Solving Spectrum Unmixing as a Multi-Task Bayesian Inverse Problem with Latent Factors for Endmember Variability

Dong Wu[1], **Mingmin Chi**[1,2*], **Xuan Zang**[1], **Bo Peng**[3]

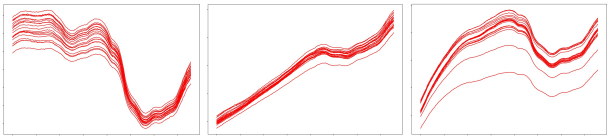
[1]Fudan University
[2]Zhongshan PoolNet Technology Co., Ltd
[3]Shanghai Ocean University

Abstract

With the increasing customization of spectrometers, spectral unmixing has become a widely used technique in fields such as remote sensing, textiles, and environmental protection. However, endmember variability is a common issue for unmixing, where changes in lighting, atmospheric, temporal conditions, or the intrinsic spectral characteristics of materials, can all result in variations in the measured spectrum. Recent studies have employed deep neural networks to tackle endmember variability. However, these approaches rely on generic networks to implicitly resolve the issue, which struggles with the ill-posed nature and lack of effective convergence constraints for endmember variability. This paper proposes a streamlined multi-task learning model to rectify this problem, incorporating abundance regression and multi-label classification with *U*nmixing as a *B*ayesian *I*nverse *P*roblem, denoted as *BIPU*. To address the issue of the ill-posed nature, the uncertainty of unmixing is quantified and minimized through the Laplace approximation in a Bayesian inverse solver. In addition, to improve convergence under the influence of endmember variability, the paper introduces two types of constraints. The first separates background factors of variants from the initial factors for each endmember, while the second identifies and eliminates the influence of non-existent endmembers via multi-label classification during convergence. The effectiveness of this model is demonstrated not only on a self-collected near-infrared spectral textile dataset (FENIR), but also on three commonly used remote sensing hyperspectral image datasets, where it achieves state-of-the-art unmixing performance and exhibits strong generalization capabilities.

Introduction

Due to the limited spatial resolution of hyperspectral images, each pixel is, in fact, a mixture of several different pure materials. The spectrum of an pure material is referred to as an endmember. Spectral unmixing (Keshava and Mustard 2002; Bioucas-Dias et al. 2012) addresses this issue by breaking down hyperspectral images into the spectral of endmembers and their respective ratios (Abundance) for each pixel. This technique has been widely used in a variety of applications, including mineral mapping (Rogge et al. 2006) and land cover change detection (Adams et al. 1995). With the continued shrinking in size of spectrometers, handheld spectrometers are now being used in the field of textile composition analysis. Spectral unmixing, denoted as $\boldsymbol{a} = \mathcal{F}(\boldsymbol{x})$, involves analyzing the abundance $\boldsymbol{a}$ of spectral materials from an input spectrum $\boldsymbol{x}$. On the other hand, spectral construction is $\hat{\boldsymbol{x}} = \mathcal{G}(\boldsymbol{a}) + \boldsymbol{\eta}$ of generating a corresponding spectrum, given the known abundance of spectral materials (where $\boldsymbol{\eta}$ represents random noise). In essence, spectral unmixing can be viewed as the inverse problem of spectral construction.



(a) Variation #1 (b) Variation #2 (c) Variation #3

Figure 1: There are three types of variations for the cotton textiles. The x-axis represents 200 equally divided spectral sampling points of the near-infrared spectrum ranging from 900 to 1700nm, while the y-axis depicts the reflectance.

Endmember variability (Zare and Ho 2013; Somers et al. 2011; Borsoi et al. 2021) is a common issue in unmixing, where changes in lighting, atmospheric conditions, temporal factors, or the intrinsic spectral characteristics of materials can all result in variations in the measured spectrum. This problem is particularly pronounced with low-cost, low-precision devices such as handheld spectrometers, as demonstrated in Figure 1, which depicts changes in the near-infrared reflectance of pure cotton textiles. Variability typically falls into two categories: intra-class variability and inter-class similarity. Traditional models use a single endmember to represent multiple variants within a class, leading to uncertainty in abundance estimation. Meanwhile, inter-class similarity can result in mismatched abundance ratios. Therefore, addressing endmember variability is crucial for accurate unmixing tasks.

Recent research (Borsoi et al. 2021) has sought to address the issue of endmember variability in unmixing tasks. One

*Corresponding author (mmchi@fudan.edu.cn). This work was supported in part by Natural Science Foundation of China under contract 62171139.

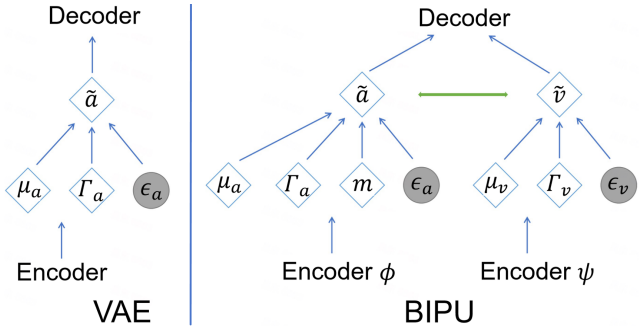


Figure 2: The conditional dependence of the VAE (left) and our proposed BIPU (right), where the diamond-shaped hollow nodes denote deterministic latent variables, the circular solid nodes denote random noises, and the total correlation network is applied to keep the distributions of the two sampled variables independent of each other.

common approach involves relying on a known spectral library (Roberts et al. 1998; Borsoi, Imbiriba, and Bermudez 2019; Borsoi et al. 2020), where all endmember variability spectra for each pure material endmember are treated as a group of spectra for unmixing. However, this approach is heavily dependent on the quality and completeness of the spectral library. Another approach involves using probability distributions to fit the endmembers of each pure material (Stein 2003; Halimi, Dobigeon, and Tourneret 2015; Zhou, Rangarajan, and Gader 2018). While these models offer better generalization capabilities, they come with unrealistic computational complexity. With the rise of deep neural networks, auto-encoder has also been applied to unmixing tasks. These methods (Su et al. 2019; Palsson et al. 2018; Su et al. 2018; Wang et al. 2019; Hong et al. 2021; Shi et al. 2021) leverage the non-linear representation of deep neural network feature space to tackle the complex unmixing problems caused by endmember variability. However, the use of a general network structure without explicit constraints on the nonlinear optimization difficulties brought about by endmember variability results in limited unmixing effectiveness.

The first issue we address is the fundamental inverse problem of unmixing, with its inherent ill-posed nature (Goh et al. 2019), making it complex and unstable to explicitly estimate the latent factors of abundance and endmembers. Accordingly, this paper aims to quantify and minimize unmixing uncertainty using the Laplace approximation, approaching it as a Bayesian inverse problem. The second issue pertains to the persistent endmember variability in unmixing, which results in diverse spectral shapes with abundance, complicating convergence during training. To mitigate this, we apply two constraints to enhance convergence. The first constraint separates the background factors of variants from the prior factor for each endmember. The second constraint improves abundance estimation by identifying non-existent endmembers through multi-label learning, thereby eliminating their influence on the convergence process and significantly reducing the solution space.

This paper presents three main contributions. Firstly, we employed joint training of spectral reconstruction, abundance regression, and multi-label classification tasks as an unified model. Our experiments demonstrate the positive impact of each task's training on the unmixing performance. Secondly, to model endmember variability in a more generalized manner, we introduced endmember variability latent factors and designed the conditional dependency relationship within the latent factors, as shown in Figure. 2. Through variational inference, we derived implementable upper bounds as the optimization objective with relatively low complexity. However, this approach relies on the assumption that endmember variability and abundance latent factors are mutually independent. To ensure the effectiveness of the optimization objective in real-world scenarios, we applied the total correction to maintain independence between the two distributions. Thirdly, we generated an abundance mask from the multi-label classification task to serve as a constraint during sampling and generating abundance, where the upper bounds of the optimization objective is tightened, resulting in more stable unmixing performance. This paper demonstrates the effectiveness of our model in analyzing the composition of textiles using near-infrared spectra on the self-collected FENIR dataset. Additionally, it verifies the generalization effect of our model on unmixing across three remote sensing hyperspectral datasets. Furthermore, we will release the FENIR dataset for near-infrared spectral research.

Related Works

Spectrum Unmixing

Traditional linear unmixing models can be primarily categorized into five types: geometric-based unmixing (Winter 1999), non-negative matrix factorization-based unmixing (Thouvenin, Dobigeon, and Tourneret 2016; Drumetz et al. 2016), prototype analysis-based unmixing (Zhao, Jia, and Zhao 2015), Bayesian-based unmixing (Chen, Nelson, and Tourneret 2016), and sparse unmixing (Tang et al. 2015). From a network architecture perspective, existing deep unmixing frameworks can be broadly classified into five types: unmixing based on autoencoders (Palsson et al. 2018; Wang et al. 2019; Su et al. 2018, 2019), convolutional neural networks (Zhang et al. 2018; Qi et al. 2020; Rasti et al. 2022; Palsson, Ulfarsson, and Sveinsson 2021; Huang et al. 2020), long short-term memory networks (Lei et al. 2020; Zhao, Yan, and Chen 2021), transformers (Ghosh et al. 2022), and deep generative networks. Specifically, (Su et al. 2019)'s work assumes that the encoder outputs the hidden layer as an abundance estimate. It makes a strong assumption that the dimension of the hidden layer is equal to the number of endmembers, directly ignoring the influence of the endmember variability.

Bayesian Inverse Problems

In scientific inverse problems, neural networks have been employed to enhance solvers. These works can be broadly classified into two categories. The first category uses deep

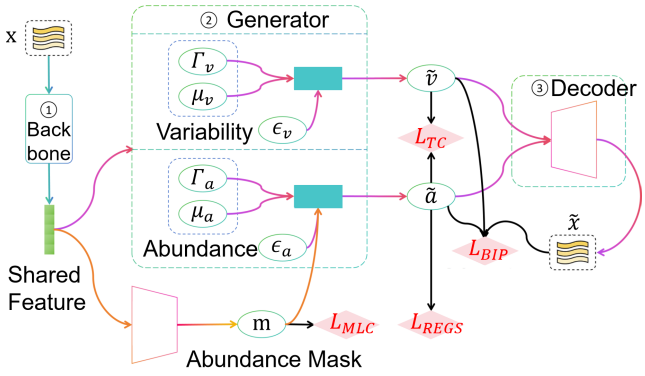


Figure 3: The BIPU model architecture comprises three stages. In the first stage, shared features are extracted for the joint learning of unmixing and multi-label classification. The second stage includes a generator that creates latent factors for both variability and abundance, with the assistance of an abundance mask from the classification task to enhance convergence. The third stage uses these latent factors to generate the reconstructed signals. Together, these three stages facilitate signal reconstruction with the loss derived from Bayesian Inversion. Additionally, the model features a prediction branch for multi-label classification at the bottom, which produces both the probability of the existence of each endmember and the abundance mask for the generator.

learning to assist in improving traditional solvers. For instance, (Adler and Öktem 2017) used convolutional neural networks to partially learn gradient descent schemes, while (Jin et al. 2017) used convolutional neural networks to replace components of convolutional operators in iterative algorithms. The second category of methods directly employs deep learning to regularize ill-posed inverse problems. (Li et al. 2020) used neural networks to replace the Tikhonov regularizer, while (Patel and Oberai 2019), as well as (Gonzalez et al. 2019), explored the use of Generative Adversarial Networks and Variational Auto-Encoder, respectively, to model prior probabilities in bayesian inverse problems. In (Chen et al. 2019)'s work, autoencoders were used to learn latent representations of unknown interests, allowing the inversion task to be redefined as solving for latent representations. Methods for measuring uncertainty using deep learning can be classified into three types. The first type (Jiang et al. 2021) employs deep learning networks to enhance traditional uncertainty measurement methods. The second type (Caldeira and Nord 2020) of method uses deep learning models with inherent stochastic properties to represent uncertainty. The third type of method directly employs deep networks to learn and model uncertainty. (Chua and Vallisneri 2020) generated one-dimensional or two-dimensional projections of Bayesian posteriors quickly through neural networks for gravitational wave astronomy. (Goh et al. 2019) modeled uncertainty from divergence-based variational inference, allowing most of the information typically present in scientific inverse problems to be fully utilized during training.

Models

Model Definition

Define $\boldsymbol{x} \in R^D$ as the input spectrum (observed spectrum), where D is the number of spectral domain sampling points; $\hat{\boldsymbol{x}} \in R^D$ represents the reconstructed spectrum; $\boldsymbol{a} \in R^A$ represents the spectral abundance, that is, the content of the observed spectrum corresponding to the physical object on A types of pure materials; $\boldsymbol{v} \in R^V$ represents the V-dimensional endmember variability latent variable, while V is a model hyperparameter, not necessarily equal to A and $V < D$. In the process of unmixing, $\boldsymbol{a} = \mathcal{F}_1(\boldsymbol{x}; \Phi)$ predicts the abundance, while $\boldsymbol{v} = \mathcal{F}_2(\boldsymbol{x}; \Psi)$ yields the representation of variability, where Φ and Ψ are the parameters of $\mathcal{F}_1$ and $\mathcal{F}_2$, respectively. In the process of generation, $\hat{\boldsymbol{x}} = \mathcal{G}(\boldsymbol{a}, \boldsymbol{v}) + \boldsymbol{\eta}$ produce the reconstructed sprectrum. As an inverse problem of spectrum generation, the objective of unmixing is

$$\begin{aligned} &\min_{\Phi,\Psi} \|\boldsymbol{x} - \mathcal{G}\left(\mathcal{F}_1(\boldsymbol{x}; \Phi), \mathcal{F}_2(\boldsymbol{x}; \Psi)\right)\|_2^2 \\ &+\frac{1}{A}\sum_{i=1}^{A} \|\boldsymbol{a}^{(i)} - \mathcal{F}_1^{(i)}(\boldsymbol{x}; \Phi)\|_2^2 \\ &+\frac{1}{V}\sum_{i=1}^{V} \|\boldsymbol{v}^{(i)} - \mathcal{F}_2^{(i)}(\boldsymbol{x}; \Psi)\|_2^2. \end{aligned} \tag{1}$$

However, this objective function is challenging to solve, and the training process can often only target a single point, making it difficult to perform uncertainty quantification. As a result, we analyze the unmixing problem as a Bayesian inverse problem.

Referring to the idea of using VAE to solve Bayesian Inverse Problems, our work combines two types of latent factors for modeling. Denote $p_1(\boldsymbol{a}|\boldsymbol{x})$, $p_2(\boldsymbol{v}|\boldsymbol{x})$ as the target posterior probability of abundance and variability respectively. Due to the assumption that the distributions of $\boldsymbol{a}$ and $\boldsymbol{v}$ are independent, $p(\boldsymbol{a}, \boldsymbol{v}|\boldsymbol{x}) = p(\boldsymbol{a}|\boldsymbol{x})p(\boldsymbol{v}|\boldsymbol{x})$. Denote $q_\Phi(\boldsymbol{a}|\boldsymbol{x})$ and $q_\Psi(\boldsymbol{v}|\boldsymbol{x})$ as the model posterior probability of abundance and variability, then $q(\boldsymbol{a}, \boldsymbol{v}|\boldsymbol{x}) = q_\Phi(\boldsymbol{a}|\boldsymbol{x})p_\Psi(\boldsymbol{v}|\boldsymbol{x})$. Denote p_{lkhd} as the likelihood model density, while $\overline{p}_{pr}$ and $\hat{p}_{pr}$ are the prior probability of abundance and variability, respectively. Suppose that the noise $\boldsymbol{\eta}$ is independent to $\boldsymbol{a}$ and $\boldsymbol{v}$, and the three latent variables follow Gaussian distributions as $\mathcal{N}(\boldsymbol{\mu}_e, \boldsymbol{\Gamma}_e)$, $\mathcal{N}(\overline{\boldsymbol{\mu}}_{pr}, \overline{\boldsymbol{\Gamma}}_{pr})$ and $\mathcal{N}(\hat{\boldsymbol{\mu}}_{pr}, \hat{\boldsymbol{\Gamma}}_{pr})$, respectively. In this way, the objective of the Bayesian inverse problem is to minimize

$$\begin{aligned} &p_{post}(\boldsymbol{a}, \boldsymbol{v}|\boldsymbol{x}) \propto p_{lkhd}(\boldsymbol{x}|\boldsymbol{a}, \boldsymbol{v})\overline{p}_{pr}(\boldsymbol{a})\hat{p}_{pr}(\boldsymbol{v}) \\ &= \exp(-\frac{1}{2}(\|\boldsymbol{x} - \mathcal{G}(\boldsymbol{a}, \boldsymbol{v}) - \boldsymbol{\mu}_e\|^2_{\boldsymbol{\Gamma}_e^{-1}}) \\ &+\|\boldsymbol{a} - \overline{\boldsymbol{\mu}}_{pr}\|^2_{\overline{\boldsymbol{\Gamma}}_{pr}^{-1}} + \|\boldsymbol{v} - \hat{\boldsymbol{\mu}}_{pr}\|^2_{\hat{\boldsymbol{\Gamma}}_{pr}^{-1}}). \end{aligned} \tag{2}$$

For the optimization process of Eq 2, to achieve the measurement and minimization of the uncertainty for the observed spectrum, Laplace approximation is applied as $\boldsymbol{\Gamma}_{post} = (\mathbf{J}_\mathcal{G}([\boldsymbol{a}, \boldsymbol{v}]_m)^{\mathrm{T}}\boldsymbol{\Gamma}_e^{-1}\mathbf{J}_\mathcal{G}([\boldsymbol{a}, \boldsymbol{v}]_m) + \overline{\boldsymbol{\Gamma}}_{pr}^{-1} + \hat{\boldsymbol{\Gamma}}_{pr}^{-1})^{-1}$, where $[\boldsymbol{a}, \boldsymbol{v}]_m$ are the maximum a posterior (MAP) estimates, and $\mathbf{J}_\mathcal{G}$ is the Jacobian matrix for $\mathcal{G}$ at the MAP

estimation. However, for this Laplace approximation, the iterative optimization process is computationally expensive and difficult to implement. In inverse problems, the measurement of the difference between the model posterior and target posterior distributions has symmetry, which is conducive to uncertainty quantification. Therefore, Jensen-Shannon divergence (JSD) is used here instead of Kullback-Leibler divergence (KLD), defined as $JS_\alpha(q(\boldsymbol{a},\boldsymbol{v}|\boldsymbol{x})\|p(\boldsymbol{a},\boldsymbol{v}|\boldsymbol{x})) = \alpha KL(q(\boldsymbol{a},\boldsymbol{v}|\boldsymbol{x})\|s(\boldsymbol{a},\boldsymbol{v}|\boldsymbol{x})) + (1-\alpha)KL(p(\boldsymbol{a},\boldsymbol{v}|\boldsymbol{x})\|s(\boldsymbol{a},\boldsymbol{v}|\boldsymbol{x}))$. where $s(\boldsymbol{a},\boldsymbol{v}|\boldsymbol{x}) = \alpha p(\boldsymbol{a},\boldsymbol{v}|\boldsymbol{x}) + (1-\alpha)q(\boldsymbol{a},\boldsymbol{v}|\boldsymbol{x})$, and $\alpha \in (0,1)$ is a hyper-parameter whose functionality will be described at the next section.

Framework of Unmixing

Figure. 3 illustrates the network structure of the model presented in this article. Initially, the Backbone network extracts the features of the input spectrum, serving as the input for two branches. The first branch generates the mean $\boldsymbol{\mu}_a$/$\boldsymbol{\mu}_v$ and variance $\boldsymbol{\Gamma}_a$/$\boldsymbol{\Gamma}_v$ of abundance and endmember variability latent variables through two fully connected layers, respectively. In contrast, the other branch generates the abundance mask $\boldsymbol{m}$ through an L_2 layer one-dimensional convolutional network. Subsequently, abundance and endmember variability latent variables are sampled to generate sampling values $\tilde{\boldsymbol{a}}$ and $\tilde{\boldsymbol{v}}$, respectively. Both are then input into the Decoder network for spectral reconstruction, generating the reconstructed spectrum $\hat{\boldsymbol{x}}$.

Within this framework, we need to solve the Bayesian inverse problem defined at Eq.**??**. Dividing both sides of Eq.**??** by α gives

$$
\begin{aligned}
&\frac{1}{\alpha}JS_\alpha(q(\boldsymbol{a},\boldsymbol{v}|\boldsymbol{x}) \\
=&\log(p(\boldsymbol{x})) - E_{\boldsymbol{a},\boldsymbol{v}\sim q}\left[\log\left(\frac{p(\boldsymbol{a},\boldsymbol{v},\boldsymbol{x})}{q(\boldsymbol{a},\boldsymbol{v}|\boldsymbol{x})}\right)\right] \\
&- E_{\boldsymbol{a},\boldsymbol{v}\sim q}\left[\log\left(\alpha + \frac{(1-\alpha)q(\boldsymbol{a},\boldsymbol{v}|\boldsymbol{x})}{p(\boldsymbol{a},\boldsymbol{v}|\boldsymbol{x})}\right)\right] \\
&- (1-\alpha)E_{\boldsymbol{a},\boldsymbol{v}\sim p}\left[\log\left(\alpha + \frac{(1-\alpha)q(\boldsymbol{a},\boldsymbol{v}|\boldsymbol{x})}{p(\boldsymbol{a},\boldsymbol{v}|\boldsymbol{x})}\right)\right] \quad (3) \\
\leq&\log(p(\boldsymbol{x})) + KL(q(\boldsymbol{a},\boldsymbol{v}|\boldsymbol{x})\|p(\boldsymbol{x})) \\
&- E_{\boldsymbol{a},\boldsymbol{v}\sim q}\log(p(\boldsymbol{x}|\boldsymbol{a},\boldsymbol{v})) \\
&- KL(q(\boldsymbol{a},\boldsymbol{v}|\boldsymbol{x})\|p(\boldsymbol{a},\boldsymbol{v}|\boldsymbol{x})) - \log(1-\alpha) \\
&+ \frac{1-\alpha}{\alpha}KL(p(\boldsymbol{a},\boldsymbol{v}|\boldsymbol{x})\|q(\boldsymbol{a},\boldsymbol{v}|\boldsymbol{x})) \\
&- \frac{(1-\alpha)\log(1-\alpha)}{\alpha}.
\end{aligned}
$$

In this way, the following formula can be applied as the objective of our variational inference process $\frac{1}{\alpha}JS_\alpha(q(\boldsymbol{a},\boldsymbol{v}|\boldsymbol{x}) + KL(q(\boldsymbol{a},\boldsymbol{v}|\boldsymbol{x})\|p(\boldsymbol{a},\boldsymbol{v}|\boldsymbol{x}))$, and considering the assumption that the distribution $\boldsymbol{a}$ is independent to $\boldsymbol{v}$, minimizing this objective is equivalent to minimize its upper bounds as

$$
\begin{aligned}
&\frac{1-\alpha}{\alpha}KL(p(\boldsymbol{a},\boldsymbol{v}|\boldsymbol{x})\|q(\boldsymbol{a},\boldsymbol{v}|\boldsymbol{x})) \\
&+ KL(q(\boldsymbol{a},\boldsymbol{v}|\boldsymbol{x})\|p(\boldsymbol{x})) - E_{\boldsymbol{a},\boldsymbol{v}\sim q}\log(p(\boldsymbol{x}|\boldsymbol{a},\boldsymbol{v})) \\
=&\frac{1-\alpha}{\alpha}KL(p(\boldsymbol{a}|\boldsymbol{x})\|q_\Phi(\boldsymbol{a}|\boldsymbol{x})) \\
&+ \frac{1-\alpha}{\alpha}KL(p(\boldsymbol{v}|\boldsymbol{x})\|q_\Psi(\boldsymbol{v}|\boldsymbol{x})) \\
&+ E_{\boldsymbol{a}\sim q_\Phi}\left[\log\left(\frac{q_\Phi(\boldsymbol{a}|\boldsymbol{x})}{p(\boldsymbol{x})}\right)\right] \\
&+ E_{\boldsymbol{v}\sim q_\Psi}\left[\log\left(\frac{q_\Psi(\boldsymbol{v}|\boldsymbol{x})}{p(\boldsymbol{x})}\right)\right] + \log(p(\boldsymbol{x})) \\
&- (E_{\boldsymbol{a}\sim q_\Phi}\log(p(\boldsymbol{x}|\boldsymbol{a})) + E_{\boldsymbol{v}\sim q_\Psi}\log(p(\boldsymbol{x}|\boldsymbol{v})) \\
&- \log(p(\boldsymbol{x})).
\end{aligned} \quad (4)
$$

In this way, the objective function over Φ and Ψ turns to

$$
\begin{aligned}
&\frac{1-\alpha}{\alpha}KL(p(\boldsymbol{a}|\boldsymbol{x})\|q_\Phi(\boldsymbol{a}|\boldsymbol{x})) \\
&+ KL(q_\Phi(\boldsymbol{a}|\boldsymbol{x})\|p(\boldsymbol{x})) - E_{\boldsymbol{a}\sim q_\Phi}\log(p(\boldsymbol{x}|\boldsymbol{a})) \\
+&\frac{1-\alpha}{\alpha}KL(p(\boldsymbol{v}|\boldsymbol{x})\|q_\Psi(\boldsymbol{v}|\boldsymbol{x})) \\
&+ KL(q_\Psi(\boldsymbol{v}|\boldsymbol{x})\|p(\boldsymbol{x})) - E_{\boldsymbol{v}\sim q_\Psi}\log(p(\boldsymbol{x}|\boldsymbol{v})),
\end{aligned} \quad (5)
$$

which demonstrates that selecting different values of α interpolates between zero-forcing $KL(p\|q_\Phi)$/$KL(p\|q_\Psi)$ and zero-avoiding $KL(q_\Phi\|p)$/$KL(q_\Psi\|p)$. As $\alpha \to 1$, Eq. 5 converges to the negative value of ELBO, while as $\alpha \to 0$, the influence of JSD gradually supersedes KLD. By adjusting α, our framework enables the optimizer to select the distance concept for directing the model posterior towards the target posterior, thereby providing control over data fitting and regularization.

Optimization

This section aims to transform the objective function Eq. 5 into an easily implementable loss function through approximation methods. Assume the model posteriors of abundance and variability follow the Gaussian distributions $\mathcal{N}(\boldsymbol{\mu}_a,\boldsymbol{\Gamma}_a)$ and $\mathcal{N}(\boldsymbol{\mu}_v,\boldsymbol{\Gamma}_v)$. For each observed spectrum $\boldsymbol{x}^{(n)}$ from N pieces of spectrum in training dataset, there is a unique corresponding abundance value $\boldsymbol{a}$ and variation $\boldsymbol{v}$. Minimizing the KL divergence between the target posteriors and the model posteriors is equivalent to maximizing the likelihood function over Φ and Ψ. Here, we apply two networks with shared parameters to fit the prior probability $[\boldsymbol{\mu}_a^{(n)},(\boldsymbol{\Gamma}_a^{(n)})^{-\frac{1}{2}}] = \mathcal{F}_\Phi(\boldsymbol{x},\boldsymbol{W}_a)$ and $[\boldsymbol{\mu}_v^{(n)},(\boldsymbol{\Gamma}_v^{(n)})^{-\frac{1}{2}}] = \mathcal{F}_\Psi(\boldsymbol{x},\boldsymbol{W}_v)$, where $\boldsymbol{W}_a,\boldsymbol{W}_v \in R^{L\times A\times D}$ are sets of parameters from L-layer networks of $\mathcal{F}_\Phi$ and $\mathcal{F}_\Psi$. Then the sampled representations of abundance and variability are defined as

$$\tilde{\boldsymbol{a}}^{(n)} = \boldsymbol{\mu}_a^{(n)} + (\boldsymbol{\Gamma}_a^{(n)})^{\frac{1}{2}}\boldsymbol{\epsilon}_a^{(n)}, \quad (6)$$

$$\tilde{\boldsymbol{v}}^{(n)} = \boldsymbol{\mu}_v^{(n)} + (\boldsymbol{\Gamma}_v^{(n)})^{\frac{1}{2}}\boldsymbol{\epsilon}_v^{(n)}, \quad (7)$$

where the noise $\boldsymbol{\epsilon}_a^{(n)} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_A)$ and $\boldsymbol{\epsilon}_v^{(n)} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_V)$. Therefore, the final objective of this Bayesian inverse problem is

$$\begin{aligned}
&\min_{\boldsymbol{W}_a, \boldsymbol{W}_v} \mathcal{L}_{BIP} \\
&= \min_{\boldsymbol{W}_a, \boldsymbol{W}_v} \frac{1}{N} \sum_{n-1}^{N} \frac{1-\alpha}{\alpha} \Big(\log |\boldsymbol{\Gamma}_a^{(n)}| + \log |\boldsymbol{\Gamma}_v^{(n)}| \\
&\quad + \|\boldsymbol{\mu}_a^{(n)} - \tilde{\boldsymbol{a}}^{(n)}\|^2_{(\boldsymbol{\Gamma}_a^{(n)})^{-1}} + \|\boldsymbol{\mu}_v^{(n)} - \tilde{\boldsymbol{v}}^{(n)}\|^2_{(\boldsymbol{\Gamma}_v^{(n)})^{-1}} \Big) \\
&\quad + \left\| \boldsymbol{x}^{(n)} - \mathcal{G}\left(\tilde{\boldsymbol{a}}^{(n)}, \tilde{\boldsymbol{v}}^{(n)}\right) - \boldsymbol{\mu}_e \right\|^2_{\boldsymbol{\Gamma}_e^{-1}} \\
&\quad + tr\left(\overline{\boldsymbol{\Gamma}}_{pr}^{-1} \boldsymbol{\Gamma}_a^{(n)}\right) + tr\left(\hat{\boldsymbol{\Gamma}}_{pr}^{-1} \boldsymbol{\Gamma}_v^{(n)}\right) \\
&\quad + \|\boldsymbol{\mu}_a^{(n)} - \overline{\boldsymbol{\mu}}_{pr}\|^2_{\overline{\boldsymbol{\Gamma}}_{pr}^{-1}} + \|\boldsymbol{\mu}_v^{(n)} - \hat{\boldsymbol{\mu}}_{pr}\|^2_{\hat{\boldsymbol{\Gamma}}_{pr}^{-1}} \\
&\quad + \log \frac{|\overline{\boldsymbol{\Gamma}}_{pr}|}{|\boldsymbol{\Gamma}_a^{(n)}|} + \log \frac{|\hat{\boldsymbol{\Gamma}}_{pr}|}{|\boldsymbol{\Gamma}_v^{(n)}|}.
\end{aligned} \tag{8}$$

Abundance Mask via Multi-Label Classification

As illustrated in Figure. 3, we employ a network $\mathcal{F}_\Theta$ that partially shares its parameters with networks $\mathcal{F}_\Phi$ and $\mathcal{F}_\Psi$, to predict the probability $\boldsymbol{h} = \mathcal{F}_\Theta(\boldsymbol{x}, \boldsymbol{W}_h)$ whether the spectrum of each pure material is in the input spectrum. This probability only indicates the likelihood of the presence of a specified pure material spectrum and is not necessarily equivalent to its abundance. $\boldsymbol{W}_h \in R^{L_2 \times A \times D}$ is the set of parameters from L_2-layer networks $\mathcal{F}_\Theta$. Cross entropy loss is applied to train the brunch of multi-label classification as

$$\mathcal{L}_{MLC} = \sum_{n=1}^{N} \sum_{k=1}^{A} \boldsymbol{a}_k^{(n)} \log(\boldsymbol{h}_k^{(n)}). \tag{9}$$

At the same time, we can use the spectral probability $\boldsymbol{h}$ from the multi-label classification task to constrain the convergence process of the Bayesian Inversion Solver, and generate masks $\boldsymbol{m}_k = \boldsymbol{I}[\boldsymbol{h}_k > \xi]$ for the k-th endmember, where $\boldsymbol{I}$ is an indicator function, indicating that if the condition inside the square brackets is satisfied, it is 1, otherwise it is 0; ξ indicates that when the probability $\boldsymbol{h}_k$ of the k-th material existing in the observed spectrum is greater than ξ, the mask is 1, otherwise the mask is 0. In this way, the abundance mask is applied at Eq. 6 to be

$$\tilde{\boldsymbol{a}}^{(n)} = \tilde{\boldsymbol{\mu}}_a^{(n)} + (\boldsymbol{\Gamma}_a^{(n)})^{\frac{1}{2}} \tilde{\boldsymbol{\epsilon}}_a^{(n)} = \boldsymbol{m}\boldsymbol{\mu}_a^{(n)} + (\boldsymbol{\Gamma}_a^{(n)})^{\frac{1}{2}} (\boldsymbol{m}\boldsymbol{\epsilon}_a^{(n)})). \tag{10}$$

Then replace $\boldsymbol{\mu}_a^{(n)}$ with $\tilde{\boldsymbol{\mu}}_a^{(n)}$ at Eq. 5.

Theoretically, when the formula converges and the number of dimensions where the mask $\boldsymbol{m}$ equals 1 is denoted as M, then $M \leq A$. Since the abundance on the material dimension where the mask is 0 is also 0, adding the mask will not significantly impact JSD or cause the newly generated loss $\mathcal{L}_{BIP}$ to lose its upper bound constraint. Simultaneously, as some dimensions are invalidated by multiplication, the newly generated loss $\mathcal{L}_{BIP}$ decreases, resulting in a tighter upper bound and prompting the model to produce stable results with higher probability.

aRMSE	$\frac{1}{N}\sum_{n=1}^{N}\sqrt{\frac{1}{A}\sum_{k=1}^{A}\left(\boldsymbol{a}_k^{(n)} - \tilde{\boldsymbol{a}}_k^{(n)}\right)^2}$
γ-accuracy	$\frac{1}{A\cdot N}\sum_{n=1}^{N}\sum_{k=1}^{A}\boldsymbol{I}\left[\left\|\boldsymbol{a}_k^{(n)} - \tilde{\boldsymbol{a}}_k^{(n)}\right\| \leq \gamma\right]$
accuracy	$\frac{1}{A\cdot N}\sum_{n=1}^{N}\sum_{k=1}^{A}\boldsymbol{I}\left[\left\|\boldsymbol{m}_k^{(n)} = \boldsymbol{I}(\boldsymbol{a}_k^{(n)} > \xi)\right\|\right]$

Table 1: The evaluation metrics used in the experiments

Total Correlation for Abundance and Variability

At the previous section of model definition, we posited that the distributions of the two latent variables, abundance $\boldsymbol{a}$ and endmember variation $\boldsymbol{v}$, are independent. However, this assumption is challenging to satisfy in real-world scenarios. Drawing inspiration from this paper, we aim to enhance the stability of our model by minimizing the KL divergence between $q(\boldsymbol{a}, \boldsymbol{v}|\boldsymbol{x})$ and $q_\Phi(\boldsymbol{a}|\boldsymbol{x})q_\Psi(\boldsymbol{v}|\boldsymbol{x})$. To achieve this, we train a classifier $\mathcal{H}$ to determine whether a set of generated latent variables $\boldsymbol{a}$ and $\boldsymbol{v}$ originate from p_Φ and q_Ψ or from $q(\boldsymbol{a}, \boldsymbol{v}|\boldsymbol{x})$. Consequently, our goal is to minimize the following formula:

$$\mathcal{L}_{TC} = \sum_n \log\left(\frac{1 - \mathcal{H}([\tilde{\boldsymbol{a}}^{(n)}, \tilde{\boldsymbol{v}}^{(n)}}{\mathcal{H}([\tilde{\boldsymbol{a}}^{(n)}, \tilde{\boldsymbol{v}}^{(n)}])}\right). \tag{11}$$

Besides, with the label of real abundance $\boldsymbol{a}^{(n)}$, we can constraint the sampled abundance with regression loss

$$\mathcal{L}_{REGR} = \sqrt{\frac{1}{A \cdot N}\sum_{n=1}^{N}\sum_{k=1}^{A}\left(\boldsymbol{a}_k^{(n)} - \tilde{\boldsymbol{a}}_k^{(n)}\right)^2}. \tag{12}$$

Finally, the loss function of our model is $\mathcal{L} = \mathcal{L}_{BIP} + \lambda_1\mathcal{L}_{MLC} + \lambda_2\mathcal{L}_{TC} + \lambda_3\mathcal{L}_{REGR}$.

For the convenience of noting, we denote the model in this paper as *BIPU*. As a comparison model for the ablation study, we denote the model with only the $\mathcal{L}_{BIP} + \mathcal{L}_{TC}$ loss as *BIPU*-base. Based on this, the model that only adds regression loss is denoted as *BIPU*-regr; the model that only adds multi-label classification is denoted as *BIPU*-mlc; and the model that adds both multi-label classification and abundance mask is denoted as *BIPU*-mask.

Experiments

Datasets

FENIR dataset [1]: a large-scale, near-infrared textile fiber dataset, compiled by professionals through the standardized collection of textile fiber product samples from various enterprises. Currently, FENIR data encompasses 12 common fiber components, accounting for approximately 80% of the market's textile fiber composition. FENIR's 12 primary fiber components comprise a total of 391,945 valid series of spectrum. The dataset primarily stores collected data information, such as wavelength, absorption rate, reflectivity, and light intensity, as well as corresponding component composition and content label information.

[1] https://drive.google.com/file/d/1_7XDhx_03TGutcPcU9GeWcwJ_f2SMs_8/view?usp=sharing

Synthetic Dataset: generated by selecting five endmembers from the USGS spectral library. The image comprises 200×200 pixels and 224 spectral bands, spanning a wavelength range of 0.4 to 2.5μm. Further details about the simulation process can be found at (Drumetz et al. 2016; Yao et al. 2021).

Jasper Ridge Dataset: This stunning hyperspectral scene was captured by the AVIRIS sensor, soaring high above the picturesque rural landscape of Jasper Ridge in sunny California, USA. The original image boasts an impressive 512×614 pixels, with a dazzling array of 224 wavelength bands spanning from 0.38 to 2.5μm, and a ground sampling distance (GSD) of just 20m. In this captivating study scene, we delved deep into the analysis of four main endmembers - the lush trees, the sparkling water, the rich soil, and the winding roads - with abundance maps sourced from the website [2].

EnMAP Munish Dataset: Crafted from the corresponding HyMap data, meticulously gathered from the bustling city of Munich in picturesque Germany, and expertly processed using the cutting-edge end-to-end EnMAP simulation tool: EeteS (Segl et al. 2012). They carefully selected an ROI that encompasses a remarkable 93×171 pixels, with a ground sampling distance (GSD) of just 30m, and an impressive 221 bands spanning a spectral range from 0.38 to 2.5μm. In this captivating scene, we delved into the analysis of the sturdy roofs, the sleek asphalt, the rich soil, the sparkling water, and the lush vegetation.

Experimental Results

Experiment Settings In this paper, set $\alpha = 0.1$, $\lambda_1 = 0.02$, $\lambda_2 = 0.9$, $\lambda_3 = 5e-3$, $\xi = 0.5$ and $V = \lceil \frac{A}{2} \rceil$. The numbers of endmember at the four datasets FENIR, Synthetic, Jasper Ridge and EnMAP Munish are 12, 5, 4 and 5, respectively, and therefore V are 6, 3, 2 and 3. We apply ResNet18 as the backbone networks, while the three-layer one-dimensional convolution network is applied to extract features for the multi-label classification. During training, the batch size is 128, and we use Adam optimizer for 500 epochs with the learning ratio 5e-3. We choose eight representative baseline models as comparison: NHU-DAE (Wang et al. 2019), DAEN (Su et al. 2019), HU-DCN (Zhang et al. 2018), CNNAEU (Palsson, Ulfarsson, and Sveinsson 2021), HU-FNNC (Lei et al. 2020), HU-LAE (Zhao, Yan, and Chen 2021), TANet (Jin et al. 2022), EGU-Net (Hong et al. 2021) and VAE (Shi et al. 2021), with slight tuned hyper-parameters at FENIR dataset. For the three remote sensing datasets, six baseline modes are used as comparison: FCLSU (Heinz et al. 2001), ALMM (Hong et al. 2018), DAEU (Palsson et al. 2018), DAEN, CNNAEU and EGU-Net. The metrics are shown as Tab. 1. Among them, aRMSE is a commonly used metric for unmixing to evaluate the overall performance of abundance estimation, γ-accuracy is used to measure the stability of abundance estimation, and accuracy is used to evaluate the overall performance of multi-label classification. The hardware platform is Ubuntu 20.04.2 LTS (GNU/Linux 5.15.0-78-generic x86_64), with two gpu cards of Nvidia GeForce 3090, and six cores of Intel(R) Xeon(R) Silver 4210R CPU @ 2.40GHz. We also use Pytorch in version of 1.8.1, running on 64G memories. The average computation time of training is about 68.4s for an epoch over 500 epochs.

Unmixing on FENIR From Tab. 2, it can be seen that, First and foremost, BIPU and its ingenious variants have achieved a remarkable improvement in performance and enhanced stability when compared to the baseline models, boasting at 18.60% to 31.48% reduction in aRMSE and a significant increase in γ-accuracy within the range of $\gamma \in [1\%, 30\%]$; In particular, VAE, EGU-Net, HU-LAE, and TANet have each proposed modeling approaches to address endmember variability. As demonstrated on the FENIR dataset, BIPU has proven to be particularly effective in handling this issue. Secondly, BIPU outshines its variant models in terms of aRMSE and γ-accuracy within the range of $\gamma \in [3\%, 30\%]$, with the difference in performance between γ-accuracy within the range of $\gamma \in [1\%, 2\%]$ and BIPU-mask being negligible, suggesting that BIPU has achieved a relatively stable SOTA performance on the FENIR dataset; Thirdly, BIPU-base surpasses the eight baseline models in terms of aRMSE and γ-accuracy within the range of $\gamma \in [1\%, 30\%]$, indicating that the Bayesian Inversion Solver proposed in this paper has made a significant leap forward in unmixing at FENIR; Fourthly, when compared to BIPU-base, BIPU-mlc sees a significant increase in aRMSE and only minor improvements within the range of $\gamma \in [1\%, 5\%]$, revealing that simply adding a multi-label classification branch does not significantly enhance unmixing performance; Fifthly, when compared to BIPU-mlc, BIPU-mask excels in all indicators, reaching its peak within the range of $\gamma \in [1\%, 2\%]$, suggesting that the multi-label classification branch can only effectively improve the stability of unmixing performance by utilizing abundance masks to constrain the upper bound; Sixthly, when compared to BIPU-base, BIPU-regr sees a significant reduction in aRMSE but its performance declines within the range of $\gamma \in [1\%, 10\%]$, indicating that while BIPU-regr can optimize unmixing performance, it needs to be combined with masks to achieve optimal unmixing stability.

Ablation Study We collected four sub-datasets for isolated training and evaluation, which are: L_C (Linen_Cotton), P_N_A (Poly_Nylon_Acrylic), N_A (Nylon_Acrylic) and CA_W_S (Cachmere_Wool_Silk), with data volumes of 81718, 40494, 9293 and 10366, respectively. From Figure. 4, it can be seen that, Firstly, by comparing BIPU-base to VAE, BIPU-mlc to VAE+MLC, and BIPU-mask to VAE+MLC+MASK, it becomes clear that the Bayesian Inversion Solver proposed in this paper has achieved a significant leap forward in the unmixing of FENIR when compared to VAE. Secondly, when compared to VAE, VAE+MLC sees significant improvements only on the L_C subset, while the enhancements on the other three subsets are not as pronounced. However, after the addition of the abundance mask, the performance has improved significantly. Similarly, BIPU-mlc does not see any notable improvements when compared to BIPU-base, but after

[2] https://rslab.ut.ac.ir/data

Models	aRMSE	1%	2%	3%	5%	10%	15%	20%	25%	30%
NHU-DAE	0.1489	8.18	14.63	20.09	28.26	37.93	43.73	47.68	51.66	55.36
DAEN	0.1468	10.12	15.71	20.98	30.07	40.22	46.18	52.00	55.88	58.19
HU-DCN	0.1431	12.09	17.38	23.00	33.19	42.88	51.95	60.07	63.39	66.90
CNNAEU	0.1322	12.95	17.10	24.84	34.68	45.23	54.03	61.25	62.87	68.99
HU-FNNC	0.1306	13.10	18.19	22.54	32.48	43.33	52.68	61.90	63.85	67.11
HU-LAE	0.1448	9.89	16.38	21.47	31.99	39.89	46.11	54.02	57.19	59.90
TANet	0.1284	14.69	19.15	24.90	35.40	48.88	55.19	62.33	65.96	68.83
EGU-Net	0.1250	16.05	18.94	24.56	36.03	50.22	57.79	62.12	65.23	69.09
VAE	0.1188	18.03	21.69	26.53	36.96	52.73	60.19	65.02	68.47	71.49
VAE+MLC+MASK	0.1079	22.18	23.94	28.73	39.20	58.41	65.79	74.68	76.33	78.90
BIPU-base	0.0875	25.09	31.58	37.83	49.57	63.89	72.81	79.06	81.45	84.72
BIPU-mlc	0.0967	26.38	33.19	37.60	51.72	62.03	70.84	72.77	76.91	79.17
BIPU-mask	0.0880	**33.13**	**39.51**	43.94	52.38	64.01	71.20	75.98	77.52	80.16
BIPU-regr	0.0848	28.92	33.17	39.40	51.21	61.82	71.57	79.84	80.99	84.33
BIPU	**0.0814**	32.94	39.47	**44.28**	**53.21**	**67.06**	**75.33**	**81.65**	**83.47**	**86.08**

Table 2: Empirical results of unmixing on FENIR dataset, with metrics of aRMSE and γ-accuracy, while VAE+MLC+MASK the multi-task training of VAE and multi-label classification with abundance mask.

the addition of the abundance mask, the performance has improved significantly. From this, it can be inferred that the multi-label classification branch can only effectively improve unmixing performance by adding a abundance mask to tighten the upper bound. Finally, on the L_C and N_A subsets, there is no significant difference in stability of unmixing between BIPU and BIPU-mask. On the P_N_A subset, BIPU-mask outperforms BIPU in terms of stability of unmixing. On the challenging CA_W_S subset, BIPU has significantly better stability of unmixing than BIPU-mask. This indicates that $\mathcal{L}_{regr}$ does not always play a role in improving performance on different subsets. However, when combined with Tab. 2, it can be seen that $\mathcal{L}_{regr}$ has a positive impact on all categories of FENIR. Therefore, we ultimately choose to add $\mathcal{L}_{regr}$ to the objective function of the BIPU model.

Performance of Multi-Label Classification In this section, we will delve deep into the improvement on multi-label classification. As can be seen from Tab. 3: Firstly, compared to only performing multi-label classification, the addition of multi-label classification and abundance mask to VAE has significantly improved classification performance, boasting an accuracy improvement range of 4.53% to 9.21%; at the same time, the accuracy improvement range of BIPU-mask compared to BIPU-mlc is 1.13% to 4.19%; this proves that the abundance mask still has a significant improvement effect on classification. Secondly, since BIPU-base does not possess multi-label classification capabilities, BIPU-mlc is used to compared with MLC here, with an accuracy improvement range of 4.75% to 9.85%; at the same time, BIPU-mask is compared with VAE+MLC+MASK, with an accuracy improvement range of 0.78% to 4.19%. This shows that the convergence process of the Bayesian inversion solver proposed in this paper has significant improvements on multi-label classification. Thirdly, compared to BIPU-mask, BIPU has an accuracy difference of -0.14%, +0.32%, -0.14% and +0.75% on the four subsets respectively, indicating that $\mathcal{L}_{regr}$ does not necessarily have an positive effect on multi-label classification performance on different datasets.

Method	L_C	P_N_A	N_A	CA_W_S
MLC	79.62	89.47	90.92	60.23
VAE+MLC+MASK	85.43	94.00	96.10	69.44
BIPU-mlc	84.89	93.52	95.75	70.08
BIPU-mask	**87.43**	95.31	**96.88**	74.27
BIPU	87.29	**95.63**	96.74	**75.02**

Table 3: Experimental results of multi-label classification on FENIR, with the metric of Accuracy(%).

Generalization on Hyperspectrum Images In order to verify the generalization ability of the model proposed in this paper, we have meticulously selected three remote sensing hyperspectral datasets and conducted extended experiments in terms of application scenarios for hyperspectrum images. As can be seen from Tab. 4: First and foremost, the unmixing performance of BIPU and its ingenious variants is significantly better than that of each baseline model, proving that the Bayesian inversion solver proposed in this paper has a stable and significant improvement effect on remote sensing hyperspectral unmixing tasks. Second, compared to BIPU-base, the unmixing performance of BIPU-mlc has significantly decreased, but the unmixing performance of BIPU-mask has significantly improved compared to BIPU-mlc, and has improved to some extent compared to BIPU-base on Jasper Ridge and EnMAP Munish; this further proves that only when the abundance mask is applied can the multi-label classification task effectively tighten the upper bound of the unmixing task, thereby effectively improving the unmixing performance. Third, the unmixing performance of BIPU-regr on remote sensing hyperspectral datasets is better than that of BIPU-base, indicating that $\mathcal{L}_{regr}$ has a certain degree of promoting effect on unmixing performance in this scenario. Fourth, BIPU is significantly

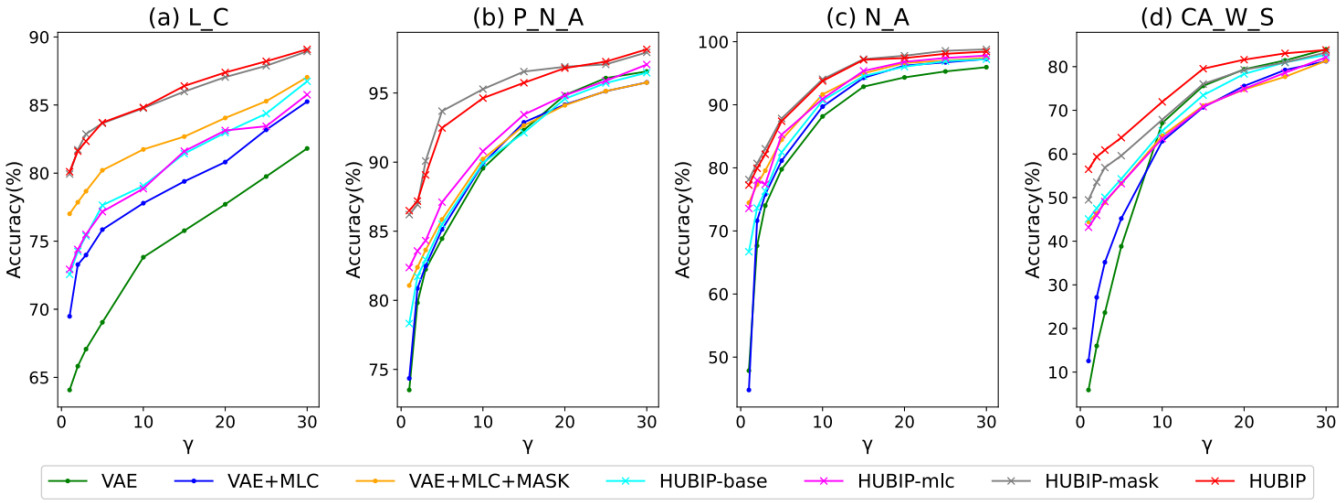


Figure 4: Empirical results of γ_accuracy on four subsets from FENIR dataset. VAE+MLC denotes the multi-task training of VAE and multi-label classification, while VAE+MLC+MASK the model with abundance mask based on VAE+MLC.

Methods	Synthetic(%)	Jasper(%)	EnMAP(%)
FCLSU	6.78±0.18	17.83±0.98	33.15±1.85
ALMM	2.47±0.19	10.15±1.72	22.97±3.15
DAEU	3.27±0.24	10.91±1.50	23.55±4.33
DAEN	2.57±0.11	9.57±2.34	21.18±4.36
CNNAEU	2.49±0.16	9.34±2.11	18.34±1.34
EGU-Net	1.83±0.07	8.61±2.11	16.80±1.07
BIPU-base	1.48±0.12	7.62±2.24	15.84±1.77
BIPU-mlc	1.76±0.10	8.04±1.49	16.97±2.45
BIPU-mask	1.53±0.17	7.54±1.96	**14.68**±2.03
BIPU-regr	1.46±0.13	7.76±1.50	15.57±1.83
BIPU	**1.35**±0.16	**7.02**±1.87	14.70±2.04

Table 4: Quantitative results of unmixing measured by aRMSE on Synthetic, Jasper Ridge and EnMAP Munish datasets.

better than BIPU-mask in terms of unmixing performance on Synthetic and Jasper Ridge datasets, and there is no significant difference between BIPU and BIPU-regr on EnMAP Munish. Therefore, it is believed that the model proposed in this paper, BIPU, can achieve a relatively stable SOTA effect.

Hyperparameters Tuning For α=[0.01, 0.05, 0.1, 0.15, 0.2, 0.3, 0.4, 0.5, 0.8], aRMSE=[0.1059, 0.0929, 0.0814, 0.0853, 0.0938, 0.1013, 0.1052, 0.1148, 0.1205]. For λ_1=[0.05, 0.01, 0.015, 0.02, 0.025, 0.03, 0.035, 0.04], aRMSE=[0.1017, 0.0939, 0.0853, 0.0814, 0.825, 0.0898, 0.0932, 0.1048]. For λ_2=[0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3], aRMSE=[0.0905, 0.087, 0.0832, 0.0822, 0.0814, 0.0818, 0.0819, 0.0841, 0.0838]. For λ_3=[5e-4, 1e-3, 2e-3, 5e-3, 1e-2, 2e-2, 5e-2], aRMSE=[0.0867, 0.0841, 0.0829, 0.0814, 0.0835, 0.0831, 0.0844]. For η=[0.3, 0.4, 0.5, 0.6, 0.7], aRMSE=[0.1131,0.1040, 0.0814, 0.0968, 0.01094]. It indicates that the performance is significantly sensitive to η, relatively sensitive to α & λ_1, and relatively stable with the settings of λ_2 and λ_3.

Conclusion

To address the issue of endmember variability, we has proposed a unified multi-task model that integrates unmixing, abundance regression, and multi-label classification.

Firstly, we validated the effectiveness of spectral reconstruction, abundance regression, and multi-label classification tasks respectively, through ablation experiments on near-infrared spectral data of textiles and remote sensing hyperspectral data. Our unified model achieved state-of-the-art unmixing performance in four scenarios across two domains, demonstrating the stability and generalization of BIPU unmixing. Secondly, by comparing the performance of BIPU-base against other baseline models that address endmember variability, we discovered that our proposed latent factors of variability can more effectively solve the problem of endmember variability when distribution correction is applied to both the latent factors of variability and abundance, verifying the effectiveness of our derived feasible upper bounds. Finally, after comparing the unmixing performance of BIPU-mlc with BIPU-base, we concluded that simply adding a multi-label classification task does not significantly improve unmixing performance. Only after incorporating an abundance mask did we observe a significant improvement in unmixing of BIPU-mask. By adding multi-label classification and abundance abundance masks to VAE, we also achieved significant improvements in unmixing performance. Furthermore, we verified that the abundance mask can also significantly improve the accuracy of multi-label classification. In future work, we will explore the use of nonlinear encoders to generate spectral feature of endmembers during reconstruction and investigate the characteristics of the factor of variability for a more interpretable model.

References

Adams, J. B.; Sabol, D. E.; Kapos, V.; Almeida Filho, R.; Roberts, D. A.; Smith, M. O.; and Gillespie, A. R. 1995. Classification of multispectral images based on fractions of endmembers: Application to land-cover change in the Brazilian Amazon. *Remote sensing of Environment*, 52(2): 137–154.

Adler, J.; and Öktem, O. 2017. Solving ill-posed inverse problems using iterative deep neural networks. *Inverse Problems*, 33(12): 124007.

Bioucas-Dias, J. M.; Plaza, A.; Dobigeon, N.; Parente, M.; Du, Q.; Gader, P.; and Chanussot, J. 2012. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE journal of selected topics in applied earth observations and remote sensing*, 5(2): 354–379.

Borsoi, R. A.; Imbiriba, T.; and Bermudez, J. C. M. 2019. Deep generative endmember modeling: An application to unsupervised spectral unmixing. *IEEE Transactions on Computational Imaging*, 6: 374–384.

Borsoi, R. A.; Imbiriba, T.; Bermudez, J. C. M.; and Richard, C. 2020. Deep generative models for library augmentation in multiple endmember spectral mixture analysis. *IEEE Geoscience and Remote Sensing Letters*, 18(10): 1831–1835.

Borsoi, R. A.; Imbiriba, T.; Bermudez, J. C. M.; Richard, C.; Chanussot, J.; Drumetz, L.; Tourneret, J.-Y.; Zare, A.; and Jutten, C. 2021. Spectral variability in hyperspectral data unmixing: A comprehensive review. *IEEE geoscience and remote sensing magazine*, 9(4): 223–270.

Caldeira, J.; and Nord, B. 2020. Deeply uncertain: comparing methods of uncertainty quantification in deep learning algorithms. *Machine Learning: Science and Technology*, 2(1): 015002.

Chen, P.; Nelson, J.; and Tourneret, J.-Y. 2016. Toward a Sparse Bayesian Markov Random Field Approach to Hyperspectral Unmixing and Classification. *IEEE Transactions on Image Processing*, vol. 26(n° 1): pp. 426–438.

Chen, P.; Wu, K.; Chen, J.; O'Leary-Roseberry, T.; and Ghattas, O. 2019. Projected Stein variational Newton: A fast and scalable Bayesian inference method in high dimensions. *Advances in Neural Information Processing Systems*, 32.

Chua, A. J.; and Vallisneri, M. 2020. Learning Bayesian posteriors with neural networks for gravitational-wave inference. *Physical review letters*, 124(4): 041102.

Drumetz, L.; Veganzones, M.-A.; Henrot, S.; Phlypo, R.; Chanussot, J.; and Jutten, C. 2016. Blind hyperspectral unmixing using an extended linear mixing model to address spectral variability. *IEEE Transactions on Image Processing*, 25(8): 3890–3905.

Ghosh, P.; Roy, S. K.; Koirala, B.; Rasti, B.; and Scheunders, P. 2022. Hyperspectral Unmixing Using Transformer Network. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–16.

Goh, H.; Sheriffdeen, S.; Wittmer, J.; and Bui-Thanh, T. 2019. Solving Bayesian inverse problems via variational autoencoders. *arXiv preprint arXiv:1912.04212*.

Gonzalez, M.; Almansa, A.; Delbracio, M.; Musé, P.; and Tan, P. 2019. Solving inverse problems by joint posterior maximization with a VAE prior. *arXiv preprint arXiv:1911.06379*.

Halimi, A.; Dobigeon, N.; and Tourneret, J.-Y. 2015. Unsupervised unmixing of hyperspectral images accounting for endmember variability. *IEEE Transactions on Image Processing*, 24(12): 4904–4917.

Heinz, D. C.; et al. 2001. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE transactions on geoscience and remote sensing*, 39(3): 529–545.

Hong, D.; Gao, L.; Yao, J.; Yokoya, N.; Chanussot, J.; Heiden, U.; and Zhang, B. 2021. Endmember-guided unmixing network (EGU-Net): A general deep learning framework for self-supervised hyperspectral unmixing. *IEEE Transactions on Neural Networks and Learning Systems*, 33(11): 6518–6531.

Hong, D.; Yokoya, N.; Chanussot, J.; and Zhu, X. X. 2018. An augmented linear mixing model to address spectral variability for hyperspectral unmixing. *IEEE Transactions on Image Processing*, 28(4): 1923–1938.

Huang, Y.; Li, J.; Qi, L.; Wang, Y.; and Gao, X. 2020. Spatial-Spectral Autoencoder Networks for Hyperspectral Unmixing. In *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, 2396–2399.

Jiang, Z.; Zhang, S.; Turnadge, C.; and Xu, T. 2021. Combining autoencoder neural network and Bayesian inversion to estimate heterogeneous permeability distributions in enhanced geothermal reservoir: Model development and verification. *Geothermics*, 97: 102262.

Jin, K. H.; McCann, M. T.; Froustey, E.; and Unser, M. 2017. Deep convolutional neural network for inverse problems in imaging. *IEEE transactions on image processing*, 26(9): 4509–4522.

Jin, Q.; Ma, Y.; Mei, X.; and Ma, J. 2022. TANet: An Unsupervised Two-Stream Autoencoder Network for Hyperspectral Unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–15.

Keshava, N.; and Mustard, J. F. 2002. Spectral unmixing. *IEEE signal processing magazine*, 19(1): 44–57.

Lei, M.; Li, J.; Qi, L.; Wang, Y.; and Gao, X. 2020. Hyperspectral Unmixing via Recurrent Neural Network With Chain Classifier. In *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, 2173–2176.

Li, H.; Schwab, J.; Antholzer, S.; and Haltmeier, M. 2020. NETT: Solving inverse problems with deep neural networks. *Inverse Problems*, 36(6): 065005.

Palsson, B.; Sigurdsson, J.; Sveinsson, J. R.; and Ulfarsson, M. O. 2018. Hyperspectral unmixing using a neural network autoencoder. *IEEE Access*, 6: 25646–25656.

Palsson, B.; Ulfarsson, M. O.; and Sveinsson, J. R. 2021. Convolutional Autoencoder for Spectral–Spatial Hyperspectral Unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 59(1): 535–549.

Patel, D.; and Oberai, A. A. 2019. Bayesian inference with generative adversarial network priors. *arXiv preprint arXiv:1907.09987*.

Qi, L.; Li, J.; Wang, Y.; Lei, M.; and Gao, X. 2020. Deep spectral convolution network for hyperspectral image unmixing with spectral library. *Signal Processing*, 176: 107672.

Rasti, B.; Koirala, B.; Scheunders, P.; and Ghamisi, P. 2022. UnDIP: Hyperspectral Unmixing Using Deep Image Prior. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–15.

Roberts, D. A.; Gardner, M.; Church, R.; Ustin, S.; Scheer, G.; and Green, R. 1998. Mapping chaparral in the Santa Monica Mountains using multiple endmember spectral mixture models. *Remote sensing of environment*, 65(3): 267–279.

Rogge, D. M.; Rivard, B.; Zhang, J.; and Feng, J. 2006. Iterative spectral unmixing for optimizing per-pixel endmember sets. *IEEE Transactions on Geoscience and Remote Sensing*, 44(12): 3725–3736.

Segl, K.; Guanter, L.; Rogass, C.; Kuester, T.; Roessner, S.; Kaufmann, H.; Sang, B.; Mogulsky, V.; and Hofer, S. 2012. EeteS—The EnMAP end-to-end simulation tool. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(2): 522–530.

Shi, S.; Zhao, M.; Zhang, L.; and Chen, J. 2021. Variational autoencoders for hyperspectral unmixing with endmember variability. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1875–1879. IEEE.

Somers, B.; Asner, G. P.; Tits, L.; and Coppin, P. 2011. Endmember variability in spectral mixture analysis: A review. *Remote Sensing of Environment*, 115(7): 1603–1616.

Stein, D. 2003. Application of the normal compositional model to the analysis of hyperspectral imagery. In *IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data, 2003*, 44–51. IEEE.

Su, Y.; Li, J.; Plaza, A.; Marinoni, A.; Gamba, P.; and Chakravortty, S. 2019. DAEN: Deep autoencoder networks for hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 57(7): 4309–4321.

Su, Y.; Marinoni, A.; Li, J.; Plaza, J.; and Gamba, P. 2018. Stacked nonnegative sparse autoencoders for robust hyperspectral unmixing. *IEEE Geoscience and Remote Sensing Letters*, 15(9): 1427–1431.

Tang, W.; Shi, Z.; Wu, Y.; and Zhang, C. 2015. Sparse Unmixing of Hyperspectral Data Using Spectral A Priori Information. *IEEE Transactions on Geoscience and Remote Sensing*, 53(2): 770–783.

Thouvenin, P.-A.; Dobigeon, N.; and Tourneret, J.-Y. 2016. Hyperspectral unmixing with spectral variability using a perturbed linear mixing model. *IEEE Transactions on Signal Processing*, 64(2): 525–538.

Wang, M.; Zhao, M.; Chen, J.; and Rahardja, S. 2019. Nonlinear unmixing of hyperspectral data via deep autoencoder networks. *IEEE Geoscience and Remote Sensing Letters*, 16(9): 1467–1471.

Winter, M. E. 1999. N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data. In Descour, M. R.; and Shen, S. S., eds., *Imaging Spectrometry V*, volume 3753, 266 – 275. International Society for Optics and Photonics, SPIE.

Yao, J.; Hong, D.; Xu, L.; Meng, D.; Chanussot, J.; and Xu, Z. 2021. Sparsity-enhanced convolutional decomposition: A novel tensor-based paradigm for blind hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–14.

Zare, A.; and Ho, K. 2013. Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing. *IEEE Signal Processing Magazine*, 31(1): 95–104.

Zhang, X.; Sun, Y.; Zhang, J.; Wu, P.; and Jiao, L. 2018. Hyperspectral Unmixing via Deep Convolutional Neural Networks. *IEEE Geoscience and Remote Sensing Letters*, 15(11): 1755–1759.

Zhao, G.; Jia, X.; and Zhao, C. 2015. Multiple endmembers based unmixing using Archetypal Analysis. In *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 5039–5042.

Zhao, M.; Yan, L.; and Chen, J. 2021. LSTM-DNN Based Autoencoder Network for Nonlinear Hyperspectral Image Unmixing. *IEEE Journal of Selected Topics in Signal Processing*, 15(2): 295–309.

Zhou, Y.; Rangarajan, A.; and Gader, P. D. 2018. A Gaussian mixture model representation of endmember variability in hyperspectral unmixing. *IEEE Transactions on Image Processing*, 27(5): 2242–2256.