Diffusion Model Based Hyperspectral Unmixing Using Spectral Prior Distribution

Diffusion Model Based Hyperspectral Unmixing Using Spectral Prior Distribution

Keli Deng, Yuntao Qian, *Senior Member, IEEE*, Jie Nie, *Member, IEEE*, and Jun Zhou, *Senior Member, IEEE*

***Abstract*—Hyperspectral unmixing is a crucial task for identifying the constituent materials and their respective distributions in a scene. Utilizing known spectral libraries as prior information, semi-blind unmixing methods (also known as spectral-library-based methods) have been proven advantageous over unblind and blind approaches. However, such methods encounter two main challenges: difficulty in handling large-scale spectral libraries and vulnerability to variabilities stemming from differences between the underlying signatures and those in the spectral library. To address these challenges, a novel approach named DiffUn, based on a diffusion model, is proposed in this article for semi-blind hyperspectral unmixing. DiffUn considers hyperspectral unmixing as a sampling process from a posterior distribution, where the prior distribution is learned from a spectral library, and the likelihood distribution is estimated from the observed data by the linear spectral mixture model. Specifically, the approach first learns the spectral prior distribution from a spectral library through an unconditional diffusion model, then integrates this prior knowledge into the reverse process of the diffusion model, and finally samples the underlying endmembers and corresponding abundances from the posterior distribution. Since spectral prior distribution estimation is not sensitive to library size, DiffUn exhibits superior unmixing performance even in a large-scale library. Furthermore, DiffUn permits sampling spectral signatures from a continuous probabilistic distribution, whereas conventional semi-blind unmixing methods only allow endmembers selected from the library, which is a discrete space. Thus, DiffUn shows greater robustness to spectral variations. Experimental results on synthetic and real-world datasets demonstrate DiffUn outperforming the state-of-the-art semi-blind unmixing methods. The code is available at https://github.com/Dmsw/DiffUn.git.**

***Index Terms*—Hyperspectral unmixing, diffusion model, spectral library.**

I. Introduction

HYPERSECTRAL images (HSIs) constitute a category of remote sensing images characterized by having numerous spectral bands, which provide a rich spectrum of information about the observed scene. Leveraging fine spectral information, HSIs have been utilized in various applications, such as land classification, object detection and tracking. However, due to the relatively low spatial resolution of the imaging instruments, many pixels within HSIs represent mixtures of multiple materials, posing challenges for advanced data analysis. Therefore, hyperspectral unmixing has been a significant task in the field of HSI processing. Its primary goal is to decompose the HSI into a collection of constituent materials, known as endmembers, along with their corresponding distributions, known as abundances. In numerous scenarios, the linear spectral mixture model (LMM) has been proven to be efficient and capable of modeling the mixing process of HSI [1], [2]. LMM assumes that the incident light interacts with only one material and is linearly mixed in the measuring instrument [1]. The unmixing task modeled by LMM can be regarded as an inverse problem to restore endmembers and abundances. According to whether the underlying endmembers are known, hyperspectral unmixing can be roughly divided into three categories, i.e., unblind unmixing, blind unmixing and semi-blind unmixing.

Unblind unmixing methods assume that the underlying endmembers are already known, and the unmixing problem is solved by estimating the abundance matrix. These methods typically treat unmixing as a linear regression problem and solve it by minimizing the least square error. The orthogonal subspace projection (OSP) has been demonstrated to be an effective solution when there is no noise in the HSI and the number of signatures is small [3]. However, in order to achieve a more optimal unmixing performance when there is noise present or the number of signatures is large, the abundance nonnegative constraint (ANC) and the abundance sum-to-one constraint (ASC) are commonly considered. The nonnegative constrained least square (NCLS) [3], [4] and the fully constrained least square (FCLS) [5] are two classical methods that have been employed to address the unblind unmixing problem in these cases. However, in practical scenarios, the endmembers are generally unknown, thereby restricting the applicability of these unblind methods.

Blind unmixing methods solve the unmixing problem without any information about the underlying endmembers. By assuming the presence in the observed HSI of at least one pure pixel per material, the endmembers can be extracted by geometrical-based algorithms, such as vertex component analysis (VCA) [6] and N-FINDR [7]. Nevertheless, the pure pixel assumption does not always hold in reality. The nonnegative matrix factorization (NMF) based methods do not depend on pure pixel assumption and they are widely used for hyperspectral unmixing [8]. To reduce the occurrence of local optima and enhance the performance of NMF-based methods, the sparsity constraint in the abundance matrix is widely explored [9]–[11]. Minimum volume constraint [12] and spatial and spectral smoothness constraints are also investigated [13]. In order to better capture the spectral-spatial joint information,

This work was supported by the National Natural Science Foundation of China 62071421. (Corresponding author: Yuntao Qian)

K. Deng and Y. Qian are with the Institute of Artificial Intelligence, College of Computer Science, Zhejiang University, Hangzhou, China. (email: kldeng@zju.edu.cn, ytqian@zju.edu.cn)

J. Nie is with the Department of Computer Science and Technology, Ocean University of China, Qingdao 266100, China (email: niejie@ouc.edu.cn)

J. Zhou is with the School of Information and Communication Technology, Griffith University, Nathan, Australia. (email: jun.zhou@griffith.edu.au)

NMF-based methods have been extended to nonnegative tensor factorization (NTF) and its variations [14]. Recently, much research has been devoted to the adoption of neural networks for solving the blind unmixing problem, leveraging the amount of training data [15]–[17]. However, rare and imprecise prior information seriously impedes the unmixing quality promotion of blind methods.

Semi-blind unmixing approaches assume that the underlying signatures exist within a known spectral library, e.g. the United States Geological Survey (USGS) spectral library. Thus, these methods amount to finding the optimal subset of signatures in the spectral library. The spectral library provides valuable spectral prior information and enhances the unmixing performance. Since the number of endmembers in the observed HSI is significantly smaller than the size of the library, most semi-blind unmixing methods are based on sparse regression [18]. The sparse unmixing by variable splitting and augmented Lagrangian (SUnSAL) utilizes the L1 norm to measure the sparsity of the abundance matrix and introduces an effective solution based on the alternating direction method of multipliers (ADMM) [19]. Other sparsity constraints involved in semi-blind unmixing include the total variation regularization [20] and L0 norm [2]. The pruning-based multi-objective sparse unmixing (PMoSU) prunes the library to reduce its size utilizing singular value decomposition (SVD) [21]. Since the object function is non-convex and incontinuous, PMoSU optimizes the problem using a genetic algorithm (GA). A Multi-object group sparse hyperspectral unmixing (MO-GSU) introduces a group normalization which provides better sparsity for abundance matrix [22]. However, most of these methods are non-convex and sensitive to the regularization parameters. Another challenge faced by semi-blind unmixing approaches is the spectral variability caused by illumination conditions, atmospheric effects, intrinsic variability and other factors [23]–[25]. Owing to this spectral variability, the underlying real signatures may not align perfectly with those in the spectral library. The sparse redundant unmixing model (SpaRedU) posits that the redundant error primarily arises from shifting and scaling of the absorption peak [26]. Assuming these variabilities are sparse within the HSI, SpaRedU models the variability by applying an L1 regularization to the residual error. The extended linear mixing model (ELMM) involves pixel-wise scaling factors for all the signatures to perceive the scaling variability [27]. The augmented linear mixing model (ALMM) utilizes a dictionary, which is orthogonal to the spectral library, to model complex variability [28]. Similarly, spectral variability augmented sparse unmixing (SVASU) extracts a variability dictionary by principal component analysis (PCA) [29]. However, these methods are incapable of fully exploring the spectral information within a large spectral library.

In response to the challenges of semi-blind unmixing in handling large-size spectral libraries and spectral variabilities between the underlying real signatures and those in the library, this article presents a novel diffusion-model-based unmixing method (DiffUn) for hyperspectral unmixing. Diffusion models (DMs) are a class of unconditional generation models that generate samples from a learned prior distribution of data [30]–[32]. Although the adoption of DMs for image inverse problems has been studied, including super-resolution [33], inpainting [34], and denoising [35], they cannot be directly extended to solve unmixing problem (which will be explained in Section II). To our knowledge, the proposed DiffUn is the first time to apply DM for hyperspectral unmixing. Specifically, DiffUn utilizes an unconditional DM trained on a pre-existing spectral library to learn the spectral prior distribution, then integrates this prior knowledge into the reverse process of DiffUn, and finally, the hyperspectral unmixing problem is addressed by sampling from a posterior distribution utilizing this reverse process. The framework is illustrated in Fig. 1. The major contributions of this article are summarized as follows:

- A novel diffusion-model-based hyperspectral semi-blind unmixing algorithm, DiffUn, is proposed, which learns the spectral prior distribution from a given spectral library. This idea is completely different from other semi-blind unmixing methods owing to its new probabilistic distribution estimation, inference, and sampling mechanisms.
- DiffUn can handle large-scale spectral libraries by formulating the hyperspectral unmixing as a generation process instead of a sparse regression problem and leveraging DM to learn the spectral prior distribution from the library.
- DiffUn has the capability of adapting to the spectral variations between the underlying real signatures and those in the spectral library, as it allows us to sample from a continuous spectral prior distribution.
- Exhaustive experiments on synthetic and real datasets demonstrate the effectiveness of DiffUn and its advantages over state-of-the-art semi-blind unmixing algorithms.

This article is organized as follows. Section II introduces the LMM, unconditional DM, and DMs for image restoration, indicating their constraints for hyperspectral unmixing. Section III derives the pipeline of DiffUn and elucidates the implementation details. Section IV reports the experimental results on both synthetic and real datasets. Section V concludes this article.

II. Background

In this section, we reveal the potential extension of DMs for hyperspectral unmixing and investigate the restriction of existing DMs in the realm of hyperspectral unmixing. Firstly, we introduce the LMM for hyperspectral unmixing. Next, we present the unconditional DM, which is capable of learning the prior distribution of data. Lastly, we introduce DMs for image restoration, leveraging the prior distribution learned by the unconditional DM, and explicate the challenges to adopt them for hyperspectral unmixing.

A. Linear Spectral Mixture Model

Let $\mathbf{X} \in \mathbb{R}^{N\times C}$ be the clean HSI with N pixels and C spectral bands, $\mathbf{Y} \in \mathbb{R}^{N\times C}$ be the corresponding observed HSI, and $\mathbf{S} = [\mathbf{S}^{(1)}, \mathbf{S}^{(2)}, \ldots, \mathbf{S}^{(R)}] \in \mathbb{R}^{C\times R}$ denotes the endmember matrix, where R is the number of endmembers and $\mathbf{S}^{(r)} \in \mathbb{R}^{C}$ is the r-th endmember signature. $\mathbf{H} \in \mathbb{R}^{N\times R}$

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2024.3408475

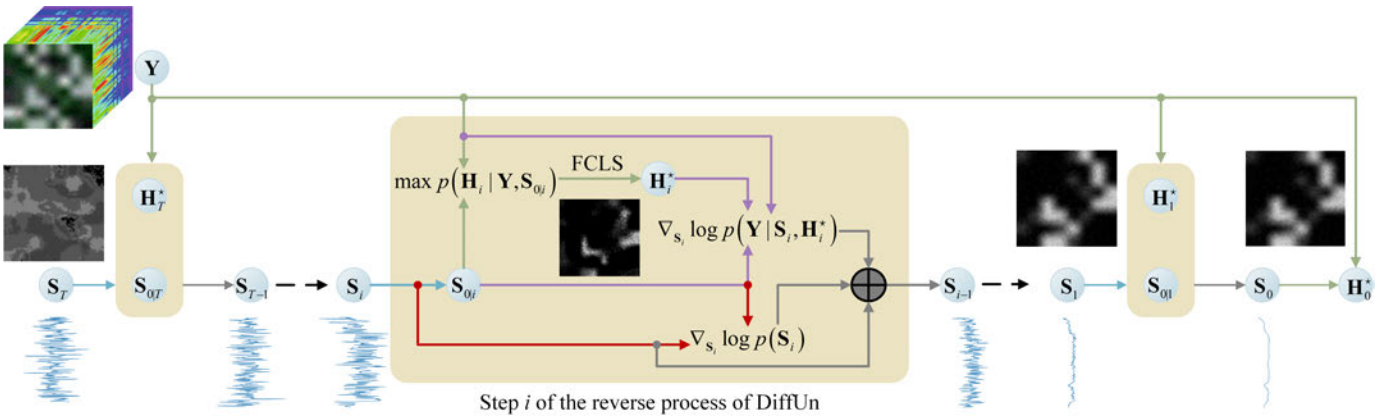


Fig. 1: The reverse process of the proposed DiffUn method. At each timestep i, the estimated endmember matrix $\mathbf{S}_{0|i}$ is firstly derived from the noisy endmember matrix $\mathbf{S}_i$ through the 1D denoising U-Net. The estimated abundance matrix $\mathbf{H}_i^\star$ is computed by maximizing the posterior probabilistic $p(\mathbf{H}_i|\mathbf{Y},\mathbf{S}_{0|i})$, which is solved by FCLS algorithm. $\nabla_{\mathbf{S}_i}\log p(\mathbf{Y}|\mathbf{S}_i,\mathbf{H}_i^\star)$ and $\nabla_{\mathbf{S}_i}\log p(\mathbf{S}_i)$ are calculated based on the values of $\mathbf{Y}$, $\mathbf{S}_{0|i}$ and $\mathbf{H}_i^\star$. Finally, $\mathbf{S}_{i-1}$ is sampled from $p(\mathbf{S}_{i-1}|\mathbf{S}_i,\mathbf{Y})$. After T iterations, the endmember matrix $\mathbf{S}_0$ is sampled from the posterior distribution $p(\mathbf{S}_0|\mathbf{Y})$, and the corresponding abundance matrix $\mathbf{H}_0^\star$ can be estimated.

is the abundance matrix and $\mathbf{E}\in\mathbb{R}^{N\times C}$ is the corresponding noise matrix. The LMM can be expressed as

$$\begin{aligned}&\mathbf{Y}=\mathbf{X}+\mathbf{E}=\mathbf{H}\mathbf{S}^{\mathrm{T}}+\mathbf{E}\\&\text{s.t.}\quad \mathbf{H}\succeq 0,\quad \mathbf{H}\mathbf{1}_R^{\mathrm{T}}=\mathbf{1}_N\end{aligned}\tag{1}$$

where $\succeq$ is the element-wise comparisons operator, $\mathbf{1}_N\in\mathbb{R}^N$ and $\mathbf{1}_R\in\mathbb{R}^R$ are the column vectors with all elements equal to one, $\mathbf{E}$ is generally modeled by Gaussian distribution and $E_{ij}\sim\mathcal{N}(0,\sigma_{\mathbf{Y}}^2)$. $\sigma_{\mathbf{Y}}$ is the standard deviation of the noise. The hyperspectral unmixing problem modeled by LMM can be considered as an inverse problem to restore the $\mathbf{S}$ and $\mathbf{H}$ from $\mathbf{Y}$.

In ordinary semi-blind unmixing, the underlying real signatures in $\mathbf{S}$ are assumed to present in a spectral library $\mathbf{A}$ with P signatures. However, it should be noted that there can be discrepancies between the underlying signatures and those in the spectral library, rendering it impractical to simply extract the signatures from the library to construct the endmember matrix $\mathbf{S}$. As a result, it becomes essential to learn the spectral prior distribution from the spectral library and leverage this prior knowledge to address the spectral variability and solve the unmixing problem.

B. Unconditional diffusion model

The unconditional DM employs a neural network to learn the prior distribution of data and stochastically generates samples from this distribution. It consists of two Markov processes, i.e., forward process and reverse process. The forward process progressively adds Gaussian noise to the data, which can be expressed as a stochastic difference equation (SDE) as follows [32],

$$d\mathbf{x}_t=-\frac{\beta_t}{2}\mathbf{x}_t dt+\sqrt{\beta_t}d\mathbf{w}\tag{2}$$

where the timestep t is in the range $[0,1]$, the initial data point $\mathbf{x}_0$ follows the distribution of the underlying dataset p_{data}, β_t controls the amount of noise added at timestep t, $\mathbf{w}$ is standard Brownian motion and $\mathbf{x}_t$ is the noisy data at timestep t. Given $\int_0^1\frac{\beta_\tau}{2}\mathbf{d}\tau\to+\infty$, we have $\mathbf{x}_1\sim\mathcal{N}(\mathbf{0},\mathbf{I})$ at the end of the forward process. Then, we can reverse the above process by starting with a randomly noisy data $\mathbf{x}_1\sim\mathcal{N}(\mathbf{0},\mathbf{I})$ and generate a sample $\mathbf{x}_0$ from p_{data} using the following SDE [36],

$$d\mathbf{x}_t=-\left(\frac{\beta_t}{2}\mathbf{x}_t+\beta_t\nabla_{\mathbf{x}_t}\log p_t(\mathbf{x}_t)\right)dt+\sqrt{\beta_t}d\mathbf{w}\tag{3}$$

where $\nabla_{\mathbf{x}_t}\log p_t(\mathbf{x}_t)$ can be approximated by a denoising neural network $s_\theta(\mathbf{x}_t,t)$. Specifically, we have $s_\theta(\mathbf{x}_t,t)\simeq\nabla_{\mathbf{x}_t}\log p_t(\mathbf{x}_t)$ by minimizing the following loss function [37],

$$L(\theta)=\mathbb{E}_{\mathbf{x}_t,\mathbf{x}_0,t}\|\nabla_{\mathbf{x}_t}\log p_t\left(\mathbf{x}_t|\mathbf{x}_0\right)-s_\theta(\mathbf{x}_t,t)\|_2^2\tag{4}$$

s_θ is typically implemented by a U-Net architecture.

C. Diffusion model for image restoration

Image restoration is a typical inverse problem, which recovers the ground truth image $\mathbf{x}$ from the observed image $\mathbf{y}$. Generally, the degradation model for image restoration can be expressed as

$$\mathbf{y}=\mathbf{H}\mathbf{x}+\mathbf{e}\tag{5}$$

where matrix $\mathbf{H}$ represent the degradation kernel, vector $\mathbf{e}\sim\mathcal{N}(\mathbf{0},\sigma_{\mathbf{y}}^2\mathbf{I})$ represent the *i.i.d.* Gaussian noise. $\sigma_{\mathbf{y}}$ is the standard deviation of the noise. The above restoration task can be regarded as an inverse problem and roughly categorized into two types based on whether $\mathbf{H}$ is known, i.e., unblind inverse problem and blind inverse problem.

The capability of DMs to solve unblind inverse problems has been widely explored. These approaches regard this problem as a generation task of sampling from the posterior distribution $p(\mathbf{x}|\mathbf{y})$. This can be accomplished by plugging the

gradient of log-likelihood distribution $\nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{x}_t)$ into the reverse SDE in Eq. (3) [38], which is

$$d\mathbf{x}_t = -\frac{\beta_t}{2}\mathbf{x}_t dt - \beta_t(\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t) + \zeta_t \nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{x}_t)) + \sqrt{\beta_t} d\mathbf{w} \tag{6}$$

where ζ_t is a weight factor and $\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t)$ can be approximated by $s_\theta(\mathbf{x}_t, t)$. Leveraging the neural network s_θ, the prior information is incorporated for solving the inverse problem.

$\nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{x}_t)$ can be estimated by $\nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{x}_{0|t})$, where $\mathbf{x}_{0|t} = \mathbb{E}[\mathbf{x}_0|\mathbf{x}_t]$ [39]–[41]. Combining with the degradation model in Eq. (5), we obtain

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{x}_t) \simeq -\frac{1}{\sigma_{\mathbf{y}}^2} \nabla_{\mathbf{x}_t} \|\mathbf{y} - \mathbf{x}_{0|t}\|_2^2 \tag{7}$$

For blind inverse problems, most methods use supervised learning with sufficient paired data $(\mathbf{y}, \mathbf{x})$ to train a conditional DM but they can only estimate $\mathbf{x}$ and ignore the degradation model. There is little research on simultaneously estimating $\mathbf{H}$ and $\mathbf{x}$. The blind diffusion posterior sampling (BlindDPS) utilizes two parallel DMs to respectively generate $\mathbf{H}$ and $\mathbf{x}$ [42], which are

$$d\mathbf{x}_t = -\frac{\beta_t}{2}\mathbf{x}_t - \beta_t(\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t) + \nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{x}_t, \mathbf{H}_t))dt + \sqrt{\beta_t} d\mathbf{w} \tag{8}$$

$$d\mathbf{H}_t = -\frac{\beta_t}{2}\mathbf{H}_t + \beta_t(\nabla_{\mathbf{H}_t} \log p(\mathbf{H}_t) + \nabla_{\mathbf{H}_t} \log p(\mathbf{y}|\mathbf{x}_t, \mathbf{H}_t))dt + \sqrt{\beta_t} d\mathbf{W} \tag{9}$$

However, training a neural network $s_\theta(\mathbf{H}_t, t)$ to approximate $\nabla_{\mathbf{H}_t} \log p(\mathbf{H}_t)$ requires a large amount of data.

As the degradation model in Eq. (5) is similar to the LMM in Eq. (1), it is possible to extend DMs for hyperspectral unmixing. However, there is limited abundance and endmember data for training a conditional DM or BlindDPS. It is necessary to develop a novel DM for hyperspectral unmixing. In the next section, we introduce the proposed DiffUn for hyperspectral unmixing. It only requires spectral library data for training an unconditional DM, which serves as the spectral prior distribution, and is then transformed into a posterior distribution for solving the unmixing problem.

III. Diffusion-model-based Unmixing Method

In this section, DiffUn for semi-blind spectral unmixing is presented. Firstly, we introduce the forward process of the DiffUn which is extended from the unconditional DM. By discretizing and reversing the forward process, we then derive the reverse process of DiffUn. It solves the hyperspectral unmixing by sampling from a posterior distribution. Next, we discuss how to incorporate the spectral prior distribution and the likelihood distribution into the reverse process. Subsequently, we elucidate the solution of the abundance matrix during the reverse process. Finally, we present the implementation details of the reverse process of DiffUn, both with and without a neural network.

A. Forward process of DiffUn

Aiming to learn the spectral prior distribution, we apply an unconditional DM for a single signature $\mathbf{s}$ with C channels, which has the following forward SDE as Eq. (2),

$$d\mathbf{s}_t = -\frac{\beta_t}{2}\mathbf{s}_t dt + \sqrt{\beta_t} d\mathbf{w} \tag{10}$$

where $\mathbf{s}_t \in \mathbb{R}^C$ is the noisy signature at timestep t. By solving the SDE in Eq. (10), we can directly yield $\mathbf{s}_t$ from a signature $\mathbf{s}_0$ in the spectral library $\mathbf{A}$, which is

$$\mathbf{s}_t = \sqrt{\bar{\alpha}_t}\mathbf{s}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon} \tag{11}$$

where $\sqrt{\bar{\alpha}_t} = \exp(-\int_0^t \frac{\beta_\tau}{2} d\tau)$ and $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is the Gaussian noise. Consequently, the loss function in Eq. (4) can be written as

$$L(\theta) = \mathbb{E}_{\boldsymbol{\epsilon}, \mathbf{s}_0, t} \|\boldsymbol{\epsilon} - \sqrt{1-\bar{\alpha}_t} s_\theta(\sqrt{\bar{\alpha}_t}\mathbf{s}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t)\|_2^2 \tag{12}$$

This is exactly a loss function for spectral denoising. Many neural networks can be used for optimizing this loss function, and we employ a U-Net to implement s_θ. After training, s_θ learns the spectral prior distribution from the spectral library. As the size of the spectral library increases, s_θ can learn more spectral information from the library, and establish a more accurate spectral prior distribution. This indicates that DiffUn is not sensitive to the size of the spectral library.

The above DM establishes the prior distribution of the spectral library $\mathbf{A}$ and is subsequently expanded to DiffUn for modeling the prior distribution of the endmember matrix. Utilizing Euler method, we can discretize the SDE in Eq. (10) as

$$\mathbf{s}_{t+\Delta t} = \mathbf{s}_t - \frac{\beta_t}{2}\mathbf{s}_t \Delta t + \sqrt{\beta_t \Delta t}\boldsymbol{\epsilon} \tag{13}$$

where Δt is the step size and each element of ϵ following the standard Gaussian distribution. Thus, we can obtain the transfer probability of a single endmember as

$$p(\mathbf{s}_{t+\Delta t}|\mathbf{s}_t) \propto \exp\left(-\frac{\|\mathbf{s}_{t+\Delta t} - \mathbf{s}_t - \frac{\beta_t}{2}\mathbf{s}_t\|_2^2}{2\beta_t \Delta t}\right) \tag{14}$$

In the forward process, it is reasonable to assume the endmembers are independent of each other. Thereby, we can derive the transfer probability of the endmember matrix $\mathbf{S}_t \in \mathbb{R}^{C \times R}$ in the forward process as

$$p(\mathbf{S}_{t+\Delta t}|\mathbf{S}_t) = \prod_{r=1}^{R} p(\mathbf{S}_{t+\Delta t}^{(r)}|\mathbf{S}_t^{(r)}) \tag{15}$$

$$\propto \exp\left(-\frac{\|\mathbf{S}_{t+\Delta t} - \mathbf{S}_t - \frac{\beta_t}{2}\mathbf{S}_t\|_{\mathrm{F}}^2}{2\beta_t \Delta t}\right) \tag{16}$$

Therefore, while $\Delta t \to 0$, we can derive the SDE for the forward process of DiffUn as

$$d\mathbf{S}_t = -\frac{\beta_t}{2}\mathbf{S}_t dt + \sqrt{\beta_t} d\mathbf{W} \tag{17}$$

where each element of $\mathbf{W}$ is a standard Brownian motion. Due to the endmembers independence assumption, we can define the function s_θ for the endmember matrix as

$$s_\theta(\mathbf{S}_t, t) \triangleq [s_\theta(\mathbf{S}_t^{(1)}, t), s_\theta(\mathbf{S}_t^{(2)}, t), \ldots, s_\theta(\mathbf{S}_t^{(R)}, t)] \tag{18}$$

and we have

$$\nabla_{\mathbf{S}_t} \log p(\mathbf{S}_t) \simeq s_\theta(\mathbf{S}_t, t) \tag{19}$$

By the forward process of DiffUn, the prior distribution $p_{\mathbf{A}}(\mathbf{S})$ undergoes a progressive perturbation and transforms into Gaussian noise. Guided by $\nabla_{\mathbf{S}_t} \log p(\mathbf{S}_t)$, we can yield samples from the original spectral prior distribution. In the next section, we will prove that this forward process can be inverted and utilized to solve the unmixing problem with the aid of a suitably designed strategy.

B. Reverse process of DiffUn

Given the gradually corrupted prior distributions $p(\mathbf{S}_t)$ derived from the forward process, we can reverse this forward process and guide it towards a posterior distribution $p(\mathbf{S}|\mathbf{Y})$. The reverse process is initialized from the standard Gaussian noise and ultimately yields the endmember matrix from $p(\mathbf{S}|\mathbf{Y})$, which is a probable solution to the hyperspectral unmixing. To obtain the reverse process of DiffUn, we first derive the transfer probability of the reverse process as

$$\begin{aligned} &p(\mathbf{S}_{t-\Delta t}|\mathbf{S}_t, \mathbf{Y}) \\ &= \int p(\mathbf{S}_{t-\Delta t}|\mathbf{S}_t, \mathbf{Y}, \mathbf{H}_t) p(\mathbf{H}_t|\mathbf{S}_t, \mathbf{Y}) d\mathbf{H}_t \qquad (20) \\ &= \int \frac{p(\mathbf{S}_t|\mathbf{S}_{t-\Delta t}) p(\mathbf{S}_{t-\Delta t}|\mathbf{Y}, \mathbf{H}_t)}{p(\mathbf{S}_t|\mathbf{Y}, \mathbf{H}_t)} p(\mathbf{H}_t|\mathbf{S}_t, \mathbf{Y}) d\mathbf{H}_t \qquad (21) \\ &\simeq p(\mathbf{S}_t|\mathbf{S}_{t-\Delta t}) \mathbb{E}_{\mathbf{H}_t \sim p(\mathbf{H}_t|\mathbf{S}_{0|t}, \mathbf{Y})} \left[\frac{p(\mathbf{S}_{t-\Delta t}|\mathbf{Y}, \mathbf{H}_t)}{p(\mathbf{S}_t|\mathbf{Y}, \mathbf{H}_t)} \right] \qquad (22) \end{aligned}$$

Eq. (20) is derived from the law of total probability by involving variable $\mathbf{H}_t$ which represents the abundance matrix at timestep t. Eq. (21) is derived from Bayes' rule and the assumption of the Marchov process. To obtain the Eq. (22) from Eq. (21), we utilize the trick in [39]–[41], which is to approximate $p(\mathbf{H}_t|\mathbf{S}_t, \mathbf{Y})$ by $p(\mathbf{H}_t|\mathbf{S}_{0|t}, \mathbf{Y})$. $\mathbf{S}_{0|t} \triangleq \mathbb{E}[\mathbf{S}_0|\mathbf{S}_t]$ represents the estimation of $\mathbf{S}_0$ at timestep t given $\mathbf{S}_t$, which can be derived by Tweedie's formula as follows [39],

$$\begin{aligned} \mathbf{S}_{0|t} &= \frac{1}{\sqrt{\bar{\alpha}_t}} (\mathbf{S}_t + (1 - \bar{\alpha}_t) \nabla_{\mathbf{S}_t} \log p(\mathbf{S}_t)) \qquad (23) \\ &\simeq \frac{1}{\sqrt{\bar{\alpha}_t}} (\mathbf{S}_t + (1 - \bar{\alpha}_t) s_\theta(\mathbf{S}_t, t)) \qquad (24) \end{aligned}$$

According to the loss function in Eq. (12), $\mathbf{S}_{0|t}$ calculated by Eq. (24) is an approximation of the ground truth endmember matrix $\mathbf{S}_0$. Therefore, we can further simplify the expectation in Eq. (22) by assuming that $p(\mathbf{H}_t|\mathbf{S}_{0|t}, \mathbf{Y})$ only has a single narrow peak at $\mathbf{H}_t^\star$, which could be estimated by a deterministic function of $\mathbf{S}_{0|t}$ and $\mathbf{Y}$. Specifically, $p(\mathbf{H}_t|\mathbf{S}_{0|t}, \mathbf{Y})$ can be approximated by a delta function $\delta(\mathbf{H}_t - \mathbf{H}_t^\star)$. This is reasonable because, for most cases, the optimal abundance matrix is deterministic while the HSI and endmember matrix are given. Utilizing the property of the delta function, we eventually obtain the transfer probability of the reverse process of DiffUn as

$$p(\mathbf{S}_{t-\Delta t}|\mathbf{S}_t, \mathbf{Y}) \simeq p(\mathbf{S}_t|\mathbf{S}_{t-\Delta t}) \frac{p(\mathbf{S}_{t-\Delta t}|\mathbf{Y}, \mathbf{H}_t^\star)}{p(\mathbf{S}_t|\mathbf{Y}, \mathbf{H}_t^\star)} \tag{25}$$

where $\mathbf{H}_t^\star$ is the optimal estimation of the abundance matrix $\mathbf{H}_t$ given $\mathbf{S}_{0|t}$ and $\mathbf{Y}$, which can be derived from the following optimization problem,

$$\mathbf{H}_t^\star = \arg\max_{\mathbf{H}_t} p(\mathbf{H}_t|\mathbf{S}_{0|t}, \mathbf{Y}) \tag{26}$$

When $\Delta t \to 0$, we yield the reverse SDE of DiffUn from the discrete-time equation as

$$d\mathbf{S}_t = -\left(\frac{\beta_t}{2} \mathbf{S}_t + \beta_t \nabla_{\mathbf{S}_t} \log p(\mathbf{S}_t|\mathbf{Y}, \mathbf{H}_t^\star) \right) dt + \sqrt{\beta_t} d\mathbf{W} \tag{27}$$

where $\mathbf{H}_t^\star$ is defined by Eq. (26) and $\mathbf{S}_t$ represents the endmember matrix at timestep t during the reverse process of DiffUn. The details of deriving the SDE in Eq. (27) from discrete-time Eq. (25) when $\Delta t \to 0$ are outlined in Appendix A.

C. Analysis of $\nabla_{\mathbf{S}_t} \log p(\mathbf{S}_t|\mathbf{Y}, \mathbf{H}_t^\star)$

Aiming to perform the reverse process of DiffUn in Eq. (27), we need to calculate $\nabla_{\mathbf{S}_t} \log p(\mathbf{S}_t|\mathbf{Y}, \mathbf{H}_t^\star)$. To decompose the posterior probability $p(\mathbf{S}_t|\mathbf{Y}, \mathbf{H}_t^\star)$, we involve the Bayes' rule and obtain

$$p(\mathbf{S}_t|\mathbf{Y}, \mathbf{H}_t^\star) = \frac{p(\mathbf{Y}|\mathbf{S}_t, \mathbf{H}_t^\star) p(\mathbf{S}_t|\mathbf{H}_t^\star)}{p(\mathbf{Y}|\mathbf{H}_t^\star)} \tag{28}$$

Under the assumption that the abundance matrix is typically independent of the endmember matrix when the observed HSI is not provided, we can infer that $\mathbf{S}_t$ is independent of $\mathbf{H}_t^\star$, thereby indicating that $p(\mathbf{S}_t|\mathbf{H}_t^\star) = p(\mathbf{S}_t)$. Similarly, we also assume that $\mathbf{Y}$ is independent of $\mathbf{H}_t^\star$, indicating that $p(\mathbf{Y}|\mathbf{H}_t^\star) = p(\mathbf{Y})$. Therefore, Eq. (28) can be rewritten as

$$p(\mathbf{S}_t|\mathbf{Y}, \mathbf{H}_t^\star) = \frac{p(\mathbf{Y}|\mathbf{S}_t, \mathbf{H}_t^\star) p(\mathbf{S}_t)}{p(\mathbf{Y})} \tag{29}$$

where $p(\mathbf{Y}|\mathbf{S}_t, \mathbf{H}_t^\star)$ is the likelihood, $p(\mathbf{S}_t)$ is the prior probability and $p(\mathbf{Y})$ is a normalization factor. Applying the logarithm and gradient operator to Eq. (29), we can derive $\nabla_{\mathbf{S}_t} \log p(\mathbf{S}_t|\mathbf{Y}, \mathbf{H}_t^\star)$ as

$$\nabla_{\mathbf{S}_t} \log p(\mathbf{S}_t|\mathbf{Y}, \mathbf{H}_t^\star) = \nabla_{\mathbf{S}_t} \log p(\mathbf{Y}|\mathbf{S}_t, \mathbf{H}_t^\star) + \nabla_{\mathbf{S}_t} \log p(\mathbf{S}_t) \tag{30}$$

Leveraging the Theorem 1 in [39], $\nabla_{\mathbf{S}_t} \log p(\mathbf{Y}|\mathbf{S}_t, \mathbf{H}_t^\star)$ can be approximated as

$$\nabla_{\mathbf{S}_t} \log p(\mathbf{Y}|\mathbf{S}_t, \mathbf{H}_t^\star) \simeq \nabla_{\mathbf{S}_t} \log p(\mathbf{Y}|\mathbf{S}_{0|t}, \mathbf{H}_t^\star) \tag{31}$$

According to the LMM in Eq. (1), we can explicitly express the likelihood $p(\mathbf{Y}|\mathbf{S}_{0|t}, \mathbf{H}_t^\star)$ as

$$p(\mathbf{Y}|\mathbf{S}_{0|t}, \mathbf{H}_t^\star) \propto \exp\left(-\frac{\|\mathbf{Y} - \mathbf{H}_t^\star \mathbf{S}_{0|t}^{\mathrm{T}}\|_{\mathrm{F}}^2}{2\sigma_{\mathbf{Y}}^2} \right) \tag{32}$$

Therefore, we can derive the gradient of the log-likelihood with respect to the r-th signature $\mathbf{S}_t^{(r)}$ as

$$\nabla_{\mathbf{S}_t^{(r)}} \log p(\mathbf{Y}|\mathbf{S}_{0|t}, \mathbf{H}_t^\star) = \frac{1}{\sigma_{\mathbf{Y}}^2} \mathbf{J}_t^{(r)} \left(\mathbf{Y} - \mathbf{H}_t^\star \mathbf{S}_{0|t}^{\mathrm{T}} \right)^{\mathrm{T}} \mathbf{H}_t^{\star(r)} \tag{33}$$

where $\mathbf{J}_t^{(r)} \in \mathbb{R}^{C\times C}$ is the Jacobian matrix of $\mathbf{S}_{0|t}^{(r)}$ with respect to $\mathbf{S}_t^{(r)}$. However, according to the experimental results in [43], $\mathbf{J}_t^{(r)}$ is time-consuming and has few benefits for solving the inverse problem. Therefore, we approximate $\mathbf{J}_t^{(r)}$ with $\frac{\sqrt{\bar{\alpha}_t}}{1-\bar{\alpha}_t}\gamma_t\mathbf{I}$ and we can derive

$$\nabla_{\mathbf{S}_t}\log p(\mathbf{Y}|\mathbf{S}_{0|t},\mathbf{H}_t^\star) = \frac{\sqrt{\bar{\alpha}_t}\gamma_t}{(1-\bar{\alpha}_t)\sigma_{\mathbf{Y}}^2}\left(\mathbf{Y}-\mathbf{H}_t^\star\mathbf{S}_{0|t}^{\mathrm{T}}\right)^{\mathrm{T}}\mathbf{H}_t^\star \tag{34}$$

The error analysis can be found in Appendix B.

To rectify the error arising from the approximation, a correction scalar for the gradient of log-likelihood has been proposed [39]–[42], [44]–[46]. Consequently, we can employ a weight ζ_t to replace $\frac{\sqrt{\bar{\alpha}_t}\gamma_t}{\sigma_{\mathbf{Y}}^2(1-\bar{\alpha}_t)}$ and ultimately obtain the following expression,

$$\nabla_{\mathbf{S}_t}\log p(\mathbf{S}_t|\mathbf{Y},\mathbf{H}_t^\star) = \zeta_t\left(\mathbf{Y}-\mathbf{H}_t^\star\mathbf{S}_{0|t}^{\mathrm{T}}\right)^{\mathrm{T}}\mathbf{H}_t^\star + s_\theta(\mathbf{S}_t,t) \tag{35}$$

By incorporating the pre-training neural network s_θ, which contains the prior information about the signatures learned from the spectral library, we transform the blind unmixing problem into a semi-blind unmixing problem. $(\mathbf{Y}-\mathbf{H}_t^\star\mathbf{S}_{0|t}^{\mathrm{T}})^{\mathrm{T}}\mathbf{H}_t^\star$ represents a gradient that directs to the increase of the likelihood $p(\mathbf{Y}|\mathbf{S}_{0|t},\mathbf{H}_t^\star)$. The guidance in the reverse process of DiffUn is the weighted sum of s_θ and $(\mathbf{Y}-\mathbf{H}_t^\star\mathbf{S}_{0|t}^{\mathrm{T}})^{\mathrm{T}}\mathbf{H}_t^\star$. Hence, it will generate the endmember matrix $\mathbf{S}_0$ that has a great probability to produce $\mathbf{Y}$ and simultaneously exhibits similarity to the signatures in the spectral library.

D. Estimation of $\mathbf{H}_t^\star$

To solve $\mathbf{H}_t^\star$ in Eq. (26), we apply Bayes' rule and the independence hypothesis between $\mathbf{H}_t$ and $\mathbf{S}_{0|t}$. Therefore, the problem can be converted to the subsequent optimization problem,

$$\mathbf{H}_t^\star = \arg\max_{\mathbf{H}_t} p(\mathbf{Y}|\mathbf{S}_{0|t},\mathbf{H}_t)p(\mathbf{H}_t) \tag{36}$$

Prior distribution $p(\mathbf{H}_t)$ is generally assumed a Dirichlet distribution. When the noise variance $\sigma_{\mathbf{Y}}^2$ in LMM is small, the specific parameter settings in the Dirichlet distribution become trivial. For the sake of efficiency and simplicity, we assume the parameters of the Dirichlet distribution are all equal to one. Consequently, Eq. (36) can be rewritten as

$$\begin{aligned}&\arg\min_{\mathbf{H}_t}\|\mathbf{Y}-\mathbf{H}_t\mathbf{S}_{0|t}^{\mathrm{T}}\|_{\mathrm{F}}^2\\ &\text{s.t.}\quad \mathbf{H}_t\succeq 0,\quad \mathbf{H}_t\mathbf{1}_R^{\mathrm{T}}=\mathbf{1}_N\end{aligned} \tag{37}$$

Several algorithms are available to address the optimization problem in Eq. (37). In this article, we adopt the FCLS algorithm [5], which is a fast and efficient algorithm.

E. Implementation of Reverse Process

Here, we solv the reverse SDE of DiffUn in Eq. (27) under the framework of denoising diffusion probability model (DDPM) [30]. The range of t is discretized into T intervals, and i represent the i-th interval. The pseudocode is presented in Algorithm 1, where $\{\tilde{\sigma}_i\}_{i=1}^T$ is the posterior variance in the DDPM. The algorithm does not require a well-designed initialization and it can start from a random state sampled from the standard Gaussian distribution. However, aiming to further shrink the solution space and accelerate the sampling process, we initialize the endmember matrix $\mathbf{S}_0$ with the result of VCA [6] and start the reverse process of DiffUn from timestep $i_{\text{start}} < T$.

Algorithm 1 Reverse process of DiffUn with learning

Input: $\mathbf{Y}$, T, R, $\{\zeta_i\}_{i=1}^T$, $\{\tilde{\sigma}_i\}_{i=1}^T$, s_θ and i_{start}.
Output: $\mathbf{S}_0$ and $\mathbf{H}_0^\star$.

1: $\boldsymbol{\epsilon}\sim\mathcal{N}(\mathbf{0},\mathbf{I})$;
2: **if** $i_{\text{start}} < T$ **then**
3: Estimate $\hat{\mathbf{S}}$ with R endmembers from $\mathbf{Y}$ using VCA;
4: $\mathbf{S}_{i_{\text{start}}} \leftarrow \sqrt{\bar{\alpha}_{i_{\text{start}}}}\hat{\mathbf{S}} + \sqrt{1-\bar{\alpha}_{i_{\text{start}}}}\boldsymbol{\epsilon}$;
5: **else**
6: $\mathbf{S}_T \leftarrow \boldsymbol{\epsilon}$;
7: **end if**
8: **for** $i \leftarrow i_{\text{start}}$ to 1 **do**
9: $\mathbf{S}_{0|i} \leftarrow \frac{1}{\sqrt{\bar{\alpha}_i}}(\mathbf{S}_i + (1-\bar{\alpha}_i)s_\theta(\mathbf{S}_i,i))$;
10: Calculate $\mathbf{H}_i^\star$ using FCLS;
11: $\boldsymbol{\epsilon}\sim\mathcal{N}(\mathbf{0},\mathbf{I})$;
12: $\mathbf{S}'_{i-1} \leftarrow \frac{\sqrt{\bar{\alpha}_{i-1}}\beta_i}{1-\bar{\alpha}_i}\mathbf{S}_{0|i} + \frac{\sqrt{\alpha_i}(1-\bar{\alpha}_{i-1})}{1-\bar{\alpha}_i}\mathbf{S}_i + \tilde{\sigma}_i\boldsymbol{\epsilon}$;
13: $\mathbf{S}_{i-1} \leftarrow \mathbf{S}'_{i-1} - \zeta_i\nabla_{\mathbf{S}_i}\log p(\mathbf{Y}|\mathbf{S}_i,\mathbf{H}_i^\star)$;
14: **end for**
15: Calculate $\mathbf{H}_0^\star$ using FCLS based on $\mathbf{S}_0$ and $\mathbf{Y}$;

Algorithm 1 requires learning s_θ with the spectral library, which is time-consuming. Alternatively, we also introduce the reverse process of DiffUn without the necessity of training s_θ, presented in Algorithm 2. In Algorithm 1, the role of s_θ is to predict $\mathbf{S}_{0|i}$ in line 9. When the spectral library is provided, we can establish the spectral prior distribution $p_{\mathbf{A}}$ with multi-point distribution and obtain $\mathbf{S}_{0|i}$ without the involvement of $\mathbf{s}_\theta$ [47], which is

$$\mathbf{S}_{0|i}^{(r)} = \frac{\int \mathbf{s}_0 p(\mathbf{S}_i^{(r)}|\mathbf{s}_0)p_{\mathbf{A}}(\mathbf{s}_0)d\mathbf{s}_0}{\int p(\mathbf{S}_i^{(r)}|\mathbf{s}_0)p_{\mathbf{A}}(\mathbf{s}_0)d\mathbf{s}_0} \tag{38}$$

We have $p(\mathbf{S}_i^{(r)}|\mathbf{s}_0)\sim\mathcal{N}(\sqrt{\bar{\alpha}_i}\mathbf{s}_0,(1-\bar{\alpha}_i)\mathbf{I})$ and

$$p_{\mathbf{A}}(\mathbf{s}) \simeq \frac{1}{P}\sum_{k=1}^{P}\delta(\mathbf{s}-\mathbf{A}^{(k)}) \tag{39}$$

where $\mathbf{A}$ is the spectral library, and $\mathbf{A}^{(k)}$ represents the k-th signature. $\mathbf{S}_{0|i}^{(r)}$ can be approximated as

$$\mathbf{S}_{0|i}^{(r)} \simeq \frac{1}{Z^{(r)}}\sum_{k=1}^{P}\mathbf{A}^{(k)}\exp\left(-\frac{\|\sqrt{\bar{\alpha}_i}\mathbf{A}^{(k)}-\mathbf{S}_i^{(r)}\|_2^2}{2(1-\bar{\alpha}_i)}\right) \tag{40}$$

where $Z^{(r)}$ is the normalization factor

$$Z^{(r)} = \sum_{k=1}^{P}\exp\left(-\frac{\|\sqrt{\bar{\alpha}_i}\mathbf{A}^{(k)}-\mathbf{S}_i^{(r)}\|_2^2}{2(1-\bar{\alpha}_i)}\right) \tag{41}$$

Different from Algorithm 1, which uses s_θ to predict $\mathbf{S}_{0|i}$, Algorithm 2 estimates $\mathbf{S}_{0|i}$ utilizing Eq. (40) and Eq. (41),

which is faster than Algorithm 1. The estimation of $p_{\mathbf{A}}$ in Eq. (39) will approach to the true spectral prior distribution as P increases, thereby demonstrating the robustness of Algorithm 2 to the library size P. Eq. (40) approximates $\mathbf{S}_{0|i}$ by a weighted average of the library signatures, where the weights are measured by the similarity between the noisy endmember and the library signatures using the Gaussian kernel. $1-\bar{\alpha}_i$ controls the shape of Gaussian function. When $i=0$ and $1-\bar{\alpha}_i=0$, $\mathbf{S}_{0|0}^{(r)}$ will be identical to one of the library signatures. This indicates that the Algorithm 2 is not capable of adapting spectral variations between the underlying real signatures and those in the library.

Algorithm 2 Reverse process of DiffUn without learning

Input: $\mathbf{Y}$, T, R, $\{\zeta_i\}_{i=1}^T$, $\{\tilde{\sigma}_i\}_{i=1}^T$, $\mathbf{A}$ and i_{start}.
Output: $\mathbf{S}_0$ and $\mathbf{H}_0^\star$.

$\boldsymbol{\epsilon}\sim\mathcal{N}(\mathbf{0},\mathbf{I})$;
if $i_{\text{start}}<T$ **then**
  Estimate $\hat{\mathbf{S}}$ with R endmembers from $\mathbf{Y}$ using VCA;
  $\mathbf{S}_{i_{\text{start}}}\leftarrow\sqrt{\bar{\alpha}_{i_{\text{start}}}}\hat{\mathbf{S}}+\sqrt{1-\bar{\alpha}_{i_{\text{start}}}}\boldsymbol{\epsilon}$;
else
  $\mathbf{S}_T\leftarrow\boldsymbol{\epsilon}$;
end if
for $i\leftarrow i_{\text{start}}$ to 1 **do**
  for $r\leftarrow 1$ to R **do**
    $Z^{(r)}=\sum_{k=1}^{P}\exp\left(-\frac{\|\sqrt{\bar{\alpha}_i}\mathbf{A}^{(k)}-\mathbf{S}_i^{(r)}\|_2^2}{2(1-\bar{\alpha}_i)}\right)$;
    $\mathbf{S}_{0|i}^{(r)}\leftarrow\frac{1}{Z^{(r)}}\sum_{k=1}^{P}\mathbf{A}^{(k)}\exp\left(-\frac{\|\sqrt{\bar{\alpha}_i}\mathbf{A}^{(k)}-\mathbf{S}_i^{(r)}\|_2^2}{2(1-\bar{\alpha}_i)}\right)$;
  end for
  Calculate $\mathbf{H}_i^\star$ using FCLS;
  $\boldsymbol{\epsilon}\sim\mathcal{N}(\mathbf{0},\mathbf{I})$;
  $\mathbf{S}'_{i-1}\leftarrow\frac{\sqrt{\bar{\alpha}_{i-1}}\beta_i}{1-\bar{\alpha}_i}\mathbf{S}_{0|i}+\frac{\sqrt{\alpha_i}(1-\bar{\alpha}_{i-1})}{1-\bar{\alpha}_i}\mathbf{S}_i+\tilde{\sigma}_i\boldsymbol{\epsilon}$;
  $\mathbf{S}_{i-1}\leftarrow\mathbf{S}'_{i-1}-\zeta_i\nabla_{\mathbf{S}_i}\log p(\mathbf{Y}|\mathbf{S}_i,\mathbf{H}_i^\star)$;
end for
Calculate $\mathbf{H}_0^\star$ using FCLS based on $\mathbf{S}_0$ and $\mathbf{Y}$;

Stemming from the noise effect in the LMM and the large size of the spectral library, there will be several local optima with non-zero posterior probability $p(\mathbf{S}|\mathbf{Y})$. To filter out the local optima, we adopt the maximum likelihood strategy to select the best results from K times of sampling. This strategy assumes that the global optimum will have the maximum likelihood $p(\mathbf{Y}|\mathbf{S},\mathbf{H}^\star)$, which is equal to the minimum value of $\|\mathbf{Y}-\mathbf{H}_0^\star\mathbf{S}_0^{\mathrm{T}}\|_{\mathrm{F}}^2$. If the posterior probability of the global optimum is $p^\star=\max p(\mathbf{S}|\mathbf{Y})$, the probability of obtaining the global optimum after K times of sampling is $1-(1-p^\star)^K$.

IV. Experiments

In this section, we comprehensively evaluate the unmixing capability of the proposed DiffUn on both synthetic and real data. Five semi-blind unmixing methods are selected as the alternative methods to be compared with our approach, including SUnSAL [19], CSUnL0 [2], ELMM [27], ALMM [28] and SVASU [29]. Among these alternative methods, SUnSAL is a classical semi-blind unmixing method solved by ADMM. CSUnL0 improves the SUnSAL by involving L0 regularization. ELMM, SVASU and ALMM are three state-of-the-art unmixing methods that model the variability of signatures. ELMM introduces pixel-wise scaling factors for all the signatures in the library and involves regularization terms to the estimated endmember matrix based on the reference library. ALMM involves dictionary learning to fit the complex spectral variations during the unmixing. SVASU utilizes a fixed dictionary extracted by PCA as the spectral variability dictionary. All relevant parameters of the compared methods are empirically and experimentally set up to maximize the performance using the grid searching around the reported optimal values. K and i_{start} for DiffUn is set to 5 and 200, respectively. The settings of ζ_i are detailed in Appendix C. s_θ is implemented with a 1D U-Net. The proposed DiffUn is implemented using PyTorch and trained on an NVIDIA GeForce GTX 1080Ti GPU. Other algorithms are running on Intel(R) i5-12400F CPU.

A. Performance Evaluation Criteria

To quantitatively assess the unmixing performance, we incorporate three criteria following [28]: abundance overall root mean square error (aRMSE), reconstructed overall root mean square error (rRMSE), and spectral angle mapper (SAM). Given the ground truth abundance matrix $\mathbf{H}$ and the corresponding estimated abundance matrix $\hat{\mathbf{H}}$, aRMSE can be defined as

$$\text{aRMSE}(\mathbf{H},\hat{\mathbf{H}})=\frac{1}{R}\sum_{r=1}^{R}\sqrt{\frac{1}{N}\sum_{l=1}^{N}(H_{lr}-\hat{H}_{lr})^2} \tag{42}$$

where l represents the l-th pixel and r represents the r-th endmember. N and R represent the number of pixels and endmembers, respectively.

Since the metrics used for directly measuring the extracted endmembers are not suitable and convenient for semi-blind unmixing performance, the quality of the reconstruction HSI can be considered as an alternative to indicating the performance of unmixing [2], [18]–[20], [22], [27]–[29]. The rRMSE is defined as

$$\text{rRMSE}(\mathbf{X},\hat{\mathbf{X}})=\frac{1}{C}\sum_{c=1}^{C}\sqrt{\frac{1}{N}\sum_{l=1}^{N}(X_{lc}-\hat{X}_{lc})^2} \tag{43}$$

where $\mathbf{X}$ is the clean ground-truth HSI, $\hat{\mathbf{X}}$ is the corresponding reconstructed HSI by the algorithm, c represents the c-th band and C represents the number of bands.

The SAM is defined as

$$\text{SAM}(\mathbf{X},\hat{\mathbf{X}})=\frac{1}{N}\sum_{l=1}^{N}\arccos\left(\frac{\mathbf{X}_l^{\mathrm{T}}\hat{\mathbf{X}}_l}{\|\mathbf{X}_l\|_2\|\hat{\mathbf{X}}_l\|_2}\right) \tag{44}$$

It is noted that the reconstruction metrics rRMSE and SAM in some scenarios should be carefully used as they are not able to directly reflect the unmixing performance, especially when used for comparing unmixing algorithms with different characteristics.

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2024.3408475

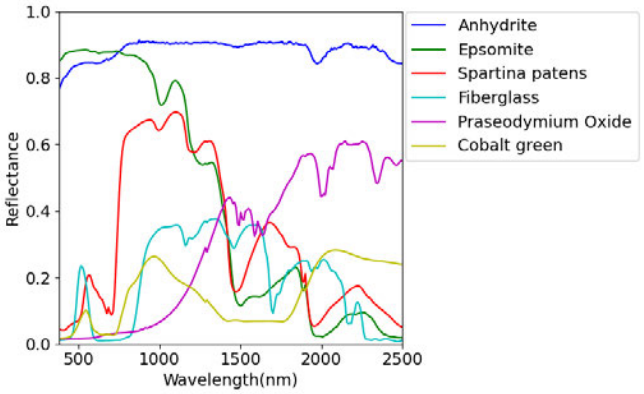


Fig. 2: Selected signatures from USGS library.

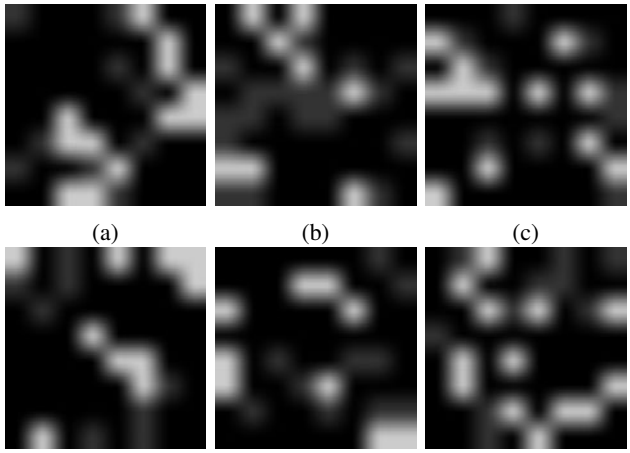


Fig. 3: Synthetic abundance maps for six endmembers: (a) Anhydrite (b) Epsomite, (c) Spartina patens, (d) Fiberglass GDS336, (e) Praseodymium Oxide, (f) Cobalt green.

B. Experiments on Synthetic Data

We utilize USGS[1] spectral library, which covers 224 bands with wavelengths ranging between 0.38 μm and 2.5 μm for the following experiments. The original USGS library contains 2457 signatures. Due to instrument artifacts, low Signal-to-Noise (SNR), atmospheric contamination, and other factors, some signatures have less than 224 available bands [48]. Removing these incomplete data, 1533 remaining signatures construct the spectral library $\mathbf{A}_0$. Six pure signatures (Anhydrite, Epsomite, Spartina patens, Fiberglass GDS336, Praseodymium Oxide, and Cobalt green) are selected from $\mathbf{A}_0$ to generate an endmember matrix $\mathbf{S}_0$ as depicted in Fig. 2. Following [49], we generate abundance matrix with a size of 64×64 for each endmember as follows,

Step 1 Split the 64×64 spatial size into 64 non-overlapping blocks with a size of 8×8.

Step 2 Randomly assign each block to two endmembers with ratios of 0.2 and 0.8, respectively.

Step 3 Adopt a low-pass Gaussian filter with a standard deviation of 2 to smooth the abundance matrix and increase the mixing degree.

Step 4 Normalize the abundance matrix to meet the ANC and ASC.

Synthetic abundance maps of six endmembers are shown in Fig. 3. After that, the synthetic clean ground-truth HSI $\mathbf{X}$ is derived from LMM, and the observed HSI is degraded by zero-mean additive Gaussian noise.

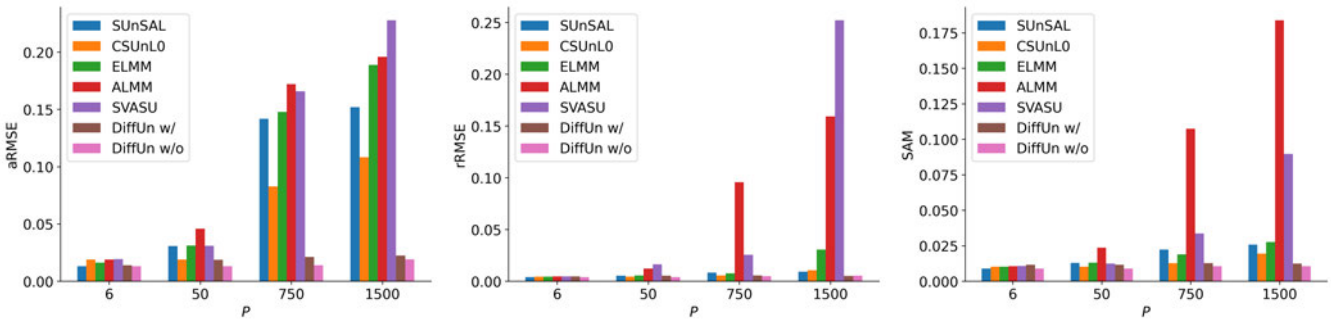
1) Robustness to Spectral Library Size: In this experiment, we compare DiffUn with learning and DiffUn without learning, referred to as "DiffUn w/" and "DiffUn w/o" respectively, with alternative methods to investigate their robustness to spectral library size P. We construct 4 sub-libraries from the USGS library, which are denoted as $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{A}_3$ and $\mathbf{A}_4$. All of these 4 sub-library includes the underlying 6 endmembers. The size of $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{A}_3$ and $\mathbf{A}_4$ are 6, 50, 750, and 1500, respectively, where the relationship holds that $\mathbf{S}_0 = \mathbf{A}_1 \subset \mathbf{A}_2 \subset \mathbf{A}_3 \subset \mathbf{A}_4 \subset \mathbf{A}_0$. The neural network s_θ in DiffUn w/ is trained on the given sub-library. For alternative approaches, different regularization weights are employed for distinct sub-libraries. These weight values are determined by minimizing the aRMSE. In this experiment, the variance $\sigma_{\mathbf{Y}}^2$ of the Gaussian used to corrupt the clean ground-truth HSI $\mathbf{X}$ is set to 10^{-3}.

Table I presents the unmixing performances of all the methods. As shown, SUnSAL and DiffUn w/o simultaneously achieve optimal unmixing performance when P is 6 as the sub-library $\mathbf{A}_1$ only consists of the underlying endmembers. Conversely, methods like CSUnL0, ELMM, ALMM and SVASU result in lower performance. When P is increased to 50, the unmixing performance of CSUnL0 remains constant due to the incorporation of L0 regularization, which enforces strong sparsity constraints on the abundance matrix. Compared to CSUnL0, methods such as SUnSAL and SVASU, which use L1 regularization, have lower performance in all metrics. ELMM offers inferior performance compared to SUnSAL as it does not impose a sparsity constraint on the abundance matrix. DiffUn w/ and DiffUn w/o outperform other methods concerning aRMSE. When P increases from 750 to 1500, methods such as SUnSAL, ELMM, ALMM, and SVASU experience a severe deterioration of unmixing performance due to inadequate ability to handle large-scale spectral libraries. Additionally, non-convexity stemming from the additional terms for spectral variations leads to a more significant decline in methods like ELMM, ALMM and SVASU. The non-convexity stemming from the L0 regularization results in a decline in the performance of CSUnL0. DiffUn w/ and DiffUn w/o outperform CSUnL0 by formulating hyperspectral unmixing as a generation task rather than a sparse regression problem. It can be found that DiffUn w/o outperforms DiffUn w/ in most cases of this experiment, which is attributed to the fact that DiffUn w/ is trained to learn a continuous spectral prior distribution, whereas DiffUn w/o estimates that distribution directly by a discrete multi-point distribution supported by the library. In this experiment, since there is no spectral variation and the underlying real endmembers are precisely matched to a subset of the spectral library, the multi-point distribution

[1]https://www.usgs.gov/labs/spec-lab/capabilities/spectral-library

TABLE I: Comparisons of unmixing methods with respect to the spectral library size P. The best and second best values are highlighted.

spectral library	Index	SUnSAL	CSUnL0	ELMM	ALMM	SVASU	DiffUn w/	DiffUn w/o
$\mathbf{A}_1$, $P=6$	aRMSE	0.0130	0.0189	0.0162	0.0188	0.0194	0.0140	0.0130
	rRMSE	0.0040	0.0045	0.0045	0.0046	0.0046	0.0047	0.0040
	SAM	0.0089	0.0103	0.0102	0.0105	0.0107	0.0115	0.0089
$\mathbf{A}_2$, $P=50$	aRMSE	0.0307	0.0189	0.0311	0.0458	0.0310	0.0187	0.0130
	rRMSE	0.0053	0.0045	0.0055	0.0122	0.0163	0.0052	0.0040
	SAM	0.0129	0.0103	0.0130	0.0236	0.0124	0.0116	0.0089
$\mathbf{A}_3$, $P=750$	aRMSE	0.1417	0.0828	0.1479	0.1350	0.1659	0.0212	0.0140
	rRMSE	0.0083	0.0056	0.0076	0.0460	0.0254	0.0055	0.0048
	SAM	0.0222	0.0127	0.0189	0.0919	0.0337	0.0127	0.0105
$\mathbf{A}_4$, $P=1500$	aRMSE	0.1520	0.1083	0.1889	0.1959	0.2278	0.0224	0.0190
	rRMSE	0.0093	0.0107	0.0306	0.1591	0.2521	0.0051	0.0054
	SAM	0.0258	0.0195	0.0277	0.1838	0.0897	0.0124	0.0106



(a) (b) (c)

Fig. 4: Comparisons of unmixing methods with respect to the spectral library size P. (a) aRMSE, (b) rRMSE, (c) SAM.

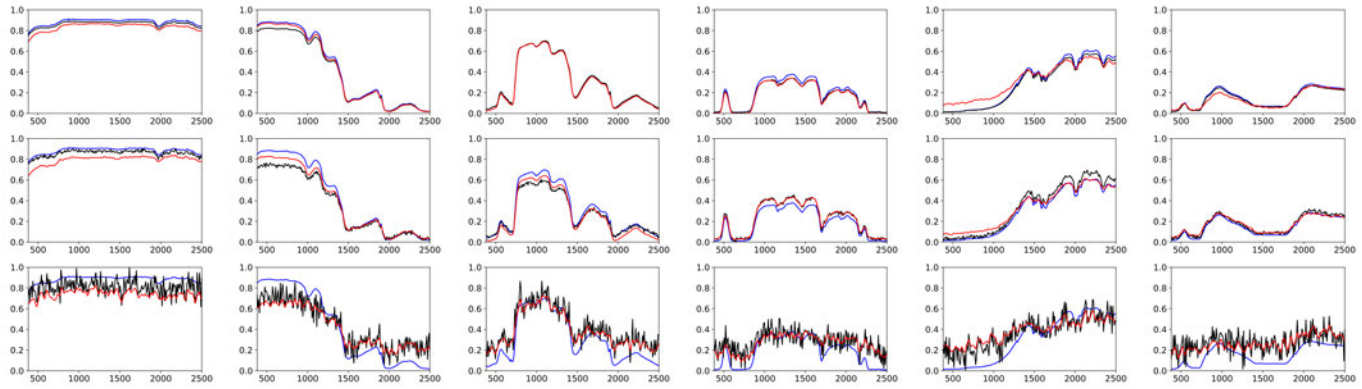
can fully cover the underlying endmembers, resulting in better unmixing performance.

To clearly compare the unmixing performance concerning the size of the spectral library, the aRMSE, rRMSE and SAM of different methods are plotted in Fig. 4. The performance of SUnSAL, CSUnL0 and ELMM exhibit stable performance in terms of rRMSE and SAM, but they are sensitive to P in terms of aRMSE. Overall, it is apparent that DiffUn w/ and DiffUn w/o demonstrate the most exceptional robustness concerning P in all metrics.

2) Robustness to Spectral Variability: We are comparing DiffUn w/ and DiffUn w/o with other methods to assess their robustness to spectral variability under different conditions. In practice, three types of variability commonly result in signature variations. The first is related to variations in illumination, and the second accounts for other complex environmental conditions and instrumental configurations [28]. The last type arises due to the differing characteristics of the spectrometers or materials between the spectral library and the observed HSI (for instance, variations between spectrometer BECK and ASDFR or between Fiberglass GDS336 and GDS337). Aiming to ensure the unmixing performance of all methods, sub-library $\mathbf{A}_2$ is utilized as the spectral library.

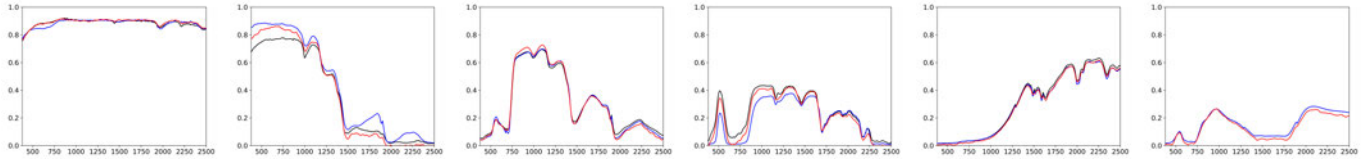
Initially, we undertake the experiment with the first two types of variability to verify the robustness of all algorithms. To be specific, we multiply the six original signatures in the synthetic HSI by randomly generated scaling factors ranging from $[1-c, 1+c]$ and then add Gaussian noise with variance σ^2 to these scaled signatures. The resulting signatures are used to construct the underlying endmember matrix for all the pixels. Consequently, the clean HSI is generated by the LMM and the observed HSI is corrupted by Gaussian noise with variance $\sigma^2_{\mathbf{Y}} = 10^{-3}$. It should be noticed that the Gaussian noise with variance σ^2 is injected into the endmembers, which serves as the spectral variations. On the other hand, the Gaussian noise with variance $\sigma^2_{\mathbf{Y}}$ is used to corrupt the clean HSI, which serves as the observation noise. The aRMSE, rRMSE and SAM of all the methods are listed in Table II. The results manifest that when the variations are small (e.g., $c = 0.1$ and $\sigma^2 = 10^{-3}$), CSUnL0 produces satisfactory unmixing performance by removing the ASC constraint. However, when the variations are pronounced, the performance of CSUnL0 worsens because it fails to model the spectral variations. ELMM surpasses SUnSAL due to its incorporation of scaling factors and endmember matrix modeling for each pixel. Nevertheless, the high complexity caused by numerous additional parameters renders its performance inferior to SVASU that utilizes a fixed variability dictionary to fit the spectral variations. The underperformance of ALMM is evident when the variations are significant, but ALMM still exhibits a certain degree of robustness compared to methods such as ELMM. Generally, DiffUn w/ outperforms other methods in most cases owing to the involvement of the unconditional DM training on the spectral library. The spectral prior distribution is smoothed and interpolated through the noising and denoising processes of DM, thus allowing DiffUn w/ to generate signatures distinct from the library. To further analyze the performance of DiffUn w/, we give the underlying signatures and those estimated by

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2024.3408475



(a) (b) (c) (d) (e) (f)

Fig. 5: Comparisons of the underlying real signatures and those estimated by DiffUn w/ with respect to cases of $(0.1, 10^{-3})$, $(0.2, 10^{-2})$, and $(0.3, 10^{-1})$ for the values of (c, σ^2), listed top to bottom. The underlying real signatures are depicted by black lines, the corresponding signatures in the spectral library $\mathbf{A}_2$ are indicated by blue lines, and the estimated endmembers are presented by the red lines. (a) Anhydrite (b) Epsomite, (c) Spartina patens, (d) Fiberglass GDS336, (e) Praseodymium Oxide (BECK), (f) Cobalt green.



(a) (b) (c) (d) (e) (f)

Fig. 6: Comparisons of the underlying real signatures in the signature set $\mathbf{A}_0 - \mathbf{A}_2$, the most similar signatures in $\mathbf{A}_2$, and those estimated by DiffUn w/. The underlying real endmembers are depicted by black lines, the corresponding signatures in the spectral library $\mathbf{A}_2$ by blue lines, and the estimated endmembers by red lines. The underlying real signatures are: (a) Albite (b) Ulexite, (c) Marsh, (d) Fiberglass GDS337, (e) Praseodymium Oxide (ASDFR), (f) Cobalt green.

DiffUn w/ in Fig. 5. Compared with the reference signatures in the library, those estimated by DiffUn w/ are more similar to the underlying signatures.

Subsequently, we compare the DiffUn w/ and DiffUn w/o with alternative methods under the third type of variability. Specifically, we substitute the endmembers (except Cobalt green as there is no signature similar to it in the USGS library) with the most similar one in the signature set $\mathbf{A}_0 - \mathbf{A}_2$ to construct the endmember matrix. The used spectral library is still $\mathbf{A}_2$ for all the algorithms. The similarity between any two signatures is measured by spectral angle distance (SAD). The selected signatures are Albite, Ulexite, Marsh, Fiberglass GDS337, Praseodymium Oxide (using the ASDFR spectrometer) and Cobalt green. These selected signatures are depicted with black lines in Fig. 6 and the corresponding most similar signatures in $\mathbf{A}_2$ are illustrated with blue lines in Fig. 6. The unmixing performances of all the methods are illustrated in Table III. From the results, it can be seen that CSUnL0 surpasses SUnSAL as L0 regularization prevents excessive involvement of signatures in the unmixing. The performances of ELMM and ALMM are comparatively poor, mainly caused by the overfitting of numerous additional parameters. Leveraging a fixed variations dictionary extracted by PCA, SVASU achieves the second-best unmixing accuracy in terms of aRMSE. Obviously, the proposed DiffUn w/ stands out among all the methods since DM is more capable of modeling the spectral variations. The signatures estimated by DiffUn w/ are illustrated in Fig. 6 with red lines. As demonstrated in Fig. 6, it is evident that the estimated signatures using DiffUn w/ are closer to the underlying real signatures, in contrast to those present in the library.

3) Robustness to Noise: To investigate the robustness of the proposed methods to noise level, we evaluate the unmixing performance under various signal-to-noise ratios (SNRs). For a fair comparison, $\mathbf{A}_2$ is used as the spectral library for all the methods, and no spectral variations are considered in this experiment. The clean HSI $\mathbf{X}$ is generated by the LMM, and the observed HSI is corrupted by Gaussian noise with the SNR ranging from 10 dB to 40 dB. The aRMSE of all the methods with different SNRs is plotted in Fig. 7. As can be seen, the results in terms of aRMSE indicate as the SNR decreases, so does the unmixing performance for all methods. When the SNR is between 35 dB and 40 dB, CSUnL0 outperforms DiffUn w/ but is inferior to DiffUn w/o. As the SNR drops to 30 dB, DiffUn w/ surpasses CSUnL0 and shows better robustness to the noise. Overall, DiffUn demonstrates the best

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2024.3408475

TABLE II: Comparisons of unmixing methods under different scaling and Gaussian spectral variabilities. The best and second best values are highlighted.

(c, σ^2)	Index	SUnSAL	CSUnL0	ELMM	ALMM	SVASU	DiffUn w/	DiffUn w/o
$(0.1, 10^{-3})$	aRMSE	0.0369	0.0307	0.0345	0.0478	0.0330	0.0259	0.0264
	rRMSE	0.0057	0.0059	0.0120	0.0056	0.0201	0.0060	0.0062
	SAM	0.0149	0.0149	0.0145	0.0249	0.0131	0.0160	0.0154
$(0.2, 10^{-2})$	aRMSE	0.0443	0.0431	0.0425	0.0492	0.0382	0.0307	0.0387
	rRMSE	0.0087	0.0084	0.0084	0.0124	0.0240	0.0085	0.0089
	SAM	0.0236	0.0227	0.0229	0.0337	0.0220	0.0228	0.0235
$(0.3, 10^{-1})$	aRMSE	0.2170	0.1545	0.1453	0.1366	0.0981	0.0534	0.1939
	rRMSE	0.0479	0.0477	0.0455	0.0504	0.0596	0.0384	0.0499
	SAM	0.1194	0.1187	0.1102	0.1260	0.1131	0.0949	0.1217

TABLE III: Comparisons of unmixing methods under variability caused by the different spectrometers and similar materials. The best and second best values are highlighted.

Index	SUnSAL	CSUnL0	ELMM	ALMM	SVASU	DiffUn w/	DiffUn w/o
aRMSE	0.0540	0.0528	0.0564	0.0639	0.0508	0.0266	0.0594
rRMSE	0.0103	0.0101	0.0097	0.0115	0.0117	0.0059	0.0221
SAM	0.0272	0.0261	0.0263	0.0292	0.0262	0.0146	0.0500

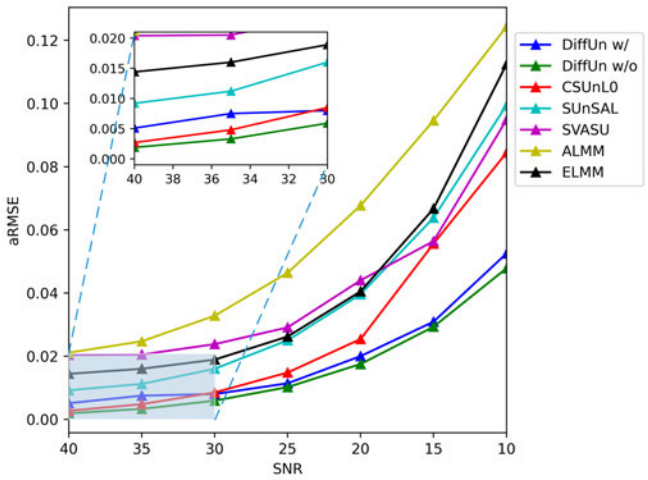


Fig. 7: The results in terms of aRMSE with different SNRs.

performance across different SNRs.

C. Experiments on Real Data

We further carried out experiments on two widely used real-world HSIs, i.e., Jasper Ridge[1] and Samson[2].

1) Jasper Ridge Dataset: It is a popular hyperspectral dataset with spatial size of 512×614. Each pixel is recorded at 224 channels ranging from 380 nm to 2500 nm with a spectral resolution of 9.46 nm. A region with 100×100 pixels is considered in this experiment. The channels 1-3, 108-112, 154-166, and 220-224 are removed owing to the dense water vapor and atmospheric effects. The retained 198 channels are used for unmixing. Fig. 8a presents the false color image of Jasper Ridge HSI. Utilizing HySime [50] and manual analysis, four kinds of materials are determined in this scenario, i.e., tree, water, soil, and road. Since the ground-truth endmembers and abundances are unavailable, four pixels that are most likely to be pure are manually selected and consist of the reference endmembers matrix. Then we use the FCLS to estimate the corresponding reference abundance maps. The first employed library with $P = 4$ is constructed by VCA, referring to [27], [28]. Then we utilize strategy in [51] to build a larger library. Specifically, we randomly split the HSI into 10 subsets without overlapping and then perform VCA to extract 4 signatures on each subset. The resulting endmembers are merged to construct a spectral library with $P = 40$.

Table IV shows the unmixing results of all the methods applied to the Jasper Ridge dataset. From the data reconstruction perspective, ALMM and ELMM demonstrate impressive performance since they consider reconstruction error as the object function and involve numerous additional parameters to fit the reconstruction error. However, it should be noted that the reconstruction error is not the main target of hyperspectral unmixing, and should be considered carefully [52], [53]. The ALMM involves an additional dictionary to further refine the residual error, which will reasonably lead to lower residual error and higher reconstruction accuracy. However, the little residual error does not always signify good estimation of endmembers and abundances. Sometimes even though the ALMM does not estimate the abundances and endmembers accurately, it is still capable of precisely reconstructing the observed HSI using the additional dictionary. From the perspective of accuracy of abundance maps, SVASU and DiffUn w/o exhibit outstanding unmixing performance when $P = 4$. DiffUn w/ and DiffUn w/o outperform other methods when $P = 40$. These results verify the robustness of DiffUn w/ and DiffUn w/o to spectral library size. Fig. 9 illustrates the reference signatures and those estimated by DiffUn w/. It is evident that the estimated signatures of DiffUn w/ bear resemblance to the reference signatures. Fig. 10 demonstrates the reference and estimated abundance maps with library size $P = 4$. As can be seen, all the algorithms show similar performance in estimating the abundance maps of trees and dirt. However, methods such as CSUnL0, ELMM, ALMM, and SVASU fall short of estimating the abundance maps of water and roads.

[1]https://rslab.ut.ac.ir/data
[2]https://www.opticks.org/

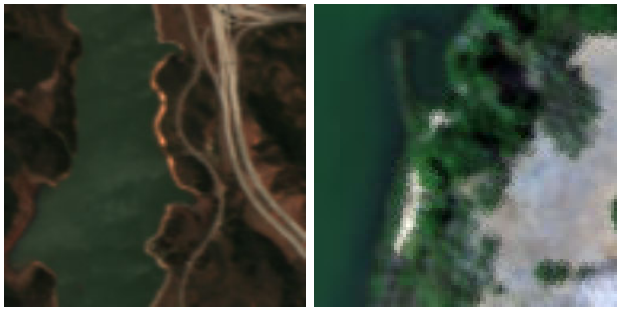
This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2024.3408475

TABLE IV: Comparisons of unmixing methods on Jasper Ridge dataset. The best and second best values are highlighted.

Spectral library	Index	SUnSAL	CSUnL0	ELMM	ALMM	SVASU	DiffUn w/	DiffUn w/o
$P = 4$	aRMSE	0.1001	0.1380	0.1094	0.1177	0.0887	0.1151	0.1001
	rRMSE	0.0425	0.0180	0.0043	0.0007	0.0227	0.0261	0.0425
	SAM	0.0951	0.0837	0.0144	0.0052	0.0954	0.0792	0.0951
$P = 40$	aRMSE	0.2048	0.1370	0.3594	0.3543	0.3549	0.1113	0.1040
	rRMSE	0.0237	0.0106	0.0025	0.0057	0.0348	0.0187	0.0247
	SAM	0.0644	0.0619	0.0174	0.0011	0.0630	0.0702	0.0832

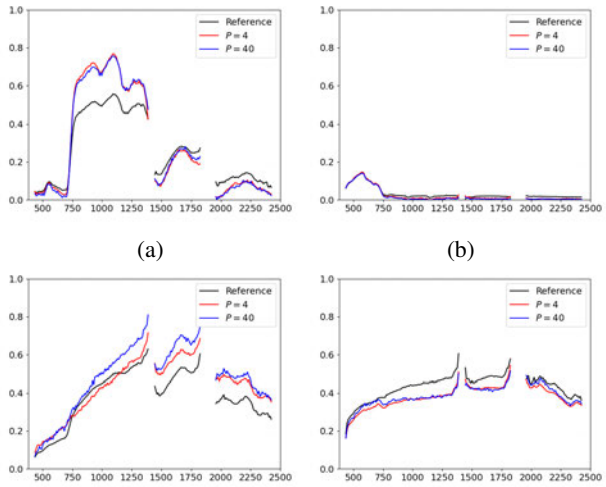
TABLE V: Comparisons of unmixing methods on the Samson dataset. The best and second best values are highlighted.

Spectral library	Index	SUnSAL	CSUnL0	ELMM	ALMM	SVASU	DiffUn w/	DiffUn w/o
$P = 3$	aRMSE	0.0763	0.1736	0.1816	0.1645	0.2142	0.1000	0.0763
	rRMSE	0.0192	0.0053	0.0151	0.0004	0.0114	0.0129	0.0192
	SAM	0.0576	0.0342	0.0788	0.0044	0.0504	0.0597	0.0576
$P = 30$	aRMSE	0.5077	0.3320	0.4411	0.4939	0.2763	0.0940	0.1012
	rRMSE	0.0100	0.0138	0.0162	0.0004	0.0083	0.0149	0.0137
	SAM	0.0509	0.2033	0.0504	0.0041	0.0495	0.0563	0.0594



(a) (b)

Fig. 8: False color images of (a) Jasper Ridge dataset and (b) Samson dataset.



(a) (b) (c) (d)

Fig. 9: Estimated and reference signatures of Jasper Ridge: (a) Tree, (b) Water, (c) Dirt, and (d) Road.

Fig. 11 illustrates the reference and estimated abundance maps with library size $P = 40$. Algorithms such as SUnSAL, ELMM, ALMM, and SVASU exhibit undesirable unmixing performance even though they have high reconstruction accuracy. Overall, the estimated abundance maps of DiffUn w/ and DiffUn w/o are more similar to the reference abundance maps.

2) Samson Dataset: It is another real-world hyperspectral dataset with spatial size of 952×952. To reduce the computational complexity and ease of analysis, a region of 95×95 pixels is considered in this experiment, which starts from the (252,332)-th pixel in the original HSI. The false color image of the selected region is presented in Fig. 8b. The spectral resolution is 3.13 nm with 156 channels covering the wavelengths from 401 nm to 889 nm. Three materials are manually selected from the HSI, i.e., soil, tree, and water. Likewise, the reference endmembers matrix is constructed by manually selecting three pure pixels and the corresponding reference abundance maps are estimated by FCLS. The first library with $P = 3$ comprises 3 signatures identified by VCA. Subsequently, we partition the HSI into 10 non-overlapping subsets and utilize VCA to derive 3 signatures from each subset, making the second library with $P = 30$ consist of 30 spectral signatures.

Table V presents the results of all the algorithms on the Samson dataset. It can be seen that ALMM manifests the best reconstruction result in terms of rRMSE and SAM. However, as mentioned above, the reconstruction error is not the primary goal of hyperspectral unmixing. The ALMM overfits the reconstruction error, which leads to an inaccurate estimation of the abundances, but still achieves a high reconstruction accuracy. In terms of aRMSE, the performances of ALMM, CSUnL0, ELMM, and SVASU are unsatisfactory whenever the value of P is 3 or 30. On the contrary, DiffUn w/ and DiffUn w/o exhibit superior unmixing performance in terms of aRMSE. The estimated abundance maps with $P = 3$ and $P = 30$ are illustrated in Fig. 12 and Fig. 13, respectively. When $P = 3$, all the methods except ELMM and ALMM show high estimation accuracy on the trees. However, methods such as CSUnL0 and SVASU fall short of estimating the abundance maps of soil and water. When $P = 30$, most algorithms fail to estimate the abundance maps of soil and water. Overall, the estimated abundance maps of DiffUn w/ and DiffUn w/o are more similar to the reference abundance maps. Fig. 14 illustrates the estimated and reference signatures. It is apparent

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2024.3408475

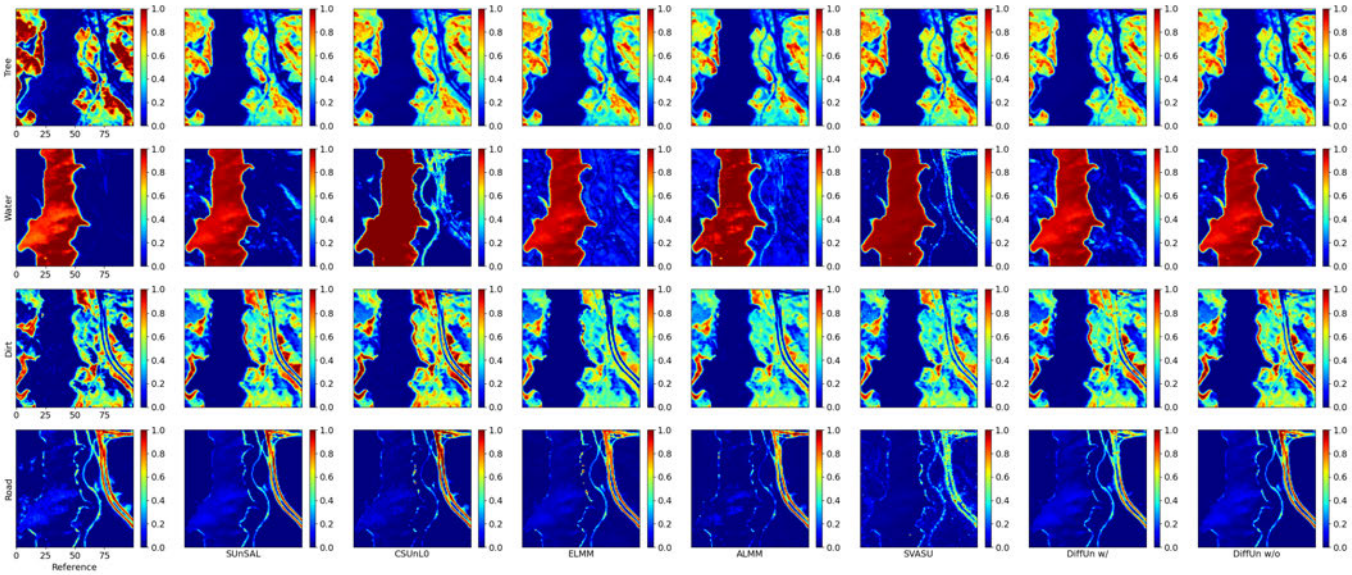


Fig. 10: Estimated and reference abundance maps of Tree, Water, Dirt, and Road of Jasper Ridge with library size $P = 4$.

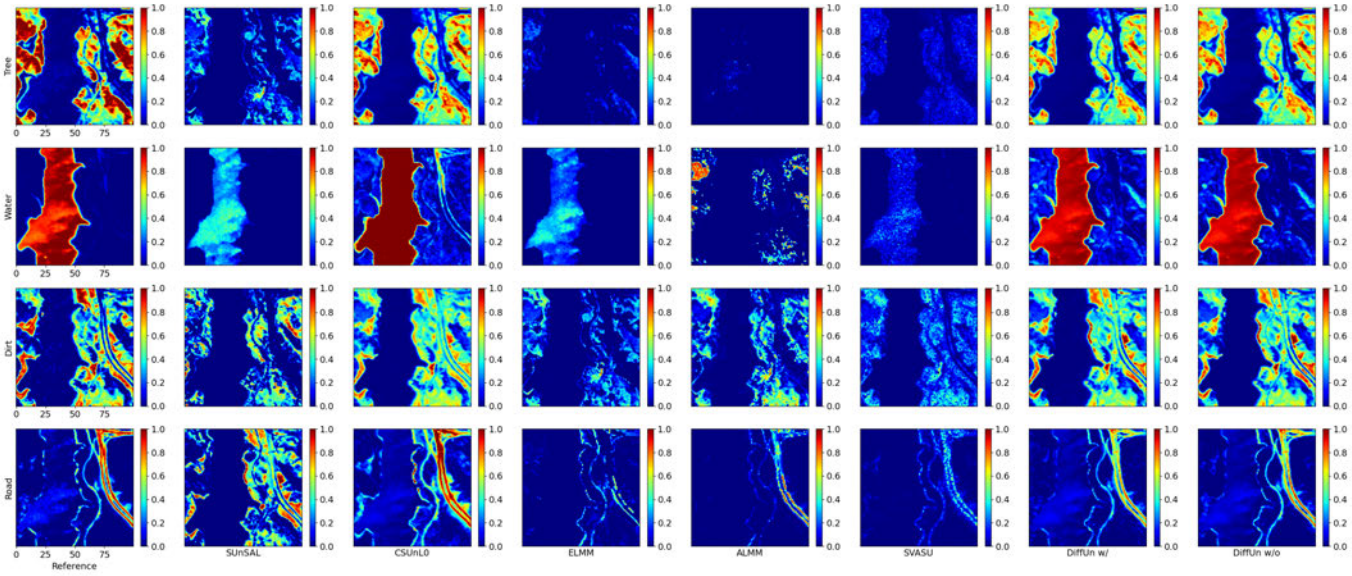


Fig. 11: Estimated and reference abundance maps of Tree, Water, Dirt, and Road of Jasper Ridge with library size $P = 40$.

that estimated signatures of DiffUn w/ exhibit high similarity to the reference signatures.

V. Conclusion

This article proposes a novel and effective diffusion-model-based hyperspectral spectral unmixing method named DiffUn. DiffUn addresses the challenges of dealing with large-size spectral libraries and handling variabilities between the underlying signatures and those in the spectral library. The method solves the hyperspectral unmixing problem by sampling from a posterior distribution, leveraging a reverse process. This process combines the spectral prior distribution, which is learned by an unconditional diffusion model, and the likelihood, which is derived from the linear spectral mixture model, to facilitate the unmixing performance. Experimental results on both synthetic and real-world HSIs demonstrate the superiority of DiffUn over several state-of-the-art methods.

Appendix A
SDE of DiffUn from the Discrete-Time Equation

We can convert the Eq. (25) to the following equation using the property of logarithm,

$$p(\mathbf{S}_{t-\Delta t}|\mathbf{S}_t, \mathbf{Y}) = p(\mathbf{S}_t|\mathbf{S}_{t-\Delta t}) e^{\log p(\mathbf{S}_{t-\Delta t}|\mathbf{Y}, \mathbf{H}_t^\star) - \log p(\mathbf{S}_t|\mathbf{Y}, \mathbf{H}_t^\star)} \quad (45)$$

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2024.3408475

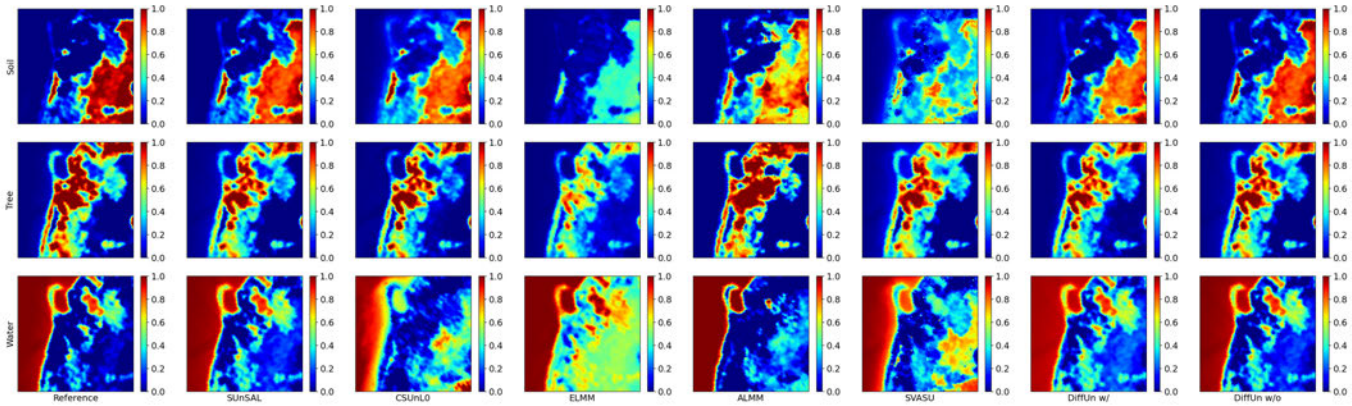


Fig. 12: Estimated and reference abundance maps of Soil, Tree Water of Samson with library size $P = 3$.

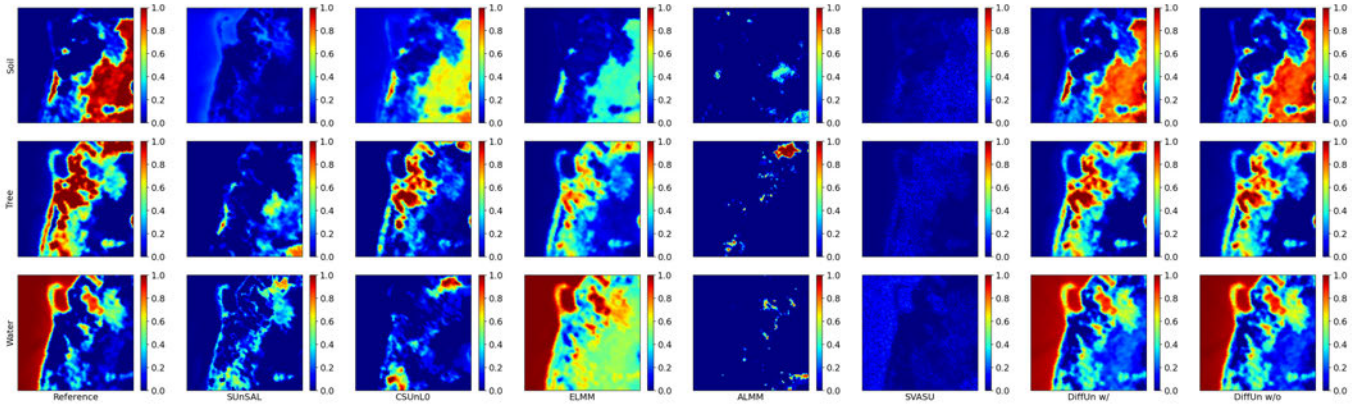
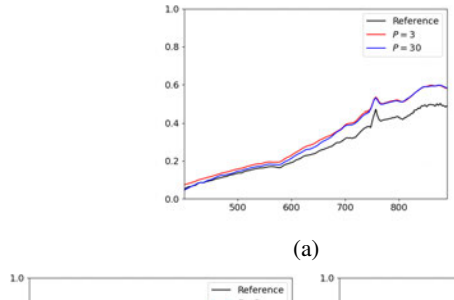
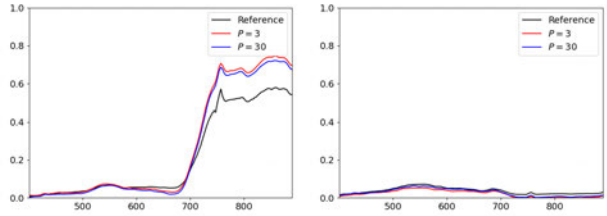


Fig. 13: Estimated and reference abundance maps of Soil, Tree Water of Samson with library size $P = 30$.



(a)



(b) (c)

Fig. 14: Estimated and reference signatures of Samson: (a) Soil, (b) Tree and (c) Water.

According to Taylor expansion, we have,

$$\log p(\mathbf{S}_t|\mathbf{Y},\mathbf{H}_t^\star) - \log p(\mathbf{S}_{t-\Delta t}|\mathbf{Y},\mathbf{H}_t^\star) = - \nabla_{\mathbf{S}_t} \log p(\mathbf{S}_t|\mathbf{Y},\mathbf{H}_t^\star)(\mathbf{S}_t - \mathbf{S}_{t-\Delta t}) + O(\Delta t) \quad (46)$$

By substituting Eq. (16) and Eq. (46) into Eq. (45), we get,

$$p(\mathbf{S}_{t-\Delta t}|\mathbf{S}_t,\mathbf{Y}) \propto e^{-\frac{\|\mathbf{S}_{t-\Delta t}-\mathbf{S}_t-(\frac{\beta_t}{2}\mathbf{S}_t+\beta_t\nabla_{\mathbf{S}_t}\log p(\mathbf{S}_t|\mathbf{Y},\mathbf{H}_t^\star))\Delta t\|_{\mathrm{F}}^2}{2\beta_t\Delta t}+O(\Delta t)} \quad (47)$$

which can be rewritten as follows,

$$\mathbf{S}_{t-\Delta t} = \mathbf{S}_t + \left(\frac{\beta_t}{2}\mathbf{S}_t + \beta_t\nabla_{\mathbf{S}_t}\log p(\mathbf{S}_t|\mathbf{Y},\mathbf{H}_t^\star)\right)dt - \sqrt{\beta_t}d\mathbf{W} \quad (48)$$

Therefore, while $\Delta t \to 0$, meaning $O(\Delta t) \to 0$, we eventually derive the reverse process of the DiffUn from Eq. (48) as follows,

$$d\mathbf{S}_t = -\left(\frac{\beta_t}{2}\mathbf{S}_t + \beta_t\nabla_{\mathbf{S}_t}\log p(\mathbf{S}_t|\mathbf{Y},\mathbf{H}_t^\star)\right)dt + \sqrt{\beta_t}d\mathbf{W} \quad (49)$$

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2024.3408475

APPENDIX B
ERROR ANALYSIS OF THE APPROXIMATION OF J

Define that $p(\mathbf{s}_t|\mathbf{s}) \sim \mathcal{N}(\mathbf{s}; \sqrt{\bar{\alpha}_t}\mathbf{s}, (1-\bar{\alpha}_t)\mathbf{I})$, we can derive the best estimation of $\mathbf{s}$ by,

$$\mathbf{r}(\mathbf{s}_t) = \frac{\int \mathbf{s} p(\mathbf{s}_t|\mathbf{s}) p(\mathbf{s}) d\mathbf{s}}{p(\mathbf{s}_t)} \tag{50}$$

Applying the gradient operator, we have,

$$\nabla_{\mathbf{s}}\mathbf{r}(\mathbf{s}_t) = \frac{\int \mathbf{s}\nabla_{\mathbf{s}_t} p(\mathbf{s}_t|\mathbf{s})^{\mathrm{T}} p(\mathbf{s}) d\mathbf{s}}{p(\mathbf{s}_t)} - \frac{\int \mathbf{s} p(\mathbf{s}_t|\mathbf{s}) p(\mathbf{s}) d\mathbf{s} \nabla_{\mathbf{s}_t} p(\mathbf{s}_t)^{\mathrm{T}}}{p(\mathbf{s}_t)^2} \tag{51}$$

According to the definition of $p(\mathbf{s}_t|\mathbf{s})$, we have,

$$\nabla_{\mathbf{s}_t} p(\mathbf{s}_t|\mathbf{s}) = -\frac{1}{1-\bar{\alpha}_t}\left(\mathbf{s}_t - \sqrt{\bar{\alpha}_t}\mathbf{s}\right) p(\mathbf{s}_t|\mathbf{s}) \tag{52}$$

and,

$$\nabla_{\mathbf{s}_t} p(\mathbf{s}_t) = \nabla_{\mathbf{s}} \int p(\mathbf{s}_t|\mathbf{s}) p(\mathbf{s}) d\mathbf{s} \tag{53}$$

$$= \frac{1}{1-\bar{\alpha}_t} \int (\sqrt{\bar{\alpha}_t}\mathbf{s} - \mathbf{s}_t) p(\mathbf{s}_t|\mathbf{s}) p(\mathbf{s}) d\mathbf{s} \tag{54}$$

$$= \frac{1}{1-\bar{\alpha}_t}\left(\sqrt{\bar{\alpha}_t}\int \mathbf{s} p(\mathbf{s}_t|\mathbf{s}) p(\mathbf{s}) d\mathbf{s} - \mathbf{s}_t p(\mathbf{s}_t)\right) \tag{55}$$

By substituting Eq. (52) and Eq. (55) into Eq. (51), we can derive,

$$\nabla_{\mathbf{s}_t}\mathbf{r}(\mathbf{s}_t) = \frac{\sqrt{\bar{\alpha}_t}}{1-\bar{\alpha}_t}\left(\frac{\int \mathbf{s}\mathbf{s}^{\mathrm{T}} p(\mathbf{s}_t|\mathbf{s})^{\mathrm{T}} p(\mathbf{s}) d\mathbf{s}}{p(\mathbf{s}_t)} - \frac{\int \mathbf{s} p(\mathbf{s}_t|\mathbf{s}) p(\mathbf{s}) d\mathbf{s} \int \mathbf{s}^{\mathrm{T}} p(\mathbf{s}_t|\mathbf{s}) p(\mathbf{s}) d\mathbf{s}}{p(\mathbf{s}_t)^2}\right) \tag{56}$$

$$= \sqrt{\bar{\alpha}_t}\frac{\int \mathbf{s}\mathbf{s}^{\mathrm{T}} p(\mathbf{s}|\mathbf{s}_t) d\mathbf{s} - \int \mathbf{s} p(\mathbf{s}|\mathbf{s}_t) d\mathbf{s} \int \mathbf{s}^{\mathrm{T}} p(\mathbf{s}|\mathbf{s}_t) d\mathbf{s}}{1-\bar{\alpha}_t} \tag{57}$$

$$= \sqrt{\bar{\alpha}_t}\frac{\mathbb{E}_{\mathbf{s}\sim p(\mathbf{s}|\mathbf{s}_t)}\left[\mathbf{s}\mathbf{s}^{\mathrm{T}}\right] - \mathbb{E}_{\mathbf{s}\sim p(\mathbf{s}|\mathbf{s}_t)}\mathbf{s}\mathbb{E}_{\mathbf{s}\sim p(\mathbf{s}|\mathbf{s}_t)}\mathbf{s}^{\mathrm{T}}}{1-\bar{\alpha}_t} \tag{58}$$

$$= \frac{\sqrt{\bar{\alpha}_t}}{1-\bar{\alpha}_t}\mathrm{cov}_{\mathbf{s}\sim p(\mathbf{s}|\mathbf{s}_t)}[\mathbf{s}] \tag{59}$$

According to Eq. (59), we can measure the upper bound of difference between $\nabla_{\mathbf{s}}\mathbf{r}(\mathbf{s}_t)$ and $\frac{\gamma\sqrt{\bar{\alpha}_t}}{1-\bar{\alpha}_t}\mathbf{I}$ under the L2 norm as follows,

$$L(\gamma) = \sup_{\|\mathbf{x}\|_2=1} \| \left(\nabla_{\mathbf{s}}\mathbf{r}(\mathbf{s}_t) - \frac{\gamma\sqrt{\bar{\alpha}_t}}{1-\bar{\alpha}_t}\mathbf{I}\right)\mathbf{x}\|_2^2 \tag{60}$$

$$= \frac{\bar{\alpha}_t}{\sigma^4}\sup_{\|\mathbf{x}\|_2=1} \|(\mathrm{cov}_{\mathbf{s}\sim p(\mathbf{s}|\mathbf{s})}[\mathbf{s}] - \gamma\mathbf{I})\mathbf{x}\|_2^2 \tag{61}$$

$$= \frac{\bar{\alpha}_t}{\sigma^4}\max_i |\lambda_i - \gamma|^2 \tag{62}$$

$$\geq \frac{\bar{\alpha}_t}{\sigma^4}\frac{|\max\lambda_i - \min\lambda_i|^2}{4} \tag{63}$$

where the last inequality hold when $\gamma = \gamma^\star \triangleq \frac{|\max\lambda_i + \min\lambda_i|}{2}$, and λ_i is the eigenvalue of $\mathrm{cov}_{\mathbf{s}\sim p(\mathbf{s}|\mathbf{s})}[\mathbf{s}]$. Therefore, we can derive the upper bound of the difference between $\nabla_{\mathbf{s}}\mathbf{r}(\mathbf{s}_t)$ and $\frac{\gamma^\star\sqrt{\bar{\alpha}_t}}{1-\bar{\alpha}_t}\mathbf{I}$ using the L2 norm of the matrix as follows,

$$\| \left(\nabla_{\mathbf{s}}\mathbf{r}(\mathbf{s}_t) - \frac{\gamma^\star\sqrt{\bar{\alpha}_t}}{1-\bar{\alpha}_t}\mathbf{I}\right)\mathbf{x}\|_2^2 \leq \frac{\bar{\alpha}_t}{\sigma^4}\frac{|\max\lambda_i - \min\lambda_i|^2}{4}\|\mathbf{x}\|_2^2 \tag{64}$$

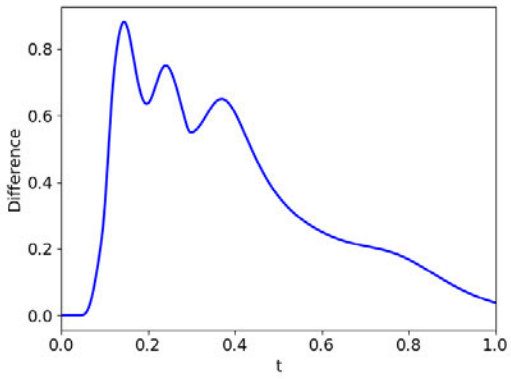


Fig. 15: The difference between $\mathbf{J}_t$ and proximate matrix.

The difference between $\nabla_{\mathbf{s}}\mathbf{r}$ and the proximate matrix $\frac{\gamma^\star\sqrt{\bar{\alpha}_t}}{1-\bar{\alpha}_t}\mathbf{I}$ is illustrated in Fig. 15. The result indicates that when t is close to 0 or 1, the difference caused by the approximation is small and bounded. Therefore, the approximation is reasonable.

APPENDIX C
SETTING OF ζ_i

From the perspective of maximum likelihood estimation, we can derive

$$\zeta_i = \arg\max_{\zeta} \log p(\mathbf{Y}|\mathbf{H}_i^\star, \mathbf{S}_i, \zeta) \tag{65}$$

$$\triangleq \arg\min_{\zeta} \|\mathbf{Y} - \mathbf{H}_i^\star(\mathbf{S}_{0|i} - \frac{\zeta}{\sqrt{\bar{\alpha}_i}}\nabla_{\mathbf{S}_i}\log p(\mathbf{Y}|\mathbf{S}_{0|i}, \mathbf{H}_i^\star))^{\mathrm{T}}\|_{\mathrm{F}}^2 \tag{66}$$

This strategy is analogous to the approach in [41]. Intuitively, at the initiation of the reverse process, the estimated endmember matrix $\mathbf{S}_{0|t}$ is far from the real endmember matrix, yielding a huge reconstruction residual. In such a situation, a large value is assigned to ζ_t to acquire more information from the likelihood, thereby increasing the rate of convergence. As t approaches 0, $\mathbf{S}_{0|t}$ is well estimated, and the reconstruction residual is small. Therefore, ζ_t is set to a relatively low value, allowing for more information to be obtained from the spectral prior distribution and eventually leading to a higher quality estimation of the endmembers.

It is evident that the optimal problem in Eq. (66) is a quadratic programming problem. Therefore, we can derive the optimal ζ_i as,

$$\zeta_i = \sqrt{\bar{\alpha}_i}\frac{tr(\mathbf{A}_i^{\mathrm{T}}\mathbf{B}_i)}{tr(\mathbf{B}_i^{\mathrm{T}}\mathbf{B}_i)} \tag{67}$$

The definition of $\mathbf{A}$ and $\mathbf{B}$ are as follows,

$$\mathbf{A}_i \triangleq \mathbf{Y} - \mathbf{H}_i^{\star}\mathbf{S}_{0|i}^{\mathrm{T}} \tag{68}$$

$$\mathbf{B}_i \triangleq \mathbf{H}_i^{\star}(\nabla_{\mathbf{S}_i} \log p(\mathbf{Y}|\mathbf{S}_{0|i}, \mathbf{H}_i^{\star}))^{\mathrm{T}} \tag{69}$$

REFERENCES

[1] 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, 2012.

[2] Z. Shi, T. Shi, M. Zhou, and X. Xu, "Collaborative sparse hyperspectral unmixing using l_0 norm," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5495–5508, 2018.

[3] C. Kwan, B. Ayhan, G. Chen, J. Wang, B. Ji, and C.-I. Chang, "A novel approach for spectral unmixing, classification, and concentration estimation of chemical and biological agents," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 2, pp. 409–419, 2006.

[4] C.-I. Chang and D. Heinz, "Constrained subpixel target detection for remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1144–1159, 2000.

[5] D. Heinz and Chein-I-Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 3, pp. 529–545, 2001.

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

[7] M. E. Winter, "N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," in *Optics & Photonics*, 1999.

[8] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999.

[9] Z. Yang, G. Zhou, S. Xie, S. Ding, J.-M. Yang, and J. Zhang, "Blind spectral unmixing based on sparse nonnegative matrix factorization," *IEEE Trans. Image Process.*, vol. 20, no. 4, pp. 1112–1125, 2011.

[10] F. Zhu, Y. Wang, B. Fan, S. Xiang, G. Meng, and C. Pan, "Spectral unmixing via data-guided sparsity," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5412–5427, 2014.

[11] S. Zhang, G. Zhang, F. Li, C. Deng, S. Wang, A. Plaza, and J. Li, "Spectral-spatial hyperspectral unmixing using nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2022.

[12] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 765–777, 2007.

[13] S. Jia and Y. Qian, "Constrained nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 161–173, 2009.

[14] F. Xiong, Y. Qian, J. Zhou, and Y. Y. Tang, "Hyperspectral unmixing via total variation regularized nonnegative tensor factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2341–2357, 2019.

[15] F. Xiong, J. Zhou, S. Tao, J. Lu, and Y. Qian, "SNMF-Net: Learning a deep alternating neural network for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, 2022.

[16] Y. Qian, F. Xiong, Q. Qian, and J. Zhou, "Spectral mixture model inspired network architectures for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7418–7434, 2020.

[17] M. Zhao, S. Shi, J. Chen, and N. Dobigeon, "A 3-D-CNN framework for hyperspectral unmixing with spectral variability," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2022.

[18] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, 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," in *2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evol. in Remote Sens.*, 2010, pp. 1–4.

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

[21] X. Xu, B. Pan, Z. Chen, Z. Shi, and T. Li, "Simultaneously multiobjective sparse unmixing and library pruning for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 3383–3395, 2021.

[22] Y. Wei, X. Xu, B. Pan, T. Li, and Z. Shi, "A multiobjective group sparse hyperspectral unmixing method with high correlation library," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 7114–7127, 2022.

[23] A. Halimi, P. Honeine, and J. M. Bioucas-Dias, "Hyperspectral unmixing in presence of endmember variability, nonlinearity, or mismodeling effects," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4565–4579, 2016.

[24] Y. Zhou, A. Rangarajan, and P. D. Gader, "A spatial compositional model for linear unmixing and endmember uncertainty estimation," *IEEE Trans. Image Process.*, vol. 25, no. 12, pp. 5987–6002, 2016.

[25] A. Zare and K. Ho, "Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 95–104, 2014.

[26] D. Wang, Z. Shi, and X. Cui, "Robust sparse unmixing for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1348–1359, 2018.

[27] L. Drumetz, M.-A. Veganzones, S. Henrot, R. Phlypo, J. Chanussot, and C. Jutten, "Blind hyperspectral unmixing using an extended linear mixing model to address spectral variability," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3890–3905, 2016.

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

[29] P.-A. Thouvenin, N. Dobigeon, and J.-Y. Tourneret, "Hyperspectral unmixing with spectral variability using a perturbed linear mixing model," *IEEE Trans. Signal Process.*, vol. 64, no. 2, pp. 525–538, 2016.

[30] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 6840–6851.

[31] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," *arxiv preprint arXiv:2010.02502*, 2020.

[32] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," in *Proc. Int. Conf. Learn. Represent.*, 2021.

[33] J. Choi, S. Kim, Y. Jeong, Y. Gwon, and S. Yoon, "ILVR: Conditioning method for denoising diffusion probabilistic models," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2021, pp. 14 347–14 356.

[34] A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte, and L. Van Gool, "Repaint: Inpainting using denoising diffusion probabilistic models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, June 2022, pp. 11 461–11 471.

[35] K. Deng, Z. Jiang, Q. Qian, Y. Qiu, and Y. Qian, "A noise-model-free hyperspectral image denoising method based on diffusion model," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2023, pp. 7308–7311.

[36] S. Särkkä and A. Solin, *Applied Stochastic Differential Equations*, ser. Institute of Mathematical Statistics Textbooks. Cambridge: Cambridge University Press, 2019.

[37] Y. Song and S. Ermon, "Generative modeling by estimating gradients of the data distribution," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019.

[38] Y. Song, L. Shen, L. Xing, and S. Ermon, "Solving inverse problems in medical imaging with score-based generative models," *arxiv preprint arXiv:2111.08005*, 2022.

[39] H. Chung, J. Kim, M. T. Mccann, M. L. Klasky, and J. C. Ye, "Diffusion posterior sampling for general noisy inverse problems," in *Proc. Int. Conf. Learn. Represent.*, 2023.

[40] H. Chung, B. Sim, D. Ryu, and J. C. Ye, "Improving diffusion models for inverse problems using manifold constraints," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 35, 2022, pp. 25 683–25 696.

[41] J. Song, A. Vahdat, M. Mardani, and J. Kautz, "Pseudoinverse-guided diffusion models for inverse problems," in *Proc. Int. Conf. Learn. Represent.*, May 2023.

[42] H. Chung, J. Kim, S. Kim, and J. C. Ye, "Parallel Diffusion Models of Operator and Image for Blind Inverse Problems," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2023, pp. 6059–6069.

[43] X. Li, Y. Ren, X. Jin, C. Lan, X. Wang, W. Zeng, X. Wang, and Z. Chen, "Diffusion models for image restoration and enhancement – a comprehensive survey," *arXiv preprint arXiv:2308.09388*, 2023.

[44] J. Ho and T. Salimans, "Classifier-free diffusion guidance," in *Proc. Adv. Neural Inf. Process. Syst. Workshops*, 2021.

[45] P. Dhariwal and A. Nichol, "Diffusion models beat GANs on image synthesis," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 8780–8794.

[46] X. Meng and Y. Kabashima, "Diffusion model based posterior sampling for noisy linear inverse problems," *arxiv preprint arXiv:2211.12343*, 2022.

[47] D. Chen, Z. Zhou, J.-P. Mei, C. Shen, C. Chen, and C. Wang, "A geometric perspective on diffusion models," *arXiv preprint arXiv:2305.19947*, 2023.
[48] R. F. Kokaly, R. N. Clark, G. A. Swayze, K. E. Livo, T. M. Hoefen, N. C. Pearson, R. A. Wise, W. Benzel, H. A. Lowers, R. L. Driscoll, and A. J. Klein, *USGS Spectral Library Version 7*. U.S. Geological Survey, 2017, no. 1035.
[49] F. Xiong, Y. Qian, J. Zhou, and Y. Y. Tang, "Hyperspectral unmixing via total variation regularized nonnegative tensor factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2341–2357, 2018.
[50] J. M. P Nascimento and J. M. Bioucas-Dias, "Hyperspectral signal subspace estimation," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2007, pp. 3225–3228.
[51] B. Somers, M. Zortea, A. Plaza, and G. P. Asner, "Automated extraction of image-based endmember bundles for improved spectral unmixing," *IEEE J. Sel. Topics in Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 396–408, 2012.
[52] R. A. Borsoi, T. Imbiriba, and J. C. M. Bermudez, "A data dependent multiscale model for hyperspectral unmixing with spectral variability," *IEEE Trans. Image Process.*, vol. 29, pp. 3638–3651, 2020.
[53] S. Shi, M. Zhao, L. Zhang, Y. Altmann, and J. Chen, "Probabilistic generative model for hyperspectral unmixing accounting for endmember variability," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2022.

Keli Deng received the B.E. degree in automatic control from Xi'an Jiaotong University, Xi'an, China, in 2022. Now he is pursuing the Ph.D. degree in artificial intelligence from Zhejiang University, Hangzhou, China. His research interests include hyperspectral image processing and diffusion models.

Yuntao Qian (Senior Member, IEEE) received the B.E. and M.E. degrees in automatic control from Xi'an Jiaotong University, Xi'an, China, in 1989 and 1992, respectively, and the Ph.D. degree in signal processing from Xidian University, Xi'an, in 1996.

From 1996 to 1998, he was a Post-Doctoral Fellow with Northwestern Polytechnical University, Xi'an. Since 1998, he has been with the College of Computer Science, Zhejiang University, Hangzhou, China, where he became a Professor in 2002. From 1999 to 2001, in 2006, in 2010, in 2013, from 2015 to 2016, and in 2018, he was a Visiting Professor with Concordia University, Montreal, QC, Canada, Hong Kong Baptist University, Hong Kong, Carnegie Mellon University, Pittsburgh, PA, USA, the Canberra Research Laboratory of NICTA, Macau University, Taipa, Macau, and Griffith University, Nathan, QLD, Australia. His research interests include machine learning, signal and image processing, pattern recognition, and hyperspectral imaging.

Dr. Qian is currently an Associate Editor of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.

Jie Nie (Member, IEEE) received the B.S. and Ph.D. degrees in computer science from Ocean University of China, Qingdao, China, in 2002 and 2011, respectively.

From 2009 to 2010, she was a Visiting Scholar with the University of Pittsburgh, Pittsburgh, PA, USA. From 2015 to 2017, she was a Postdoctoral Fellow with Tsinghua University, Beijing, China. She is currently a Professor with Ocean University of China. Her current research interests include artificial intelligence and visual analysis of marine Big Data.

Jun Zhou (Senior Member, IEEE) received the B.S. degree in computer science and the B.E. degree in international business from Nanjing University of Science and Technology, Nanjing, China, in 1996 and 1998, respectively. He received the M.S. degree in computer science from Concordia University, Montreal, Canada, in 2002, and the Ph.D. degree from the University of Alberta, Edmonton, Canada, in 2006.

He is a professor in the School of Information and Communication Technology at Griffith University, Nathan, Australia. Previously, he had been a research fellow in the Research School of Computer Science at the Australian National University, Canberra, Australia, and a researcher at the Canberra Research Laboratory, NICTA, Australia. His research interests include pattern recognition, computer vision and hyperspectral image processing with their applications to remote sensing and environmental informatics.

He is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and Pattern Recognition journal.