

Jiang, T.X., Zhuang, L., Huang, T.Z., Le Zhao, X., Bioucas-Dias, J.M., 2021. Adaptive Hyperspectral Mixed Noise Removal. IEEE Trans. Geosci. Remote Sens. https://doi.org/10.1109/TGRS.2021.3085779.

Julien, Y., Sobrino, J.A., 2019. Optimizing and comparing gap-filling techniques using simulated NDVI time series from remotely sensed global data. Int. J. Appl. Earth Obs. Geoinf. 76, 93–111. https://doi.org/10.1016/j.jag.2018.11.008.

A. Karami, R. Heylen, P. Scheunders, Hyperspectral image noise reduction and its effect on spectral unmixing, Work. Hyperspectral Image Signal Process. Evol. Remote Sens. 2014-June (2014). https://doi.org/10.1109/WHISPERS.2014.8077632.

Khajehrayeni, F., Ghassemian, H., 2020. Hyperspectral Unmixing Using Deep Convolutional Autoencoders in a Supervised Scenario. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 567–576. https://doi.org/10.1109/JSTARS.460944310.1109/JSTARS.2020.2966512.

R.F. Kokaly, M.R. Johnson, G.E. Graham, T.M. Hoefen, K.D. Kelley, B.E. Hubbard, Imaging spectrometer reflectance data for Nabesna, Alaska - ScienceBase-Catalog, 2018. https://www.sciencebase.gov/catalog/item/59ee7cc9e4b0220bbd976362 (accessed February 14, 2021).

Koziol, P., Raczkowska, M.K., Skibinska, J., Urbaniak-Wasik, S., Paluszkiewicz, C., Kwiatek, W., Wrobel, T.P., 2018. Comparison of spectral and spatial denoising techniques in the context of High Definition FT-IR imaging hyperspectral data. Sci. Rep. 8 (1) https://doi.org/10.1038/s41598-018-32713-7.

Li, S., Song, W., Fang, L., Chen, Y., Ghamisi, P., Benediktsson, J.A., 2019. Deep learning for hyperspectral image classification: An overview. IEEE Trans. Geosci. Remote Sens. 57 (9), 6690–6709. https://doi.org/10.1109/TGRS.3610.1109/TGRS.2019.2907932.

S. Liu, J. Feng, Z. Tian, Variational Low-Rank Matrix Factorization with Multi-Patch Collaborative Learning for Hyperspectral Imagery Mixed Denoising, Remote Sens. 2021, Vol. 13, Page 1101. 13 (2021) 1101. https://doi.org/10.3390/RS13061101.

K.E. Livo, R.N. Clark, Open-File Report 2013-1300 The Tetracorder User Guide-Version 4.4, 2014. https://doi.org/10.3133/ofr20131300.

J. Mao, X. Wang, H. Li, Interpolated Convolutional Networks for 3D Point Cloud Understanding, Proc. IEEE Int. Conf. Comput. Vis. 2019-October (2019) 1578–1587. http://arxiv.org/abs/1908.04512 (accessed October 26, 2020).

Mei, S., Zhang, G.e., Li, J., Zhang, Y., Du, Q., 2020. Improving Spectral-Based Endmember Finding by Exploring Spatial Context for Hyperspectral Unmixing. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 3336–3349. https://doi.org/10.1109/JSTARS.460944310.1109/JSTARS.2020.3003456.

Ozkan, S., Kaya, B., Akar, G.B., 2019. EndNet: Sparse AutoEncoder Network for Endmember Extraction and Hyperspectral Unmixing. IEEE Trans. Geosci. Remote Sens. 57 (1), 482–496. https://doi.org/10.1109/TGRS.2018.2856929.

Palsson, B., Ulfarsson, M.O., Sveinsson, J.R., 2019. Convolutional Autoencoder for Spatial-Spectral Hyperspectral Unmixing, in. In: Institute of Electrical and

Electronics Engineers (IEEE), pp. 357–360. https://doi.org/10.1109/igarss.2019.8900297.
Pan, Z., Huang, J., Wang, F., 2013. Multi range spectral feature fitting for hyperspectral imagery in extracting oilseed rape planting area. Int. J. Appl. Earth Obs. Geoinf. 25, 21–29. https://doi.org/10.1016/j.jag.2013.03.002.
Qi, L., Li, J., Wang, Y., Lei, M., Gao, X., 2020. Deep spectral convolution network for hyperspectral image unmixing with spectral library. Signal Processing. 176, 107672. https://doi.org/10.1016/j.sigpro.2020.107672.
Y. Qian, F. Xiong, Q. Qian, J. Zhou, Spectral Mixture Model Inspired Network Architectures for Hyperspectral Unmixing, IEEE Trans. Geosci. Remote Sens. 58 (2020) 7418–7434. https://doi.org/10.1109/tgrs.2020.2982490.
Qian, Y., Xiong, F., Qian, Q., Zhou, J., 2020. Spectral Mixture Model Inspired Network Architectures for Hyperspectral Unmixing. IEEE Trans. Geosci. Remote Sens. 58 (10), 7418–7434. https://doi.org/10.1109/TGRS.3610.1109/TGRS.2020.2982490.
B. Rasti, B. Koirala, SUnCNN: Sparse Unmixing Using Unsupervised Convolutional Neural Network, IEEE Geosci. Remote Sens. Lett. (2021) 1–5. https://doi.org/10.1109/LGRS.2021.3100992.
Rasti, B., Scheunders, P., Ghamisi, P., Licciardi, G., Chanussot, J., 2018. Noise Reduction in Hyperspectral Imagery: Overview and Application. Remote Sens. 10, 482. https://doi.org/10.3390/rs10030482.
Ren, Longfei, Ma, Zheng, Bovolo, Francesca, 2021. A Novel Dual-Alternating Direction Method of Multipliers for Spectral Unmixing. IEEE Geosci. Remote Sens. Lett. 18 (3), 528–532. https://doi.org/10.1109/LGRS.885910.1109/LGRS.2020.2980429.
Richter, D.H., Jones, D.L., 1973. Reconnaissance geologic map of the Nabesna A-2 Quadrangle. Alaska. https://doi.org/10.3133/i749.
Rossman, G.R., Ehlmann, B.L., 2019. Electronic Spectra of Minerals in the Visible and Near-Infrared Regions. In: Remote Compos. Anal. Cambridge University Press, pp. 3–20. https://doi.org/10.1017/9781316888872.003.
S. Sabour, N. Frosst, G.E. Hinton, Dynamic Routing Between Capsules, Adv. Neural Inf. Process. Syst. 2017-December (2017) 3857–3867. http://arxiv.org/abs/1710.09829 (accessed October 26, 2020).
Salamat, Nadeem, Missen, Malik Muhammad Saad, Surya Prasath, V.B., 2021. Recent developments in computational color image denoising with PDEs to deep learning: a review. Artif. Intell. Rev. 54 (8), 6245–6276. https://doi.org/10.1007/s10462-021-09977-z.
W. Samek, G. Montavon, S. Lapuschkin, C.J. Anders, K.R. Müller, Toward interpretable machine learning: Transparent deep neural networks and beyond, ArXiv. (2020). http://arxiv.org/abs/2003.07631.
Shah, D., Trivedi, Y.N., Zaveri, T., 2020. Non-linear spectral unmixing: A case study on mangalore aviris-ng hyperspectral data, 2020 IEEE Bombay Sect. Signat. Conf. IBSSC 2020, 11–15. https://doi.org/10.1109/IBSSC51096.2020.9332215.
Shahid, K.T., Schizas, I.D., 2021. Unsupervised Hyperspectral Unmixing via Nonlinear Autoencoders. IEEE Trans. Geosci. Remote Sens. https://doi.org/10.1109/TGRS.2021.3077833.
Shankar, V., 2015. Field Characterization by Near Infrared (NIR) Mineral Identifiers- A New Prospecting Approach. Procedia Earth Planet. Sci. 11, 198–203. https://doi.org/10.1016/j.proeps.2015.06.025.
S.M. Shankaranarayana, D. Runje, ALIME: Autoencoder Based Approach for Local Interpretability, Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics). 11871 LNCS (2019) 454–463. http://arxiv.org/abs/1909.02437 (accessed November 14, 2020).
C. Shekhar, On simplified application of multidimensional Savitzky-Golay filters and differentiators, in: AIP Conf. Proc., American Institute of Physics Inc., 2016: p. 020014. https://doi.org/10.1063/1.4940262.
Shi, Q., Tang, X., Yang, T., Liu, R., Zhang, L., 2021. Hyperspectral Image Denoising Using a 3-D Attention Denoising Network. IEEE Trans. Geosci. Remote Sens. https://doi.org/10.1109/TGRS.2020.3045273.
Signoroni, A., Savardi, M., Baronio, A., Benini, S., 2019. Deep Learning Meets Hyperspectral Image Analysis: A Multidisciplinary Review. J. Imaging. 5, 52. https://doi.org/10.3390/jimaging5050052.
K. Simonyan, A. Vedaldi, A. Zisserman, Deep inside convolutional networks: Visualising image classification models and saliency maps, in: 2nd Int. Conf. Learn. Represent. ICLR 2014 - Work. Track Proc., 2014.
Song, J., Xia, S., Wang, J., Patel, M., Chen, D., 2021. Uncertainty Quantification of Hyperspectral Image Denoising Frameworks Based on Sliding-Window Low-Rank Matrix Approximation. IEEE Trans. Geosci. Remote Sens. https://doi.org/10.1109/TGRS.2021.3065570.
Su, Y., Li, J., Plaza, A., Marinoni, A., Gamba, P., Chakravortty, S., 2019. DAEN: Deep Autoencoder Networks for Hyperspectral Unmixing. IEEE Trans. Geosci. Remote Sens. 57 (7), 4309–4321. https://doi.org/10.1109/TGRS.3610.1109/TGRS.2018.2890633.
Sun, Qiaoqiao, Liu, Xuefeng, Bourennane, Salah, Liu, Bin, 2021. Multiscale denoising autoencoder for improvement of target detection. Int. J. Remote Sens. 42 (8), 3002–3016. https://doi.org/10.1080/01431161.2020.1856960.
Swayze, G.A., Clark, R.N., Goetz, A.F.H., Livo, K.E., Breit, G.N., Kruse, F.A., Sutley, S.J., Snee, L.W., Lowers, H.A., Post, J.L., Stoffregen, R.E., Ashley, R.P., 2014. Mapping advanced argillic alteration at Cuprite, Nevada, using imaging spectroscopy. Econ. Geol. 109 (5), 1179–1221. https://doi.org/10.2113/econgeo.109.5.1179.
Tjoa, Erico, Guan, Cuntai, 2021. A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI. IEEE Trans. Neural Networks Learn. Syst. 32 (11), 4793–4813. https://doi.org/10.1109/TNNLS.2020.3027314.
van der Meer, Freek, 2004. Analysis of spectral absorption features in hyperspectral imagery. Int. J. Appl. Earth Obs. Geoinf. 5 (1), 55–68. https://doi.org/10.1016/j.jag.2003.09.001.
Wang, Chunying, Liu, Baohua, Liu, Lipeng, Zhu, Yanjun, Hou, Jialin, Liu, Ping, Li, Xiang, 2021. A review of deep learning used in the hyperspectral image analysis for agriculture. Artif. Intell. Rev. 54 (7), 5205–5253. https://doi.org/10.1007/s10462-021-10018-y.
Wang, X., Luo, Z., Li, W., Hu, X., Zhang, L., Zhong, Y., 2021. A Self-Supervised Denoising Network for Satellite-Airborne-Ground Hyperspectral Imagery. IEEE Trans. Geosci. Remote Sens. https://doi.org/10.1109/TGRS.2021.3064429.
F. Xiong, J. Zhou, M. Ye, J. Lu, Y. Qian, NMF-SAE: An Interpretable Sparse Autoencoder for Hyperspectral Unmixing, in: Institute of Electrical and Electronics Engineers (IEEE), 2021: pp. 1865–1869. https://doi.org/10.1109/icassp39728.2021.9414084.
Xiong, F., Zhou, J., Tao, S., Lu, J., Qian, Y., 2021. SNMF-Net: Learning a Deep Alternating Neural Network for Hyperspectral Unmixing. IEEE Trans. Geosci. Remote Sens. https://doi.org/10.1109/TGRS.2021.3081177.
Xu, Y., Peng, Y., 2021. (2021). Joint Spatial-Spectral Encoder for Hyperspectral Image Denoising, 2021 IEEE 3rd Int. Conf. Commun. Inf. Syst. Comput. Eng. CISCE 524–528. https://doi.org/10.1109/CISCE52179.2021.9446024.
Q. Yuan, Q. Zhang, J. Li, H. Shen, L. Zhang, Hyperspectral image denoising employing a spatial-spectral deep residual convolutional neural network, IEEE Trans. Geosci. Remote Sens. 57 (2019) 1205–1218. https://doi.org/10.1109/TGRS.2018.2865197.
Yuan, Y., Ma, H., Liu, G., 2021. Partial-DNet: A Novel Blind Denoising Model With Noise Intensity Estimation for HSI. IEEE Trans. Geosci. Remote Sens. https://doi.org/10.1109/TGRS.2021.3071799.
Yuan, Qiangqiang, Zhang, Qiang, Li, Jie, Shen, Huanfeng, Zhang, Liangpei, 2019. Hyperspectral image denoising employing a spatial-spectral deep residual convolutional neural network. IEEE Trans. Geosci. Remote Sens. 57 (2), 1205–1218. https://doi.org/10.1109/TGRS.2018.2865197.
Yuan, Y., Zhang, Z., Wang, Q.i., 2020. Improved Collaborative Non-Negative Matrix Factorization and Total Variation for Hyperspectral Unmixing. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 998–1010. https://doi.org/10.1109/JSTARS.460944310.1109/JSTARS.2020.2977399.
Zhang, H., Cai, J., He, W., Shen, H., Zhang, L., 2021. Double Low-Rank Matrix Decomposition for Hyperspectral Image Denoising and Destriping. IEEE Trans. Geosci. Remote Sens. https://doi.org/10.1109/TGRS.2021.3061148.
Zhang, Z., Duan, F., Sole-Casals, J., Dinares-Ferran, J., Cichocki, A., Yang, Z., Sun, Z., 2019. A Novel Deep Learning Approach with Data Augmentation to Classify Motor Imagery Signals. IEEE Access. 7, 15945–15954. https://doi.org/10.1109/ACCESS.2019.2895133.
Zhang, F., Zhang, K., Sun, J., Feng, Z., Wan, W., Wu, Q., 2021. Hyperspectral Image Denoising Using 3-D Geometrical Kernel with Local Similarity Prior. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 3303–3317. https://doi.org/10.1109/JSTARS.2021.3064243.
A.X. Zhao, X.J. Tang, Z.H. Zhang, J.H. Liu, The parameters optimization selection of Savitzky-Golay filter and its application in smoothing pretreatment for FTIR spectra, in: Proc. 2014 9th IEEE Conf. Ind. Electron. Appl. ICIEA 2014, Institute of Electrical and Electronics Engineers Inc., 2014: pp. 516–521. https://doi.org/10.1109/ICIEA.2014.6931218.
Y. Zhao, J. Yang, C. Yi, Y. Liu, Joint denoising and unmixing for hyperspectral image, Work. Hyperspectral Image Signal Process. Evol. Remote Sens. 2014-June (2014). https://doi.org/10.1109/WHISPERS.2014.8077483.
Zhao, M., Wang, X., Chen, J., Chen, W., 2021. A Plug-and-Play Priors Framework for Hyperspectral Unmixing. IEEE Trans. Geosci. Remote Sens. https://doi.org/10.1109/TGRS.2020.3047479.
Zhao, M., Wang, M., Chen, J., Rahardja, S., 2021. Hyperspectral Unmixing for Additive Nonlinear Models With a 3-D-CNN Autoencoder Network. IEEE Trans. Geosci. Remote Sens. https://doi.org/10.1109/TGRS.2021.3098745.