UnDIP: Hyperspectral Unmixing Using Deep Image Prior

Behnood Rasti, *Senior Member, IEEE,* Bikram Koirala, *Student Member, IEEE,*
Paul Scheunders, *Senior Member, IEEE,* and Pedram Ghamisi, *Senior Member, IEEE*

***Abstract*—In this paper, we introduce a deep learning-based technique for the linear hyperspectral unmixing problem. The proposed method contains two main steps. First, the endmembers are extracted using a geometric endmember extraction method, i.e., a simplex volume maximization in the subspace of the dataset. Then, the abundances are estimated using a deep image prior. The main motivation of this work is to boost the abundance estimation and to make the unmixing problem robust to noise. The proposed deep image prior uses a convolutional neural network to estimate the fractional abundances, relying on the extracted endmembers and the observed hyperspectral dataset. The proposed method is evaluated on simulated and three real remote sensing data for a range of SNR values (i.e., from 20 to 50 dB). The results show considerable improvements compared to state-of-the-art methods. The proposed method was implemented in Python (3.8) using PyTorch as the platform for the deep network and is available online: https://github.com/BehnoodRasti/UnDIP.**

***Index Terms*—Hyperspectral image, unmixing, convolutional neural network, deep learning, deep prior, endmember extraction**

I. Introduction

SPECTRAL unmixing is one of the major hyperspectral image analysis tasks. Hyperspectral cameras have the ability to capture the spectral signature of materials. This ability allows us to distinguish different materials within a scene. However, due to the limited spatial resolution and scattering of the light, a pixel spectrum is generally a complex mixture of the pure spectra of its constituent materials, i.e. the endmember spectra [1], [2]. Unmixing is the task of estimating the fractional abundances of those endmembers within the spectral pixels. From a modeling point of view, unmixing techniques can be divided into two main groups: linear and nonlinear unmixing [3], [4].

In linear unmixing, it is assumed that the light only interacts with one pure material before reaching the sensor. In remote sensing applications, hyperspectral images are of low spatial resolution, and pixels typically contain large homogeneous regions of single materials. For this situation, the linear mixture model is a good approximation [3]. In microscopic scenarios (i.e. close-range scenarios), the pure materials are intimately mixed within the pixel, and the light undergoes multiple reflections by several materials. In these situations, the linear approximation often fails and one has to apply nonlinear models [3].

In this paper, we aim at remote sensing applications and focus on the linear hyperspectral unmixing methods. The linear unmixing methods can be categorized as unsupervised, supervised, and semi-supervised. Unsupervised methods either sequentially extract the endmembers from the image and then estimate the fractional abundances, or simultaneously estimate both endmembers and abundances from the image (so-called blind unmixing). Supervised methods only estimate the abundances from the image assuming that endmembers are known a priori. If not known a priori, the endmembers need to be extracted from a large endmember library and one refers to semi-supervised (so-called sparse unmixing). In the latter case, the number of endmembers need not to be known a priori [1], [3].

Endmembers can be extracted from hyperspectral images based on geometrical principles. This can be done by relying on either the existence of pure spectra for each material, located at the vertices of the data simplex or the existence of sufficient spectra on the facets of the data simplex, to allow to geometrically locate the vertices of the data simplex. Approaches such as pixel purity index (PPI) [5], N-FINDR [6], and the vertex component analysis (VCA) algorithm [7] use geometrical concepts for endmember extraction. After endmember extraction, the abundance fractions are generally estimated by using optimization algorithms such as non-negative constrained least-squares [8], satisfying the abundance non-negativity constraint (ANC) or fully constrained least-squares [9], satisfying both ANC and the abundance sum-to-one constraint (ASC). This step is also referred to as inversion in the literature [3].

In blind unmixing, both endmembers and abundances are estimated simultaneously. Two major paradigms in blind unmixing are constrained penalized (or regularized) least squares (CPLS) methods, such as [10] and statistical approaches such as [11]. Examples of CPLS algorithms are minimum volume simplex analysis [12], simplex identification via variable splitting and augmented Lagrangian (SISAL), and collaborative nonnegative matrix factorization (CoNMF) [13]. They often solve a penalized least-squares problem, subject to (either or both) ASC and ANC. These algorithms often involve a data-

Behnood Rasti (corresponding author) is with Helmholtz-Zentrum Dresden-Rossendorf, Helmholtz Institute Freiberg for Resource Technology, Machine Learning Group, Chemnitzer Straße 40, 09599 Freiberg, Germany; b.rasti@hzdr.de; p.ghamisi@hzdr.de.

Bikram Koirala and Paul Scheunders are with Imec-Visionlab, University of Antwerp (CDE) Universiteitsplein 1, B-2610 Antwerp, Belgium; Bikram.Koirala@uantwerpen.be; paul.scheunders@uantwerpen.be

Pedram Ghamisi is with (1) Helmholtz-Zentrum Dresden-Rossendorf, Helmholtz Institute Freiberg for Resource Technology, Machine Learning Group, Chemnitzer Straße 40, 09599 Freiberg, Germany; (2) Institute of Advanced Research in Artificial Intelligence (IARAI), Landstraßer Hauptstraße 5, 1030 Vienna, Austria; p.ghamisi@gmail.com.

Manuscript received .....

fitting term and a minimum volume-based regularization term. A major issue with these algorithms is that the regularization parameter needs to be tuned. In [14], a geometrical constraint (the squared of the simplex volume) was enforced as a regularizer to the fully constrained least squares problem to simultaneously estimate the abundances and endmembers. In [15], the regularization parameter for the minimum volume-based regularization term was automatically selected by determining the simplex which encloses the whole data. The statistical approaches often formulate the unmixing problem in a Bayesian way and use different estimators, such as the joint maximum a posteriori (MAP) estimator in [16]. It is worth mentioning that both groups are related, as a CPLS can be derived using a MAP estimator [17]. Due to the inherent nonconvexity of blind unmixing methods, they are highly vulnerable to the initialization and therefore they are always initialized using a geometrical endmember extraction approach.

In sparse unmixing, the fractional abundances are estimated using sparse regression techniques. These methods describe each spectrum as a sparse linear combination of the elements of a rich library of pure spectra, a problem that can be generally formulated using CPLS. This results in either a convex or a non-convex problem, depending on the selected sparsity promoting penalty to be applied on the abundances [18]. Sparse unmixing by variable splitting and augmented Lagrangian (SUnSAL), constrained SUnSAL (C-SUnSAL) [19] and collaborative sparse unmixing [20] are examples of sparse unmixing methods. Both SUnSAL and C-SUnSAL apply an ℓ_1 penalty on the fractional abundances. SUnSAL utilizes ℓ_2 for the fidelity term while C-SUnSAL assumes a constraint to enforce the data fidelity. Collaborative sparse unmixing is similar to SUnSAL, but applies $\ell_{2,1}$ (i.e., the sum of ℓ_2 on the abundances) to promote the sparsity on the abundances. SUnSAL was improved in [21] by incorporating spatial information through applying a total variation penalty on the abundances (SUnSAL-TV).

The spectral variability of the endmembers (i.e. the intra-class variability of the materials) is taken into account by using a dictionary of endmember bundles, generated from the data (as opposed to the above-mentioned sparse regression-based techniques where the dictionary is made from spectral libraries and does not rely on the data itself). When using endmember bundles, four different penalties: the group LASSO (Least Absolute Shrinkage and Selection Operator) [22], the collaborative LASSO [23], the elitist LASSO [24], and the fractional LASSO [25]) were proposed in the framework of sparse regression in [25], where all take the ASC into account. The main difference between those techniques is the selection of the penalty term applied on the abundances.

Deep learning-based networks are state-of-the-art in machine learning and computer vision applications. Inevitably, most of the remote sensing applications, involving machine learning and image processing have been inspired by deep networks [26]. Recently, a variety of deep neural networks has been proposed for hyperspectral unmixing, mainly based on variations of deep encoder-decoder networks, for which the inputs are the spectra and the outputs are the abundances. The abundances are then decoded to the spectra again using linear layers, with the endmembers as the weights. EndNet [27], SNSA [28], DAEN [29], DeepGUn [30], and uDAS [31] are a few examples of such unmixing techniques. EndNet proposes a loss function with several terms, including a Kullback-Leibler divergence term, a SAD similarity, and a sparsity term which makes the parameter selection very challenging. SNSA utilizes a stack of nonnegative sparse autoencoders from which the last one performs the task of unmixing and the others are exploited to improve the robustness with respect to the outliers. DAEN exploits a stacked autoencoder to initialize a variational autoencoder which performs the unmixing task. In [30], a variational autoencoder is used to generate the endmembers. uDAS exploits an additional denoising constraint on the decoder and a $\ell_{2,1}$ sparsity constraint on the decoder. In all these methods, the spatial information is ignored.

The advantage of incorporating the spatial information for spectral unmixing has been confirmed in the literature. Training a network based on a single spectrum at a time ignores the spatial information. Therefore, patch-wise or cube-wise CNN was proposed to utilize the spatial information. First, the image was spatially divided into patches and then the convolution is applied on small patches of spectra. In [32], it was shown that cube-wise CNN outperforms pixel-wise CNN. In [33], spatial information has been exploited for unmixing, by improving the encoder-decoder architecture proposed in [34], by applying parallel encoder-decoders on HSI patches. In [35], a CNN was proposed based on a spatial-spectral model, which is trained using HSI patches. Most recently, a convolutional autoencoder was proposed for supervised hyperspectral unmixing in [36], exploiting 3D convolutional filters. The patchwise approach was found useful for endmember estimation since it supports the idea of endmember bundles and captures the variability of the spectra. However, it degrades (and blurs) the estimated abundances [35], since small patches do not contain enough structure for the convolutions (filters) to perform better than merely a mean filter.

The supervised CNN exploited in the above-mentioned techniques requires spectral signatures to train the CNN. In this paper, we propose an unsupervised CNN that does not need spectral signatures for training. The convolutional encoder-decoder network proposed in this paper is a more general network than the autoencoders often used in the literature, in the sense that the input can have any distribution regardless of the output.

A. Contributions and Novelties

The main motivation of this work is to improve the abundance estimation and to make the unmixing problem robust to noise. Hence, we propose a method called "hyperspectral unmixing using deep image prior" (UnDIP), which utilizes a conventional geometrical approach for endmember extraction and a new unmixing deep image prior using a deep convolutional neural network for abundance estimation. The main novelty of this paper is the introduction of a new unmixing deep prior for the inversion task. Deep image prior (DIP) was recently proposed for conventional inverse problems in the

area of image processing, such as denoising, inpainting, and super resolution [37], [38]. In [39], DIP was applied for hyperspectral image denoising, inpainting, and super resolution. In this work, the DIP is adjusted to the unmixing problem to generate fractional abundances. Starting from input noise, the abundances are generating by iteratively minimizing an implicitly regularized loss function. The proposed network is applicable in supervised unmixing scenarios, where the endmembers are available.

UnDIP has the following attributes that distinguish it from the other deep learning-based unmixing techniques proposed in the literature:

- It uses DIP as a deep learning procedure. UnDIP is designed to solve a regularized inverse problem, in which the regularizer is implicitly incorporated in the cost function. This controls the overfitting of the fidelity term and makes the method robust to noise.
- It incorporates spatial information globally, unlike the pixelwise or patchwise (convolutional) autoencoder-based approaches in the literature. UnDIP does not need spectral signatures for training. The input of the network has the same spatial size as the observed image and is given by Gaussian noise which is fixed throughout the learning process. Then, the network iteratively learns to map that input to abundance maps. This unsupervised learning framework has the advantage that the convolutional network can be applied globally on the entire spatial domain of an image, which leads to sharper abundance maps and enhances the robustness to noise.
- It combines a geometrical endmember estimation approach with deep unmixing. The majority of the proposed blind unmixing techniques, including deep techniques need to be initialized by a geometrical endmember estimation approach, confirming the importance of this step. Here, for the first time, UnDIP proposes a collaborative framework in which a geometrical endmember estimation is performed prior to the deep unmixing. The endmembers are then used in the loss function for training the deep network. In this way, the deep network can focus on the improvement of the abundance estimation, while the endmembers remain fixed.

The remaining of this paper is organized as follows. The unmixing methodology is explained in detail in Section II. The experimental results are shown and discussed in Section III. Section IV concludes the paper.

II. Methodology

A. Notation

Before discussing the proposed methodology we explain the notations used in the paper. Matrices, column vectors, and scalars are denoted in bold and capital letters, bold letters, and letters, respectively. $\hat{\mathbf{X}}$ represents the estimate of the variable $\mathbf{X}$. $\|.\|_F$ and $|.|$ denote the Frobenius norm and the absolute value. $\mathbf{x}_{(i)}$ and $\mathbf{x}_i^T$ denote the ith column and the ith row of matrix $\mathbf{X}$, respectively. $\mathbf{X}_{ij}$ denotes the matrix element located at ith row and the jth column. $\mathbf{1}_n$ is an n-component column vector of ones. The notation $(r)!$ denotes the factorial of the positive integer r and det($\mathbf{X}$) indicates the determinant of matrix $\mathbf{X}$.

B. Hyperspectral Modeling

We assume a linear model for HSI:

$$\mathbf{Y} = \mathbf{X} + \mathbf{N}, \tag{1}$$

where $\mathbf{Y} \in \mathbb{R}^{p\times n}$ is the observed HSI, with n pixels and p bands, $\mathbf{X} \in \mathbb{R}^{p\times n}$ is the unknown image to be estimated, and $\mathbf{N} \in \mathbb{R}^{p\times n}$ is the model error, including noise. In spectral unmixing, we assume that:

$$\mathbf{Y} = \mathbf{E}\mathbf{A} + \mathbf{N}, \tag{2}$$

where $\mathbf{E} \in \mathbb{R}^{p\times r}$and $\mathbf{A} \in \mathbb{R}^{r\times n}$, $r \ll p$, contain the r endmembers and their fractional abundances, respectively. The main goal is to estimate the fractional abundances $\mathbf{A}$, however, this is not possible without either having prior knowledge about the endmembers $\mathbf{E}$ or estimating/extracting them from the image.

C. Endmember Extraction

When the endmembers are extracted from the data, one often relies on the geometry of the data. Due to spectral redundancy, an HSI often lives in a low-dimensional subspace [40], [41]. Therefore, the data can be projected onto an (r-1)-dimensional subspace and represented by a (r-1)-dimensional simplex whose vertices are the endmembers $\mathbf{e}_m$ ($m = 1, \cdots, r$). When pure spectra are available in the data, the endmembers can be extracted by maximizing the volume of the data simplex [42]:

$$\arg\max_{\mathbf{E}} V(\mathbf{E}) = \arg\max_{\mathbf{E}} \frac{\mathbf{1}}{(\mathbf{r}-\mathbf{1})!} \left|\det \begin{bmatrix} 1 & \dots & 1 \\ \mathbf{e}_{(1)} & \dots & \mathbf{e}_{(\mathbf{r})} \end{bmatrix}\right|, \tag{3}$$

where $\mathbf{E} = [\mathbf{e}_{(i)}]$. In this paper, we use an algorithm, called simplex volume maximization (SiVM) [43] to extract the endmembers from the dataset. SiVM selects the endmembers by iteratively maximizing the simplex volume of the data:

$$\arg\max_{\mathbf{E}} V(\mathbf{E}) = \arg\max_{\mathbf{E}} \sqrt{\frac{(-\mathbf{1})^{\mathbf{r}} \cdot \text{ cmd } (\mathbf{E})}{\mathbf{2}^{\mathbf{r-1}}(\mathbf{r}-\mathbf{1})!}}, \tag{4}$$

where cmd is the Cayley–Menger determinant:

$$\text{cmd } (\mathbf{E}) = \det \begin{bmatrix} 0 & 1 & 1 & 1 & \dots & 1 \\ 1 & 0 & d_{1,2}^2 & d_{1,3}^2 & \dots & d_{1,r}^2 \\ 1 & d_{2,1}^2 & 0 & d_{2,3}^2 & \dots & d_{2,r}^2 \\ 1 & d_{3,1}^2 & d_{3,2}^2 & 0 & \dots & d_{3,r}^2 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & d_{r,1}^2 & d_{r,2}^2 & d_{r,3}^2 & \dots & 0 \end{bmatrix},$$

and $d_{i,j}^2$ is the Euclidean distance between endmembers $\mathbf{e}_i$ and $\mathbf{e}_j$. Since (4) does not take into account nuisances such as noise, we first project the data on the subspace obtained by the spectral eigenvectors of a singular value decomposition.

D. Deep Image Prior

In this section, we first explain the general concept of deep image prior and in the next section, we adapt this concept to the unmixing problem. CNNs are the most popular deep learning networks for inverse problems such as image restoration. They show excellent performances as long as a large training dataset is available.

Recently, DIP was proposed as an unsupervised deep learning alternative, in which the network is entirely trained based on the observed image. DIP generates an image $\mathbf{X}$ using a random initialization $\mathbf{Z}$ and utilizing the deep network as a parametric function $\mathbf{X} = f_\theta(\mathbf{Z})$. Then, the network is optimized over its parameters (i.e., θ) to generate the optimal image $\hat{\mathbf{X}} = f_{\hat{\theta}}(\mathbf{Z})$.

Generally, inverse image reconstruction tasks, such as denoising, super resolution, and inpainting can be formulated as an optimization problem:

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} Q(\mathbf{Y}, \mathbf{X}) + \lambda R(\mathbf{X}) \tag{5}$$

where the function Q often controls the fidelity towards the observed data and is chosen to fit the reconstruction task. R is a regularizer (or penalty) function selected based on prior knowledge, and λ is the tuning parameter to trade-off between the two terms. One major drawback of this framework is that the selection of a good regularizer depends on the application and the available prior knowledge, which can considerably vary in the case of natural images. A widely used regularizer is total variation, which promotes piece-wise smoothness on $\mathbf{X}$.

In [38], it was shown that the regularizer can be implicitly substituted by a deep network:

$$\hat{\theta} = \arg\min_{\theta} Q(\mathbf{Y}, f_\theta(\mathbf{Z})) \quad \text{s.t.} \quad \hat{\mathbf{X}} = f_{\hat{\theta}}(\mathbf{Z}). \tag{6}$$

where the selection of a proper regularizer is taken off the hands of the user and the optimization is shifted towards optimizing the network parameters, i.e., weights and biases. The minimization problem (6) is solved using the network's optimizer, e.g., a gradient descent, applied to the network's parameters θ. A common choice for the function Q is the least squares term and, hence, the problem to solve becomes:

$$\hat{\theta} = \arg\min_{\theta} \frac{1}{2}||\mathbf{Y} - f_\theta(\mathbf{Z})||_F^2 \quad \text{s.t.} \quad \hat{\mathbf{X}} = f_{\hat{\theta}}(\mathbf{Z}). \tag{7}$$

E. Abundance Estimation using DIP

In this subsection, we adapt DIP to solve the unmixing problem. Unlike the majority of the deep learning-based unmixing techniques proposed in the literature, we propose to use a deep network for estimating the abundances $\mathbf{A}$ only, given fixed endmembers $\mathbf{E}$. The widely used classical method to estimate the abundances is to solve the optimization problem:

$$\hat{\mathbf{A}} = \arg\min_{\mathbf{A}} \frac{1}{2}||\mathbf{Y} - \mathbf{E}\mathbf{A}||_F^2 \quad \text{s.t.} \quad \mathbf{A} \geq 0, \mathbf{1}_r^T\mathbf{A} = \mathbf{1}_n^T, \tag{8}$$

i.e., the fully constrained least squares unmixing (FCLSU) due to the use of both the ASC and ANC. It has been shown that the regularized (or penalized) least squares techniques can take into account prior knowledge of the data and, therefore, provides a better estimation of the abundances [3]:

$$\hat{\mathbf{A}} = \arg\min_{\mathbf{A}} \frac{1}{2}||\mathbf{Y} - \mathbf{E}\mathbf{A}||_F^2 + \lambda R(\mathbf{A}) \text{s.t.} \mathbf{A} \geq 0, \mathbf{1}_r^T\mathbf{A} = \mathbf{1}_n^T, \tag{9}$$

where $R(\mathbf{A})$ is the regularizer or penalty term and λ is the regularization parameter. The choice of R is dependent on the available prior knowledge which can vary considerably in remote sensing images. However, the regularizer can be implicitly substituted by a deep network and the problem is transformed into an optimization of the network's parameters:

$$\hat{\theta} = \arg\min_{\theta} \frac{1}{2}||\mathbf{Y} - \mathbf{E}f_\theta(\mathbf{Z})||_F^2 \quad \text{s.t.} \quad \hat{\mathbf{A}} = f_{\hat{\theta}}(\mathbf{Z}). \tag{10}$$

Therefore, problem (10) can be solved using a deep network. The only issue left to solve is to enforce the constraints. The constraints in (9) can be easily enforced by using a softmax function in the final layer of the network, given by:

$$softmax(\mathbf{A}) = \frac{e^{\mathbf{A}_{ij}}}{\sum_{i=1}^{r} e^{\mathbf{A}_{ij}}} \quad \forall i, j \tag{11}$$

As a result, the unmixing problem (8) can be solved using DIP. Fig. 1 depicts the concept of UnDIP. The random input image $\mathbf{Z}$ is fixed. f_θ is a deep network with parameters θ, which are initialized using random weights θ_0 and updated through the learning process. The core idea of UnDIP is to map $\mathbf{Z}$ to $\hat{\mathbf{A}}$, using a deep network f_θ such that $\hat{\mathbf{A}} = f_{\hat{\theta}}(\mathbf{Z})$. Therefore, $\hat{\theta}$ should be estimated. As can be seen from Fig. 1, UnDIP optimizes the network's parameters θ iteratively by computing the gradient of the loss function (10), which relies on the endmembers ($\mathbf{E}$) extracted by SiVM.

When a network is overtrained, overfitting occurs, and the network will not reach the optimal solution for a test set. Since the design of UnDIP is not based on training and testing sets, UnDIP is robust to overfitting of the network. The optimization is done by iterating based on a fixed input and by optimizing the output until the loss function has converged. On the other hand, since UnDIP is an iterative algorithm, the stopping point becomes an important hyperparameter, which will be discussed in Subsection II-G.

F. Convolutional Neural Network for UnDIP

DIP requires the selection of a network. The description of DIP in subsection II-D did not specify a specific network selection. In [38], the convolutional encoder-decoder network was suggested as the best option for DIP. Here, we discuss in detail the network (i.e., f_θ) shown in Fig. 2 used for UnDIP. The CNN, f_θ, in UnDIP has a few major differences with the other deep (convolutional) networks, typically used for unmixing. First, the entire network is only used for the abundance estimation, as the endmembers are extracted using a geometrical approach and are fixed throughout the unmixing. This framework allows using an unsupervised CNN for unmixing where the convolutions can be applied globally on the entire spatial domain to extract the spatial information. Second, the autoencoder network, generally used for deep spectral unmixing reconstructs spectra as the output of the network

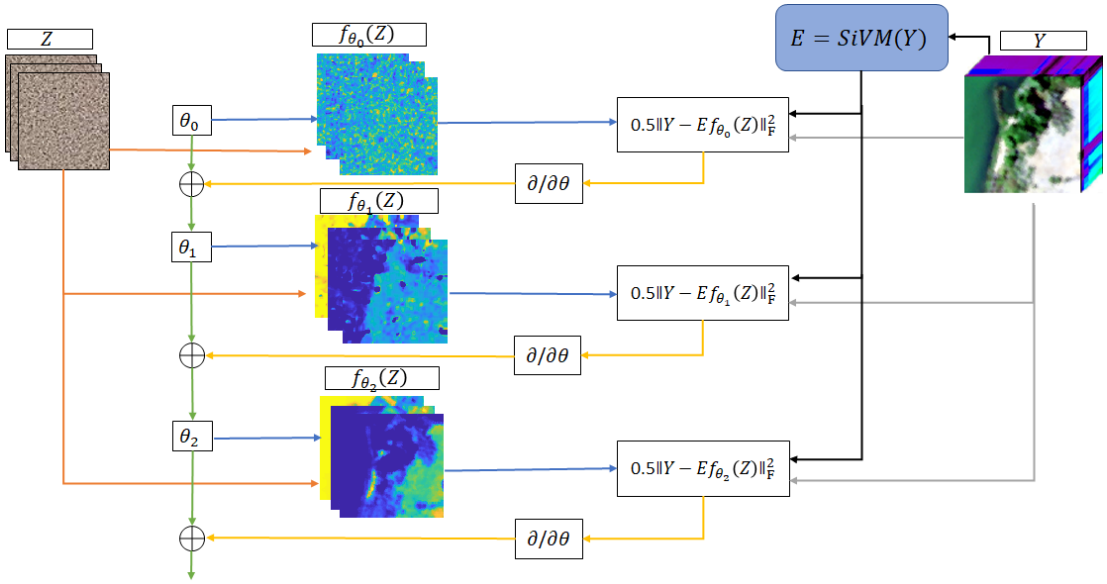


Fig. 1. Graphical illustration of UnDIP. UnDIP maps a random noise input image $\mathbf{Z}$ to $\hat{\mathbf{A}}$ using a deep network f_θ such that $\hat{\mathbf{A}} = f_{\hat{\theta}}(\mathbf{Z})$. To estimate the network's parameters $\hat{\theta}$, UnDIP starts with randomized weights (θ_0) and optimizes θ iteratively by computing the gradient of the loss function (10) which utilizes the endmembers ($\mathbf{E}$), extracted by SiVM.

using the observed spectra as the input of the network. To do so, different loss functions such as spectra angle distance and mean squared error were used to minimize the reconstruction error. As we will show later in the experiments, minimizing the reconstruction error w.r.t. both endmembers and abundances does not necessarily provide a good abundance estimation, which is the main goal in unmixing. On the other hand, in UnDIP, the input is Gaussian noise and the output is given by the abundance maps. The network is trained to minimize the loss function, w.r.t. the abundances solely.

The core of the UnDIP network is based on the convolutional encoder-decoder (also called hourglass) with some skip connections, as proposed in [38], however, with two major differences. First, UnDIP uses only one downsampling block, one upsampling block, and one skip block while DIP uses five blocks for each. From our experiments, we found that the use of several downsampling blocks downgrades the spatial resolution for the unmixing application. Additionally, as can be observed in Fig. 3, the UnDIP network converges much faster and leads to better abundance estimations than DIP. The other main difference is the activation function used in the final layer of UnDIP. While the leaky activation function is used in all layers of DIP, UnDIP uses the Leaky ReLU activation function for all the layers except the final layer. For the final layer, UnDIP exploits a softmax activation function to hold the constraints as discussed before.

The main part of the forward pass (the plain network without the skip connection) starts with two blocks of 3 layers: a convolution layer (Conv), a batch normalization (BN) layer, and a Leaky ReLU nonlinear activation layer, which are followed by a bilinear upsampling layer to account for the stride factor used in the convolutions. This type of three-layer blocks (i.e., conv, BN, and activation) is the most common one used in the CNN architectures in the literature. The convolutional layers extract different spatial features by using different filters. The BN speeds up the learning process and also provides more robustness in terms of the hyperparameter selection. The activation layer promotes the nonlinearity on the prediction in every layer. Deep networks are hard to train due to vanishing gradients. The skip connection is a solution to this problem and enables to train a deep network by using an activation from one layer and add it to a deeper layer. In this way, the network can easily learn the identity function when the parameters become zero. The network exploits two more blocks of convolution, batch normalization, and Leaky ReLU, followed by a convolution layer and softmax which finally generates the abundances.

G. Network Component and Hyperparameter Selection

In this work, Leaky ReLU was used as the activation function (except in the last layer), which often speeds up the learning process since the derivative is either one or close to zero. We compared the performance of Leaky ReLU with the use of Sigmoid, ELU, ReLU activation functions, and found that both Leaky ReLU and ReLU provide the best results. Leaky ReLU was selected since it is the default for the DIP network. The negative slope of Leaky ReLU was set to 0.1, which is also the default value in the DIP network. For the filter size of the convolutional layers, we used the default values proposed in [38], i.e., 3×3 in the forward connections and 1×1 in the skip connections. Downsampling is often applied using pooling and/or stride inside the CNN. For downsampling, we only used the stride within the convolution module as is the default in [38]. For upsampling, we experimented with both the nearest neighborhood and bilinear interpolation, and found that bilinear interpolation performs the best. Reflection padding was used in the convolution to preserve the size of the image. The number of filters used is

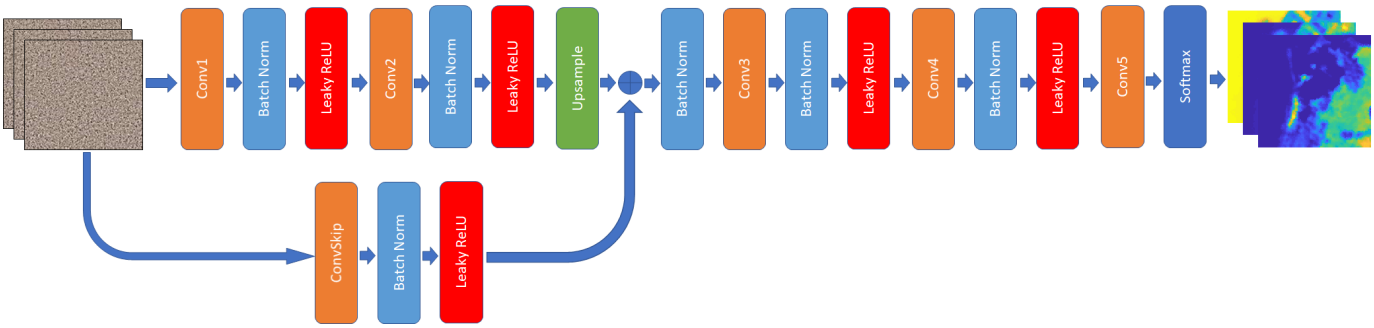


Fig. 2. The proposed convolutional network architecture with one skip connection. This network is used as f_θ for UnDIP in the experiments. Different layers in the network are shown with specific colors.

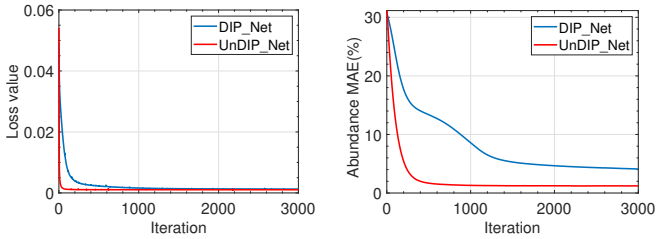


Fig. 3. Comparison of the network architecture of DIP versus UnDIP, applied on the Jasper Ridge data. (a) Loss function value and (b) Abundance MAE.

4 in the skip connection and 256 in the forward connections. The hyperparameters of the network are listed in Table I. We

TABLE I
HYPERPARAMETERS USED IN THE EXPERIMENTS FOR UNDIP.

Hyperparameters				
	Input Ch.	Ouput Ch.	Filter Size	Stride
Conv1	r	256	3x3	2
Conv2	256	256	3x3	1
Conv3	260	256	3x3	1
Conv4	256	256	1x1	1
Conv5	256	r	1x1	1
ConvSkip	r	4	1x1	1
	Negative Slope			
Leaky ReLU	0.1			
	Scale Factor		Mode	
Upsample	2		Bilinear	
	Type	Learning Rate		Iterations
Optimizer	Adam	0.001		3000

should emphasize that we do not optimize the hyperparameters according to the dataset and/or the SNR since this would be unfair to the competing methods used in the experiments. Therefore, the values mentioned in Table I are not optimal and careful tuning according to the noise level and dataset could possibly lead to better results and probably faster convergence. Since UnDIP is an iterative algorithm (as opposed to the other CNN-based algorithms which use training sets for learning) the stopping point or the number of iterations becomes an important hyperparameter to set. To deal with this issue, we use (as also suggested in [38]) exponentially weighted averaging over the outputs and set the number of iterations to a large number (3000). This makes the algorithm very robust to this parameter since the overall average is very close to the minimum solution, even if there is a considerable jump in the loss function at the stopping iteration. Finally, an Adam optimizer was used with a learning rate of 0.001 and PyTorch was used as the platform for the network implementation.

III. Experimental Results

The experiments were performed on a simulated dataset and three real datasets. The description of the datasets is given below.

A. Hyperspectral Data Description

1) Simulated Dataset: A dataset of 60×75 pixels is simulated by generating linear mixtures of three minerals, i.e., Fe_2O_3, SiO_2, and CaO. The endmembers, which are shown in Fig. 4(s), were measured by an AgriSpec spectrometer (manufactured by ASD [Analytical Spectral Devices]) and contain 200 reflection values in the wavelength range [1000-2500] nm. The ground truth abundance maps are shown in Fig. 4(b). These contain 20 squares of 5 × 5 pixels with different binary and ternary linear mixtures. The background contains binary mixtures of 50% of Fe_2O_3 and 50% of SiO_2.

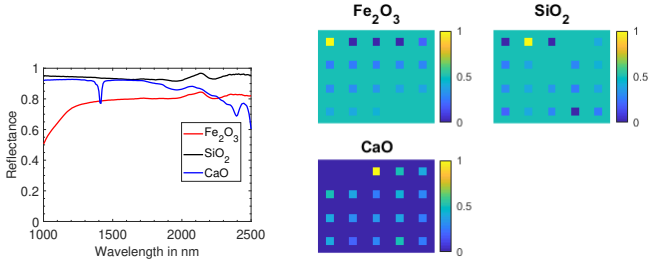
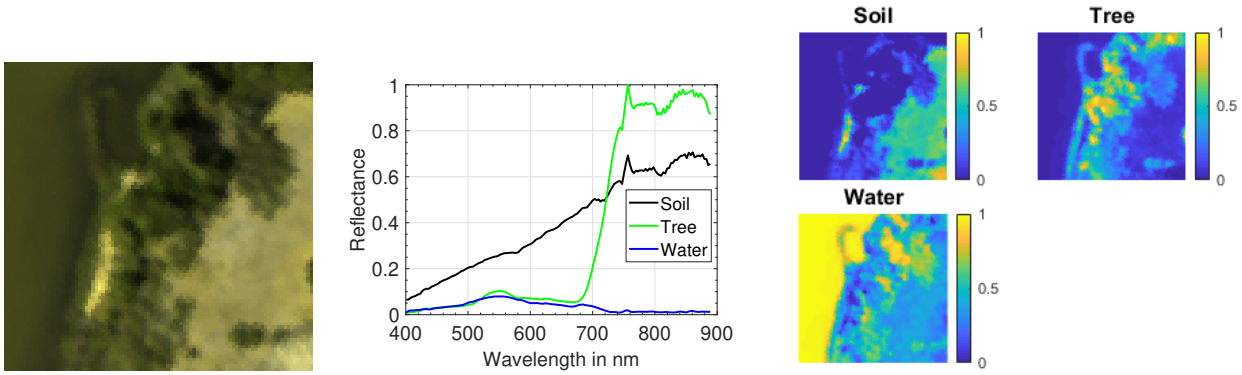


Fig. 4. The simulated image: (a) Endmembers; (b) Abundance maps.

2) Samson image: The Samson hyperspectral dataset is shown in Fig. 5(a) and contains 95×95 pixels. The spectral signatures contain 156 bands in the wavelength range [401-889] nm. There are three main materials (i.e., Soil, Tree, and Water). The ground truth endmembers were extracted using SiVM and the ground truth fractional abundances were generated using FCLSU. Both are shown in Fig. 5(b) and Fig. 5(c)), respectively.



(a) (b) (c)

Fig. 5. Samson image: (a) True-color image (Red: 571.01 nm, Green: 539.53 nm, and Blue: 432.48 nm); Endmembers; (c) Abundance maps.

3) Jasper Ridge image: The Jasper Ridge dataset contains 100×100 pixels and is shown in Fig. 6(a). The dataset contains 224 bands covering the wavelength range [380-2500] nm. The water absorption bands (1–3, 108–112, 154–166, and 220–224) were removed and 198 channels were retained. There are four endmembers (i.e., Tree, Water, Soil, and Road, shown in Fig. 6(b)), which are extracted using SiVM. The ground truth fractional abundances (Fig. 6(c)) were estimated using FCLSU.

4) Apex Dataset: The cropped image used in the paper contains 111×122 pixels (shown in Fig. 7(a)) and 285 bands which cover the wavelength range [413–2420] nm. In this dataset, there are four ground measured endmembers (i.e., Water, Grass, Road, and Roof, shown in Fig. 7(b)). The scene is influenced by variable illumination conditions and contains a shadow covered area. Therefore, to create the ground truth fractional abundances, we added a shadow endmember (a zero spectrum) to the list of ground truth endmembers, and then applied FCLSU.

5) Washington DC Mall Dataset: Washington DC Mall is an airborne hyperspectral image, captured over the Washington DC Mall using the HYDICE sensor. The cropped image (Fig. 8(a)) used in this paper contains 319 × 292 pixels in 191 bands over the spectral range from 0.4 to 2.4 μm. The ground truth is available online [1] and contains 7 classes: Grass, Tree, Roof, Road, Water, Trail, and Shadow. The ground truth endmembers are selected manually for this dataset (shown in Fig. 8(b)) and FCLSU was used to estimate the ground truth fractional abundances.

B. Experimental Setup

Seven unmixing methods from different categories were used as competing methods in the experiments:

- The baseline FCLSU [9],
- A blind unmixing method: NMF-QMV [15],
- A sparse unmixing method Collab, which is based on a group sparsity inducing mixed norm using the collaborative LASSO [25],
- Three deep unmixing methods: uDAS [31], SNSA [28], and DAEN [29].

[1]https://engineering.purdue.edu/ landgreb/Hyperspectral.Ex.html

All the parameters for the competing methods were selected according to the reported default values.

Hyperspectral images generally contain different levels and types of noise [44]. It has been shown that hyperspectral unmixing techniques are often remarkably robust to noise and can be used as denoisers [45]. To compare the robustness of the techniques w.r.t. the image SNR, we added white zero-mean Gaussian noise to the data to generate the observed data $\mathbf{Y}$. Images are generated with SNR= 20, 30, 40, and 50 dB, on all datasets, except for the Apex and the Washington DC Mall images. All experiments are repeated five times with random noise realizations. Mean results and standard deviations are shown.

For all the datasets, ground truth abundance maps are available and therefore, quality assessment metrics are applied to compare the results. In the experiments, the results are compared based on the abundance mean absolute error (MAE), the reconstruction error (RE), the spectral RMSE, and the spectral angle distance (SAD). All results, except for SAD are reported as percentages. The abundance MAE is given by the mean of the absolute errors (in percent) between the estimated abundances and the ground truth abundances:

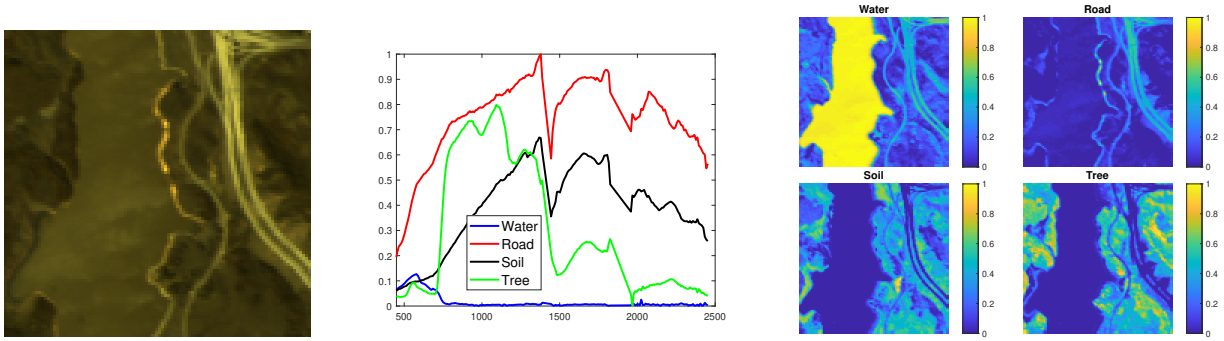
$$\text{Abundance MAE} = \frac{1}{rn}\sum_{k=1}^{r}\sum_{i=1}^{n}\left|\hat{\mathbf{A}}_{ki} - \mathbf{A}_{ki}\right| \times 100, \quad (12)$$

the reconstruction error is the RMSE (in percent) between the obtained reconstructed image $\hat{\mathbf{X}}$ and the observed (noisy) image $\mathbf{Y}$:

$$\text{RE} = \sqrt{\frac{1}{pn}\sum_{j=1}^{p}\sum_{i=1}^{n}\left(\hat{\mathbf{X}}_{ji} - \mathbf{Y}_{ji}\right)^2} \times 100, \quad (13)$$

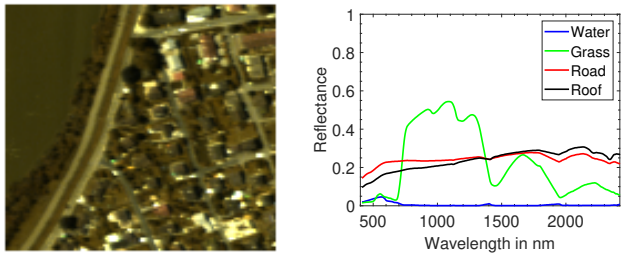
The Spectral RMSE is the RMSE (in percent) between the obtained reconstructed image $\hat{\mathbf{X}}$ and the original noise-free image $\mathbf{X}$:

$$\text{Spectral RMSE} = \sqrt{\frac{1}{pn}\sum_{j=1}^{p}\sum_{i=1}^{n}\left(\hat{\mathbf{X}}_{ji} - \mathbf{X}_{ji}\right)^2} \times 100, \quad (14)$$



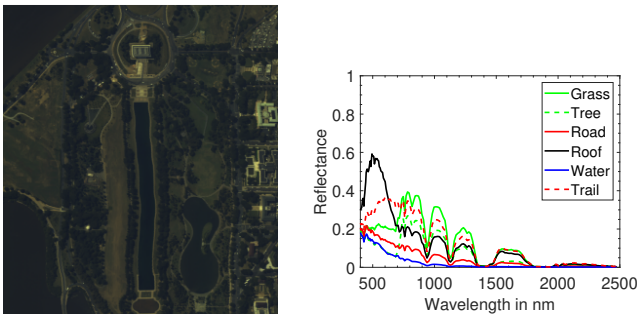
(a) (b) (c)

Fig. 6. Jasper Ridge image. (a):True-color image (Red: 570.14 nm, Green: 532.11 nm, Blue: 427.53 nm); (b): Endmembers; (c): Abundance maps.



(a) (b)

Fig. 7. Apex image: (a) True-color image (Red: 572.2 nm, Green: 532.3 nm, Blue: 426.5 nm); (b) Endmembers.



(a) (b)

Fig. 8. Washington DC Mall image: (a) True-color image (Red: 572.7 nm, Green: 530.1 nm, Blue: 425.0 nm); (b) Endmembers.

SAD (in degree) is used to measure the spectral angle distance between an estimated and the ground truth endmember as:

$$\mathrm{SAD}(\mathbf{e}_{(i)}, \hat{\mathbf{e}}_{(i)}) = \arccos\left(\frac{\langle \mathbf{e}_{(i)}, \hat{\mathbf{e}}_{(i)} \rangle}{\|\mathbf{e}_{(i)}\| \, \|\hat{\mathbf{e}}_{(i)}\|}\right)\frac{180}{\pi},$$

We should note that, although a lower Abundance MAE denotes a better abundance estimation and a lower Spectral RMSE denotes a better signal reconstruction, a lower RE does not necessarily mean a better abundance estimation performance or a better signal reconstruction. According to the linear model, the RE depends on the linear combination of the endmembers and abundances. The multiplication of both may be close to the observed spectra, but individually, they might not represent the true endmembers and abundances. Additionally, the RE includes model errors (nonlinearities) and noise. Only if the data contains insignificant levels of model errors and noise, a lower RE denotes an improved performance, and then, the RE will be close to the spectral RMSE since the observed data is close to the original data. The RE should be interpreted along with the Abundance MAE. If the abundance estimation is satisfactory, then a lower RE indicates a better performance. Otherwise, the spectral RMSE is more informative for validating the performance.

C. Unmixing Experiments

1) Experiments on Simulated Dataset: Fig. 9 shows the results of the unmixing techniques applied on the simulated data. As can be observed from Fig. 9 (a), UnDIP and FCLSU obtain the lowest Abundance MAE for all SNR values. DAEN slightly outperforms the remaining techniques and Collaborative LASSO provides the poorest results for 20 dB. The RE for all techniques is low, despite the poor abundance estimation of some of the methods, e.g. sparse unmixing. Therefore, the Spectral RMSE is more informative (Fig. 9 (c)). In the case of simulated data, the error is only induced by the noise, since no other model errors were simulated. SNSA, UnDIP, FCLSU, and NMF-QMV obtain the lowest RMSE, confirming, along with the good abundance estimation performance, that these methods are able to reconstruct the data. Fig. 9 (d) shows the performance of the endmember estimation by the different techniques, in terms of SAD. Both UnDIP and FCLSU apply SiVM for the extraction of the endmembers. It can be observed that SiVM outperforms the other techniques in terms of SAD for all SNRs.

Fig. 10 visually compares the obtained abundance maps using the different unmixing techniques for SNR=20. The visual comparison reveals that UnDIP is less sensitive to noise than the other techniques, and generates abundance maps that are very close to the ground truth abundances, even for SNR values as low as SNR=20 dB. In supervised CNN, image patches are extracted to train the network and, therefore, the convolutional operator is only applied on a spatial subset of the data. Depending on the size of the patches, the spatial information can be considerably degraded. On the other hand, UnDIP applies the convolutional operator on the entire spatial domain, since it is an unsupervised CNN. As can be seen from Fig. 10, the proposed method successfully preserves the structures and provides better abundance estimations.

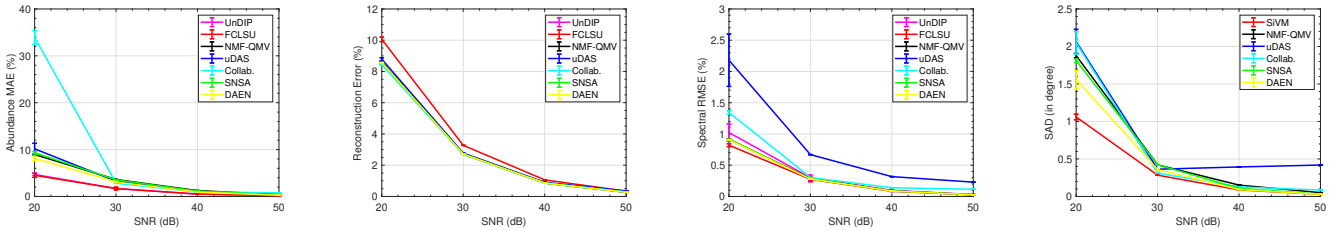


Fig. 9. Simulated data - The results of unmixing in terms of (a) Abundance MAE, (b) Reconstruction Error, (c) Spectral RMSE, and (d) SAD (in degree) w.r.t. different noise levels of the observed image (in SNR).

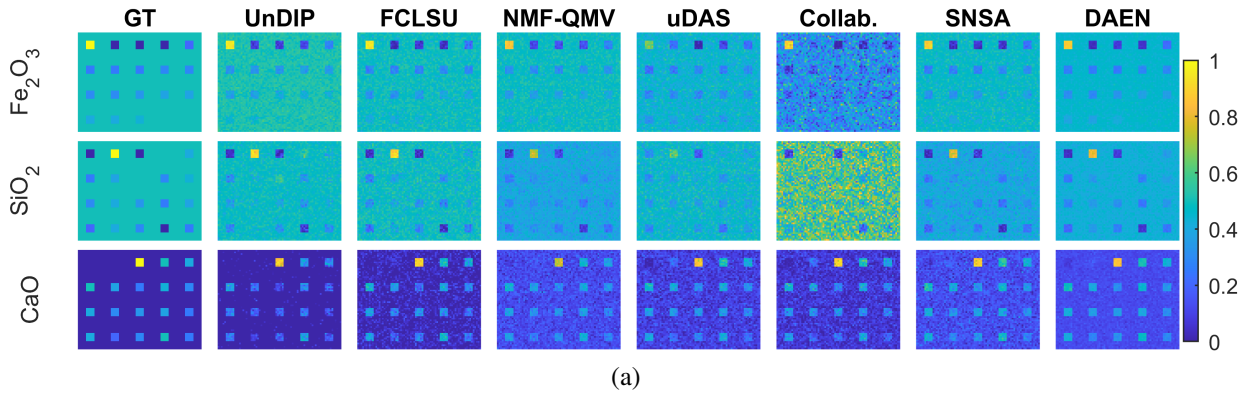


Fig. 10. Simulated data - Abundance maps obtained by applying different unmixing techniques (20 dB).

2) Experiments on Samson Dataset: Fig. 11 shows the results of the unmixing experiments applied on the Samson dataset, and Fig. 12 shows the estimated abundance maps. It can be observed that FCLSU, UnDIP, and NMF-QMV obtain the best abundance estimation performances (Fig. 11 (a)), and produce similar abundance maps, close to the ground truth (Fig. 12). However, NMF-QMV is more sensitive to noise. Both UnDIP and NMF-QMV obtain a lower RE and Spectral RMSE than FCLSU. The Abundance MAE of uDAS increases with increasing noise power, although the RE and Spectral RMSE remain low. One can conclude that uDAS performs better as a denoiser than as an unmixer. This is due to the denoising constraint applied on the encoder in the uDAS network. DAEN performs better in terms of abundance estimation than uDAS for low SNR but worse for high SNR. SNSA obtains a moderate abundance estimation, and the poorest of all methods for 20 dB, which shows that it is not robust w.r.t. noise. The abundance estimation performance of Collaborative unmixing is poor for all SNRs, which makes it very sensitive to noise (notice the large variance for 20 dB), as can also be observed from the abundance maps in Fig. 12. Fig. 11 (d) shows that SiVM and uDAS perform better for the estimation of endmembers than the other methods and show robustness to the noise. In terms of SAD, DAEN, SNSA and NMF-QMV show sensitivity to the noise power . A very low SAD is obtained by collaborative unmixing for 20 dB, but the abundance MAE and the visual comparison in Fig. 12 reveal a poor abundance estimation. The good performance of collaborative unmixing in terms of SAD can be attributed to the averaging effect of endmember bundles that considerably helps to decrease the SAD.

3) Experiments on Jasper Ridge Dataset: All the unmixing techniques were applied to the Jasper Ridge image. The results are given in Fig. 13 and the abundance maps are shown in Fig. 14. For this dataset, FCLSU and UnDIP perform the best in terms of Abundance MAE. FCLSU, however, obtains poor RE and Spectral RMSE. DAEN, SNSA, and NMF-QMV obtain lower RE and Spectral RMSE but are less performant in terms of abundance estimation. Collaborative unmixing obtains the poorest abundance estimation. SNSA is not robust to the noise, despite very low RE and Spectral RMSE. As can be observed from Fig. 14, uDAS mixes the Water and Road classes. Collaborative unmixing can hardly distinguish Soil from Road. The Water and Tree abundance maps are well estimated by all techniques, which can be attributed to their unique endmembers. From Fig. 13(d) one can observe that SiVM outperforms the other techniques w.r.t. endmember extraction. Both NMF-QMV and Collaborative unmixing give poor results. uDAS and SNSA have a similar moderate performance.

4) Experiments on Apex Dataset: To further evaluate the unmixing techniques, they were applied to the Apex dataset, for which ground truth endmembers are available. In this experiment, we did not add artificial noise to the dataset.

The results of abundance estimations are given in Table II and abundances are compared visually in Fig. 15. The lowest overall MAE is obtained by UnDIP which also obtained the best estimations of the abundances for Road and Shadow. Collaborative unmixing also performs well (0.2% higher error

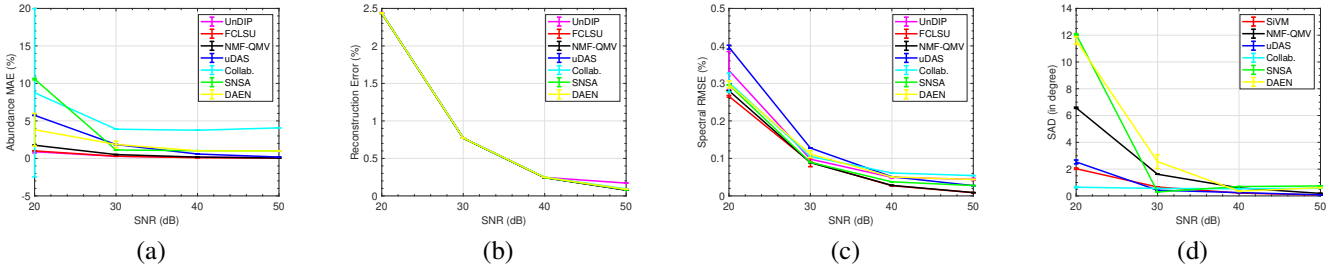


Fig. 11. Samson dataset - The results of unmixing in terms of (a) Abundance MAE, (b) Reconstruction Error, (c) Spectral RMSE, and (d) SAD (in degree) w.r.t. different noise level of the observed image (in SNR).

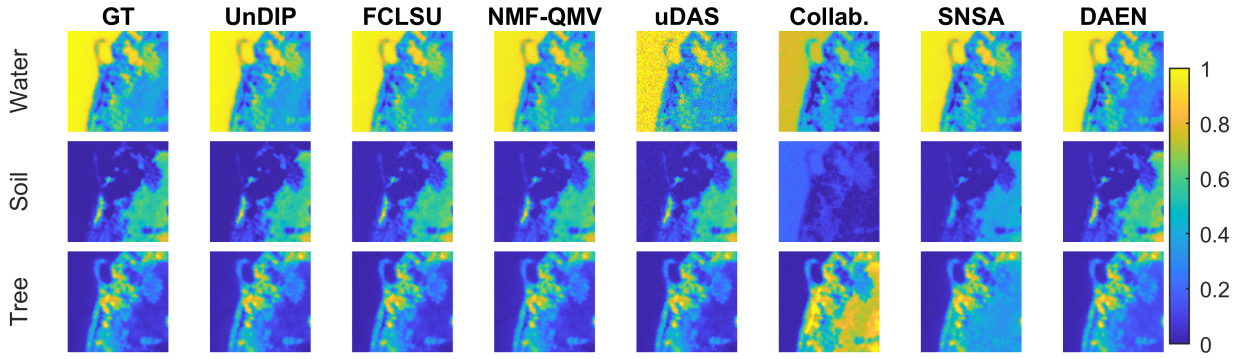


Fig. 12. Samson dataset - Abundance maps obtained by applying different unmixing techniques (20 dB).

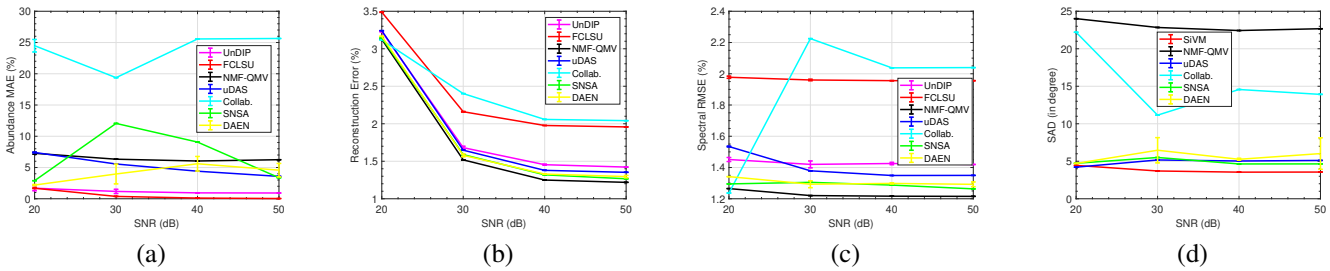


Fig. 13. Jasper Ridge dataset - The results of unmixing in terms of (a) Abundance MAE, (b) Reconstruction Error, (c) Spectral RMSE, and (d) SAD (in degree) w.r.t. the different noise levels of the observed image (in SNR).

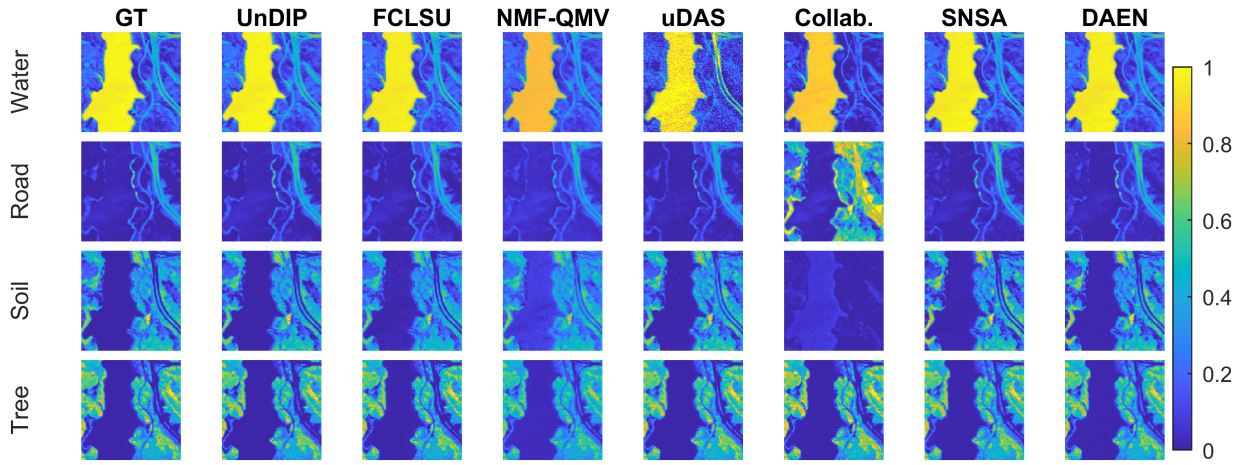


Fig. 14. Jasper Ridge dataset - Abundance maps obtained by applying different unmixing techniques (20 dB).

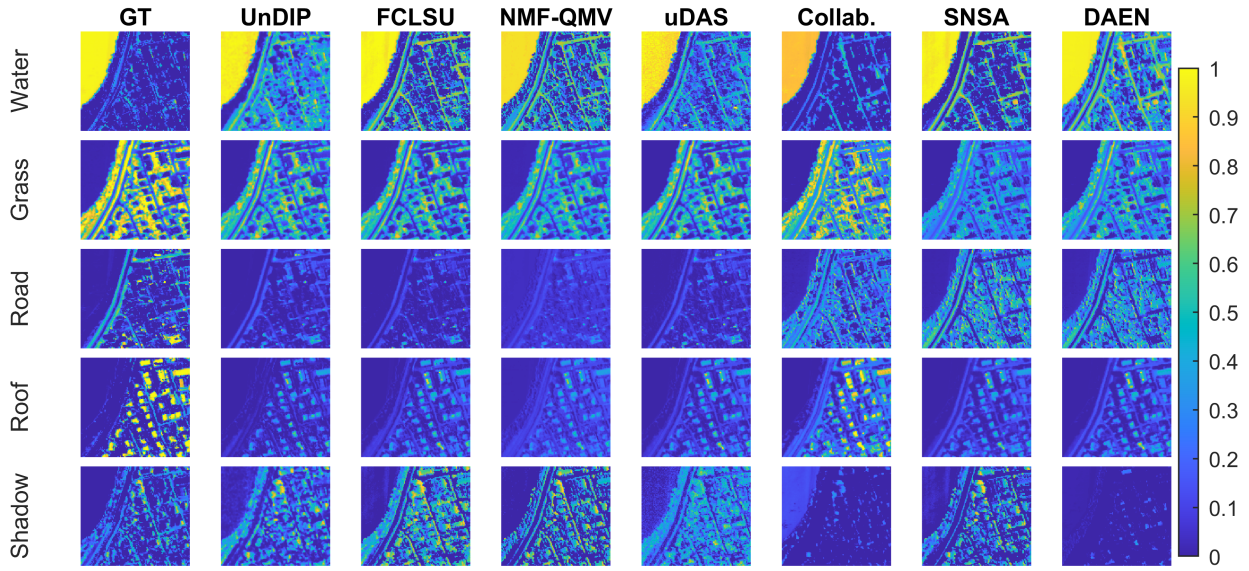


Fig. 15. Apex dataset - Abundance maps obtained by applying different unmixing techniques.

than UnDIP) and obtains the best estimations for Water and Grass. uDAS and FCLSU perform similarly with 0.9 and 0.8% higher error than UnDIP, respectively. NMF-QMV, DAEN and SNSA obtain abundance errors which are considered poor compared to the other competing techniques.

The visual comparison in Fig. 15 confirms the results reported in the table. Although Collaborative unmixing provides the lowest MAE for Water, a visual comparison reveals that it is the only technique which considerably mixes Water with Shadow, while for Grass, it shows the best performance, also visually. UnDIP show the best performance for Road, while all the other techniques mix the abundances of Road and Roof. UnDIP outperforms the others on Shadow, also visually. The performances on the endmember estimation are compared in Table IV. It can be observed that SiVM outperforms the other techniques in terms of SAD. NMF-QMV gives the highest SAD.

TABLE II
ABUNDANCE MEAN ABSOLUTE ERROR (IN %) OF THE APEX DATASET. THE BEST PERFORMANCES ARE SHOWN IN BOLD.

Abundance	UnDIP	FCLSU	NMF-QMV	uDAS	Collab.	SNSA	DAEN
Water	21.4	18.4	22.4	18.8	**10.7**	15.0	22.4
Grass	11.7	12.3	13.9	12.7	**11.1**	21.2	17.5
Road	**8.3**	8.6	10.5	8.9	19.7	28.0	26.6
Roof	11.6	**11.5**	14.1	11.8	11.7	13.3	13.0
Shadow	**11.5**	18.1	18.1	16.4	12.2	13.7	11.7
Overall	**12.9**	13.8	15.8	13.7	13.1	18.2	18.2

5) Experiments on Washington DC Mall Dataset: The unmixing techniques were applied on the Washington DC Mall dataset and the results are compared in Table III. Collaborative unmixing provides the best MAE. SNSA, UnDIP, and FCLSU perform similarly in terms of MAE and can be considered as the second best results in the table. uDAS provides the worst results on this dataset. The visual comparison in Fig. 16 reveals that all the methods fail to adequately estimate the abundances. This is due to the poor endmember estimation or extraction, as can be observed in Table IV.

TABLE III
ABUNDANCE MEAN ABSOLUTE ERROR (IN %) OF THE WASHINGTON DC DATASET. THE BEST PERFORMANCES ARE SHOWN IN BOLD.

Abundance	UnDIP	FCLSU	NMF-QMV	uDAS	Collab.	SNSA	DAEN
Grass	20.4	20.4	25.1	**19.4**	20.3	20.5	19.8
Tree	27.6	27.7	**10.5**	25.2	27.2	28.4	27.2
Rood	9.9	**9.8**	16.0	14.9	11.1	12.6	11.6
Roof	3.0	3.7	7.0	1.9	**1.7**	3.9	5.8
Water	29.7	30.0	**14.6**	22.3	21.3	24.3	30.9
Trail	5.5	5.8	23.8	7.4	**5.3**	6.7	12.3
Shadow	4.0	4.3	1.5	38.4	2.1	2.3	**1.1**
Overall	14.3	14.5	16.7	18.5	**12.7**	14.1	15.5

TABLE IV
SAD OF THE APEX AND WASHINGTON DC MALL DATASETS. THE BEST PERFORMANCES ARE SHOWN IN BOLD.

SAD	SiVM	NMF-QMV	uDAS	Collab.	SNSA	DAEN
Apex	**10.80**	41.68	18.28	27.14	20.93	19.47
WDC	**12.26**	26.04	21.14	12.92	26.49	14.70

D. Discussion

Here, we summarize and discuss the results obtained from the experiments.

- In all experiments, a very low Abundance MAE, RE, and Spectral RMSE was obtained by UnDIP compared to all competing methods. This can be partially attributed to its ability to globally incorporating the spatial information, as can visually be observed from e.g. the abundance maps of the simulated data. The results also clearly indicate that UnDIP is very robust to noise, which is due to the implicit application of a regularizer in the network. The incorporation of a geometrical endmember

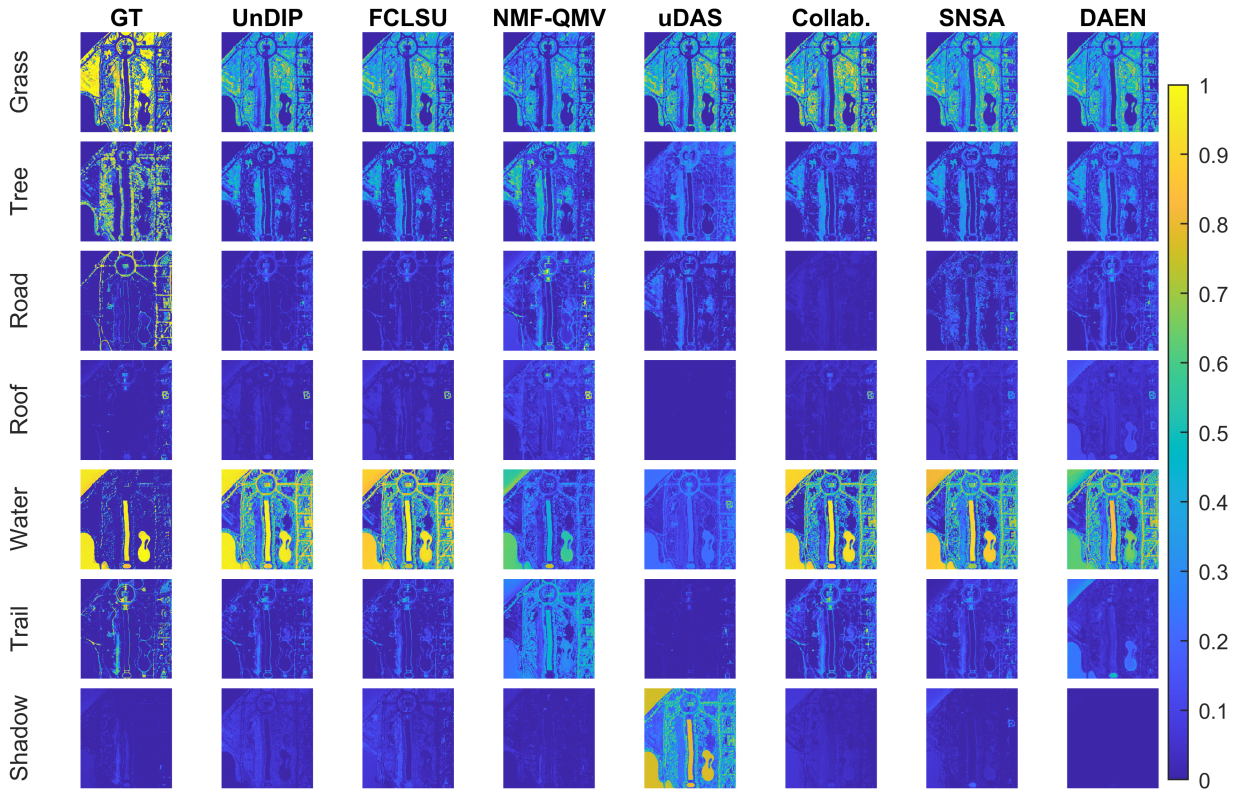


Fig. 16. Washington DC Mall dataset - Abundance maps obtained by applying different unmixing techniques.

estimation approach assures that it is entirely devoted to the abundance estimation. Other methods that jointly estimate the endmembers and the abundances obtain low RE values, but do not necessarily perform well on the abundance estimation. Since the abundance estimation highly depends on the quality of the endmembers, a poor endmember estimation evidently leads to a poor abundance estimation.

- FCLSU performs equally well for estimating fractional abundances, but obtains higher RE and Spectral RMSE, making it more sensitive to noise compared to UnDIP. We should note that FCLSU is used to generate the ground truth abundances from the noiseless images. and therefore the Abundance MAE of FCLSU can be considered as the benchmark.
- uDAS, and NMF-QMV obtain moderate results. On the simulated dataset, they perform equally well. On the Samson dataset, NMF-QMV performs better, while uDAS performs better on Jasper Ridge and Apex. NMF-QMV is more robust to noise and obtains lower Spectral RMSE. This can be attributed to the regularization term for which the regularization parameter was optimally selected. uDAS provides low RE and moderate Spectral RMSE which can be attributed to the denoising constraint inside the deep network. Although uDAS is designed to optimize the RE, the experimental results show that this does not guarantee an optimal abundance estimation.
- SNSA obtains good Spectral RMSE but is not as robust as the competing methods for abundance estimation. SNSA is based on stacked encoder-decoders and does not exploit the spatial information. Moreover, the tuning parameter of the minimum volume regularizer in the cost function is fixed and not automatically selected and cannot perform well for all the noise levels.
  Overall, DAEN performs moderately. DAEN utilizes a variational auto encoder-decoder to improve the abundance and endmember estimation by employing a regularizer into the loss function. Additionally, DAEN exploits stacked encoder-decoders to reduce the sensitivity to the noise, which can be clearly observed in the experimental results.
- Collaborative unmixing obtains the worst results and is shown to be very sensitive to noise throughout the experiments. This may be attributed to the fact that the endmember bundles are not available a priori but rather are generated from the data.
- The reported results in terms of SAD reveal the significant role of the estimated/extracted endmembers on the abundance estimation. The results confirm that poor endmember estimation leads to poor abundance estimation. SiVM consistently outperforms the other techniques in all the experiments performed in this paper and shows robustness with respect to the noise power. However, for both Apex and Washington DC datasets, none of the methods could estimate/extract the endmembers satisfactorily. This can be attributed to the occurrence of highly mixed pixels and nonlinearities in those datasets.
- Notice that all reported standard deviations are very

small, except in some cases at 20dB. It seems that all randomness, from different noise realizations, initializations (all methods except UnDIP use VCA to initialize the endmembers, and the random initialization of the UnDIP network) is well overcome by the applied methods. In particular, almost always the same endmembers were extracted, irrespective of the noise level.

E. Sensitivity Analysis to Hyperparameters

In the concept of the deep image prior, it is important that all the hyperparameters are tuned with respect to the application to obtain a better performance [38]. Here, we evaluate the performance of UnDIP with respect to the hyperparameters of the network. The results for the Jasper Ridge dataset (50 dB) are depicted in Fig. 17. Fig. 17 (a) shows the performance of UnDIP with respect to the spatial size of the convolutional filter. It can be seen that the size of 3×3 is optimal. 5×5 filters perform similarly in terms of MAE but at a higher computational cost. Fig. 17 (b) plots the MAE values in function of the number of convolutional filters. As can be seen, the use of 256 filters provides the best result. Fig. 17 (c) plots the loss function in function of the number of iterations for three different learning rates (LRs). It can be seen that a learning rate of LR = 0.001 provides the fastest convergence for the proposed algorithm. Fig. 17 (d) compares the performance of UnDIP in terms of MSE for different activation functions. Both Leaky ReLU and ReLU outperform the Sigmoid and ELU activation functions.

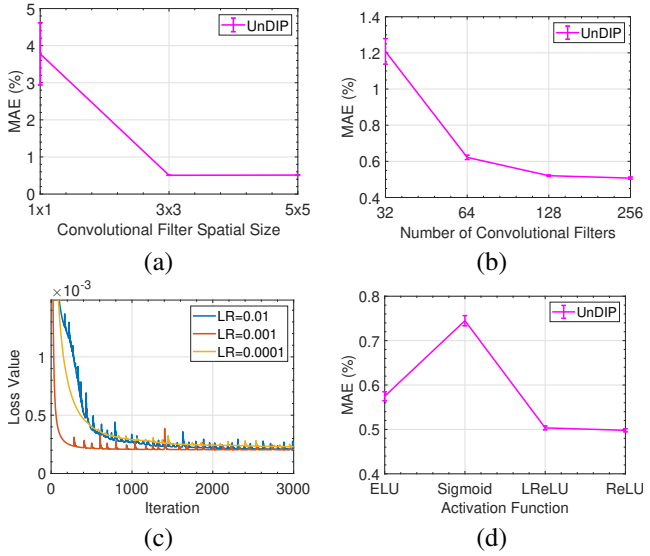


Fig. 17. Sensitivity of UnDIP to the hyperparameters of the network. The experiments was performed on Jasper dataset (50 dB).

F. Processing Time

Table V reports the processing times for the different unmixing techniques applied to the Apex and Washington DC Mall datasets. All the algorithms were implemented in Matlab (2020b), except UnDIP which was implemented in Python (3.8). The reported processing times were obtained using a computer with an Intel(R) Core(TM) i9-10980 HK processor (2.4 GHz), 32GB of memory, a 64-bit Operating System and an NVIDIA GEFORCE RTX (2080 Super) graphical processing unit. The results are averaged over five experiments. From the table, it can be observed that, partially due to the efficiency of GPU programming, the proposed deep learning method is very competitive to geometric, blind, and sparse unmixing, in terms of computational time.

TABLE V
PROCESSING TIME (IN SECONDS) OF THE UNMIXING TECHNIQUES APPLIED TO THE APEX AND WASHINGTON DC MALL DATASETS.

	UnDIP	**FCLSU**	**NMF-QMV**	**uDAS**	**Collab.**	**SNSA**	**DAEN**
Apex	49.32	22.75	**14.18**	235.40	23.87	131.61	678.6
Wash. DC	262.51	**69.39**	543.9	944.57	204.61	1.34e+03	7.17e+03

IV. Conclusion

In this paper, we proposed a deep prior unmixing technique called UnDIP. UnDIP first extracts the endmembers using a geometrical simplex volume maximization technique. Relying on the extracted endmembers, UnDIP estimates the fractional abundances using a deep convolutional network. The network is inspired by the theory behind the deep image prior that implicitly induces a regularizer on the cost function via the network parameters. Experiments were carried out on a simulated dataset and three real datasets. Comparative assessments were performed using sparse, geometrical, deep, and blind unmixing methods. Experimental results confirm that UnDIP outperforms all the other techniques used in the experiments based on quality metrics and visual assessment. Additionally, the experiments showed that UnDIP not only performs very well on abundance estimation but also successfully reconstructs the data. Moreover, UnDIP is considerably robust to the noise power and does not rely on any spectral library. The experimental results also showed that UnDIP is computationally very competitive to the conventional methods used in the experiments due to the efficiency of GPU programming.

Acknowledgment

The work of B. Rasti was funded by the Alexander-von-Humboldt-Stiftung/foundation. The work of B. Koirala was funded by BELSPO (Belgian Science Policy Office) in the frame of the STEREO III programme – project GEOMIX (SR/06/357). We would like to thank Prof. Jun Li and Dr. Yuanchao Su for providing the Matlab code of SNSA and DAEN.

References

[1] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geoscience and Remote Sensing Magazine*, vol. 1, no. 2, pp. 6–36, 2013.

[2] P. Ghamisi, N. Yokoya, J. Li, W. Liao, S. Liu, J. Plaza, B. Rasti, and A. Plaza, "Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 37–78, 2017.

[3] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, April 2012.

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

[5] J. Boardman, F. A. Kruse, and R. Green, "Mapping target signatures via partial unmixing of aviris data: in summaries," in *JPL Airborne Earth Sci. Workshop*, 1995, pp. 23–26.

[6] Michael E. Winter, "N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data," in *Imaging Spectrometry V*, Michael R. Descour and Sylvia S. Shen, Eds. International Society for Optics and Photonics, 1999, vol. 3753, pp. 266 – 275, SPIE.

[7] J. Nascimento and J. Bioucas-Dias, "Vertex component analysis: A˜fast algorithm to extract endmembers spectra from hyperspectral data," in *Pattern Recognition and Image Analysis*, Francisco José Perales, Aurélio J. C. Campilho, Nicolás Pérez de la Blanca, and Alberto Sanfeliu, Eds., Berlin, Heidelberg, 2003, pp. 626–635, Springer Berlin Heidelberg.

[8] Chein-I Chang and D. C. Heinz, "Constrained subpixel target detection for remotely sensed imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 3, pp. 1144–1159, 2000.

[9] D. C. Heinz and Chein-I-Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 3, pp. 529–545, 2001.

[10] J. Sigurdsson, M. O. Ulfarsson, and J. R. Sveinsson, "Blind hyperspectral unmixing using total variation and ℓ_q sparse regularization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 11, pp. 6371–6384, 2016.

[11] N. Dobigeon, S. Moussaoui, M. Coulon, J. Tourneret, and A. O. Hero, "Joint bayesian endmember extraction and linear unmixing for hyperspectral imagery," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4355–4368, 2009.

[12] J. Li and J. M. Bioucas-Dias, "Minimum volume simplex analysis: A fast algorithm to unmix hyperspectral data," in *IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium*, 2008, vol. 3, pp. III – 250–III – 253.

[13] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Collaborative nonnegative matrix factorization for remotely sensed hyperspectral unmixing," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2012, pp. 3078–3081.

[14] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 3, pp. 765–777, 2007.

[15] L. Zhuang, C. Lin, M. A. T. Figueiredo, and J. M. Bioucas-Dias, "Regularization parameter selection in minimum volume hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 12, pp. 9858–9877, 2019.

[16] L. C. Parra, C. Spence, P. Sajda, A. Ziehe, and K.-R. Müller, "Unmixing hyperspectral data," in *Advances in Neural Information Processing Systems 12*. 2000, pp. 942–948, MIT Press.

[17] M. Elad, P. Milanfar, and Ron Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Problems*, vol. 23, no. 3, pp. 947–968, apr 2007.

[18] M. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 6, pp. 2014–2039, 2011.

[19] J. M. Bioucas-Dias and M. A. T. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," 2010.

[20] M. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Collaborative sparse regression for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 1, pp. 341–354, 2014.

[21] M. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 11, pp. 4484–4502, 2012.

[22] L. Meier, S. V. D. Geer, and P. Bühlmann, "The group lasso for logistic regression," *Journal of the Royal Statistical Society. Series B*, vol. 70, no. 1, pp. 53–71, 2008.

[23] M. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Collaborative sparse regression for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 1, pp. 341–354, 2014.

[24] M. Kowalski and B. Torrésani, "Sparsity and persistence: mixed norms provide simple signal models with dependent coefficients," *Signal, Image and Video Processing*, vol. 3, no. 3, pp. 251–264, Sept. 2009.

[25] L. Drumetz, T. R. Meyer, J. Chanussot, A. L. Bertozzi, and C. Jutten, "Hyperspectral image unmixing with endmember bundles and group sparsity inducing mixed norms," *IEEE Transactions on Image Processing*, vol. 28, no. 7, pp. 3435–3450, 2019.

[26] B. Rasti, D. Hong, R. Hang, P. Ghamisi, X. Kang, J. Chanussot, and J. A. Benediktsson, "Feature extraction for hyperspectral imagery: The evolution from shallow to deep (overview and toolbox)," *IEEE Geoscience and Remote Sensing Magazine*, pp. 0–0, 2020.

[27] S. Ozkan, B. Kaya, and G. B. Akar, "Endnet: Sparse autoencoder network for endmember extraction and hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 1, pp. 482–496, 2019.

[28] Y. Su, A. Marinoni, J. Li, J. Plaza, and P. Gamba, "Stacked nonnegative sparse autoencoders for robust hyperspectral unmixing," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 9, pp. 1427–1431, 2018.

[29] Y. Su, J. Li, A. Plaza, A. Marinoni, P. Gamba, and S. Chakravortty, "Daen: Deep autoencoder networks for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 7, pp. 4309–4321, 2019.

[30] R. A. Borsoi, T. Imbiriba, and J. C. M. Bermudez, "Deep generative endmember modeling: An application to unsupervised spectral unmixing," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 374–384, 2020.

[31] Y. Qu and H. Qi, "udas: An untied denoising autoencoder with sparsity for spectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 3, pp. 1698–1712, 2019.

[32] X. Zhang, Y. Sun, J. Zhang, P. Wu, and L. Jiao, "Hyperspectral unmixing via deep convolutional neural networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 11, pp. 1755–1759, 2018.

[33] B. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Spectral-spatial hyperspectral unmixing using multitask learning," *IEEE Access*, vol. 7, pp. 148861–148872, 2019.

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

[35] B. Palsson, M. O. Ulfarsson, and J. R. Sveinsson, "Convolutional autoencoder for spectral-spatial hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–15, 2020.

[36] F. Khajehrayeni and H. Ghassemian, "Hyperspectral unmixing using deep convolutional autoencoders in a supervised scenario," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 567–576, 2020.

[37] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky, "Deep image prior," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[38] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," *International Journal of Computer Vision*, vol. 128, no. 7, pp. 1867–1888, Mar 2020.

[39] O. Sidorov and J. Y. Hardeberg, "Deep hyperspectral prior: Single-image denoising, inpainting, super-resolution," in *IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, 2019, pp. 3844–3851.

[40] J.M. Bioucas-Dias and J.M.P. Nascimento, "Hyperspectral subspace identification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 8, pp. 2435–2445, 2008.

[41] B. Rasti, M.O. Ulfarsson, and J.R. Sveinsson, "Hyperspectral subspace identification using SURE," *IEEE Geoscience and Remote Sensing Letter*, vol. 12, no. 12, pp. 2481–2485, Dec. 2015.

[42] M. D. Craig, "Minimum-volume transforms for remotely sensed data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no. 3, pp. 542–552, 1994.

[43] R. Heylen, D. Burazerovic, and P. Scheunders, "Fully constrained least squares spectral unmixing by simplex projection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4112–4122, Nov 2011.

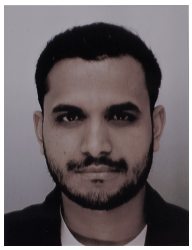
[44] B. Rasti, P. Scheunders, P. Ghamisi, G. Licciardi, and J. Chanussot, "Noise reduction in hyperspectral imagery: Overview and application," *Remote Sensing*, vol. 10, no. 3, pp. 482, 2018.

[45] B. Rasti, B. Koirala, P. Scheunders, and P. Ghamisi, "How hyperspectral image unmixing and denoising can boost each other," *Remote Sensing*, vol. 12, no. 11, pp. 1728, May 2020.

Behnood Rasti (Senior Member, IEEE) has received the B.Sc. and M.Sc. degrees both in Electronics-Electrical Engineering from the Electrical Engineering Department, University of Guilan, Rasht, Iran, in 2006, and 2009, respectively, and received the Ph.D. degree in Electrical and Computer Engineering from the University of Iceland, Reykjavik, Iceland, in 2014. He was the Valedictorian as an M.Sc. Student in 2009 and he won the Doctoral Grant of The University of Iceland Research Fund and was awarded "The Eimskip University fund", in 2013. In 2015 and 2016, Dr. Rasti worked as a postdoctoral researcher and a seasonal lecturer with the Electrical and Computer Engineering Department, University of Iceland. From 2016 to 2019, he has been a lecturer at the Center of Engineering Technology and Applied Sciences, Department of Electrical and Computer Engineering, University of Iceland, where he has taught several core courses such as Linear Systems, Control Systems, Sensors and Actuators, Data Acquisition and Processing, Circuit Theories, Electronics, and PLC programming.

In 2019, Dr. Rasti won the prestigious "Alexander von Humboldt Research Fellowship Grant" and started his work in 2020 as a Humboldt research fellow with the Machine Learning Group, Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Freiberg, Germany. His current research interests include signal and image processing, machine/deep learning, remote sensing image fusion, hyperspectral feature extraction and classification, spectral unmixing, remote sensing image denoising and restoration.

Dr. Rasti serves as an Associate Editor for the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS (GRSL) and MDPI-Remote Sensing.



Bikram Koirala received the B.S. degree in Geomatic Engineering from the Purbanchal University, Nepal, and the M.S. degree in Geomatic Engineering from the University of Stuttgart, Germany in 2011 and in 2016 respectively. In 2017, he joined Vision Lab, Department of Physics, the University of Antwerp as a Ph.D. researcher. He is a student IEEE member. His research interest includes machine learning and hyperspectral image processing.

Paul Scheunders (M'98) received the M.S. degree and the Ph.D. degree in physics, with work in the field of statistical mechanics, from the University of Antwerp, Antwerp, Belgium, in 1986 and 1990, respectively. In 1991, he became a research associate with the Vision Lab, Department of Physics, University of Antwerp, where he is currently a full professor. His current research interest includes remote sensing and hyperspectral image processing. He has published over 200 papers in international journals and proceedings in the field of image processing, pattern recognition, and remote sensing. Paul Scheunders is Associate Editor of the IEEE Transactions on Geoscience and Remote Sensing and has served as a program committee member in numerous international conferences. He is a senior member of the IEEE Geoscience and Remote Sensing Society.

Pedram Ghamisi (Senior Member, IEEE) graduated with a Ph.D. in electrical and computer engineering at the University of Iceland in 2015. He works as the head of the machine learning group at Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Germany and as the CTO, co-founder of VasoGnosis Inc, USA, and visiting professor at Institute of Advanced Research in Artificial Intelligence (IARAI), Vienna, Austria. He is also the co-chair of IEEE Image Analysis and Data Fusion Committee (IEEE IADF). Dr. Ghamisi was a recipient of the IEEE Mikio Takagi Prize for winning the Student Paper Competition at IEEE International Geoscience and Remote Sensing Symposium (IGARSS) in 2013, the first prize of the data fusion contest organized by the IEEE IADF in 2017, the Best Reviewer Prize of IEEE Geoscience and Remote Sensing Letters in 2017, and the IEEE Geoscience and Remote Sensing Society 2020 Highest-Impact Paper Award. His research interests include interdisciplinary research on machine (deep) learning, image and signal processing, and multisensor data fusion. For detailed info, please see http://pedram-ghamisi.com/.