Mixture-Net: Low-Rank Deep Image Prior Inspired by Mixture Models for Spectral Image Recovery

Tatiana Gelvez-Barrera, *Member, IEEE* and Jorge Bacca, *Member, IEEE,* and Henry Arguello, *Senior Member, IEEE*

Abstract—This paper proposes a non-data-driven deep neural network for spectral image recovery problems such as denoising, single hyperspectral image super-resolution, and compressive spectral imaging reconstruction. Unlike previous methods, the proposed approach, dubbed Mixture-Net, implicitly learns the prior information through the network. Mixture-Net consists of a deep generative model whose layers are inspired by the linear and non-linear low-rank mixture models, where the recovered image is composed of a weighted sum between the linear and non-linear decomposition. Mixture-Net also provides a low-rank decomposition interpreted as the spectral image abundances and endmembers, helpful in achieving remote sensing tasks without running additional routines. The experiments show the Mixture-Net effectiveness outperforming state-of-the-art methods in recovery quality with the advantage of architecture interpretability.

Index Terms—Spectral Image Recovery, Model-based Optimization, Deep Image Prior, Low-rank Spectral Mixture Models.

I. INTRODUCTION

Spectral imaging sensors acquire spatial-spectral information of a scene, organized in a datacube known as a spectral image (SI), which is helpful in remote sensing applications such as unmixing and material identification [1]. The SI acquisition through specialized optical systems can be affected by artifacts linked with external conditions and optical aberrations that corrupt, degrade, or diminish the number of measurements. Therefore, recovering the underlying scene is a critical step in the SI analysis [2].

The atmospheric conditions, light source nature, and photon efficiency can add noise, outlines, and striping, yielding a denoising problem [3], [4]. The technology limitation inhibits the acquisition of high spatial-spectral resolution images, leading to a hyperspectral super-resolution (HSI-SR) problem [5], [6]. The compressive spectral imaging (CSI) paradigm provides a small number of projected measurements, yielding a reconstruction problem [7], [8]. Using prior information is crucial in current recovery approaches to solve the aforementioned ill-posed problems, including *greedy* algorithms, *model-based* optimization, *data-driven* and *learning-based* strategies.

T.Gelvez-Barrera and J.Bacca both are co-first authors with equal contribution. T. Gelvez-Barrera is with the Department of Electrical Engineering, Universidad Industrial de Santander, Bucaramanga, Colombia, 680002, e-mail: tatiana.gelvez@correo.uis.edu.co. J. Bacca and H. Arguello are with the Department of Computer Science, Universidad Industrial de Santander, Bucaramanga, Colombia, 680002 e-mail: jorge.bacca1@correo.uis.edu.co; henarfu@uis.edu.co. This work was supported by the Sistema General de Regalías (SGR) under Project 8933.

Greedy algorithms include the prior information by following a sequence of reasonable steps, where only locally optimal solutions can be guaranteed [9]. Meanwhile, model-based optimization considers the prior information by designing hand-crafted regularizers to reduce the feasible search set and find globally optimal solutions [10]–[12]. For instance, the low-rank regularizer promotes low-dimensionality by modeling the SI as a low-rank structure [13]–[16]. Nonetheless, hand-crafted regularizers are often insufficient to handle the wide spectral information variety. Next, data-driven methods such as deep-learning (DL) learn the prior information by training a black-box non-linear mapping from a dataset. The black-box nature has been tackled by connecting model-based iterative algorithms and deep neural networks [17]–[19]. Nonetheless, data-driven methods are still impractical because of the expensive acquisition cost of several SI datasets.

Deep image prior (*DIP*) is a non-data-driven DL approach that overcame the data dependency limitation, showing that a single generator network is sufficient to capture the low-level SI statistics [20]–[22]. In particular, DIP approaches usually entail autoencoder-based networks that learn the underlying image's coding and decoding process, providing promising results for remote sensing tasks [23]. Still, the network structure is a black-box that lacks interpretability.

Further, [24]–[27] proposed network architectures including physical factors as the abundance and the endmembers with a linear mixture model, mitigating the lack of interpretability. However, the non-linear interactions and the benefits of using an interpretable network to solve different tasks have not been considered. Therefore, this paper proposes a non-data-driven SI recovery method that, contrary to previous methods, considers the non-linear interactions and adds interpretability to the network using linear and non-linear mixture models.

The proposed SI recovery method contains three main components, the input structure, the network architecture, and the customized loss function. The input structure uses a low-rank decomposition of the network input based on the SI low-dimensionality prior. The network architecture, termed Mixture-Net, consists of a sequence of interpretable deep-blocks composed of three trainable block-layers representing the abundance, the endmember, and the non-linear model learning. The customized loss function consists of each deep-block loss sum with custom regularizers considering spatial-spectral correlations. Remark that, Mixture-Net is interpretable in that the learned features and weights can be interpreted as the abundances and endmembers, supporting remote sensing tasks such as unmixing and material identification without

running complementary routines. Simulations over six datasets and three SI recovery tasks demonstrate the effectiveness of the proposed interpretable Mixture-Net architecture.

The contributions of the paper are summarized as follows,

- Mixture-Net: An interpretable network architecture inspired by linear and non-linear low-rank mixture models, where the learned weights and features can be interpreted as the SI abundances and endmembers (Section III-B).
- A scheme of multiple deep-blocks, whose loss functions contain particular model-based regularizers to improve the SI learned spatial-spectral correlations (Section III-C).
- A non-data-driven DIP-based approach for SI denoising, HSI-SR, and CSI reconstruction (Section III).
- A significant improvement for SI recovery in terms of recovery quality while reducing the processing complexity through the non-data-driven approach (Section IV).

II. MATHEMATICAL SPECTRAL IMAGING BACKGROUND

Let $\mathbf{f} \in \mathbb{R}^n, n = N \times N \times L$, be the vector form of a SI with $N \times N$ spatial pixels and L spectral bands. A SI with hundred of bands is referred to as a hyperspectral image (HSI) [28]. A SI usually contains a few number $r \ll L$ of different materials uniquely represented by the spectral response. Then, the ith spatial pixel $\mathbf{f}_i \in \mathbb{R}^L$ can be modeled as the linear combination of the form $\mathbf{f}_i = \mathbf{E}\mathbf{a}_i$, known as the linear mixture model (LMM). $\mathbf{E} \in \mathbb{R}^{L \times r}$ denotes an endmember matrix, whose columns contain a unique spectral response, and $\mathbf{a}_i \in \mathbb{R}^r$ denotes an abundance vector, whose elements contain the fractional proportions of each endmember at the ith spatial pixel. Consequently, $\mathbf{f}$ can be low-rank represented as

$$\mathbf{f} = (\mathbf{E} \otimes \mathbf{I}_{N^2})\bar{\mathbf{a}} = \bar{\mathbf{E}}\bar{\mathbf{a}}, \tag{1}$$

where $\bar{\mathbf{a}} \in \mathbb{R}_+^{N^2 r} = [\mathbf{a}_1^T \ldots \mathbf{a}_i^T \ldots \mathbf{a}_{N^2}^T]^T$ stacks the abundances; $\bar{\mathbf{E}} \in \mathbb{R}_+^{N^2 L \times N^2 r}$ encompasses the endmembers spanning the SI; $\mathbf{I}_{N^2}$ is an identity matrix of size $N^2 \times N^2$; and $\otimes$ denotes the Kronecker product, introduced to apply the endmembers along the spatial pixels in vector notation. For details of the low-rank representation refer to [11]. Usually, the LMM does not reliably describe the complex spatial-spectral endmember interactions [29]. Therefore, non-linear mixture models (NLMMs) aim to consider non-linear interactions and scattering factors via flexible models described by

$$\mathbf{f} = \mathcal{N}(\mathbf{E}, \bar{\mathbf{a}}), \tag{2}$$

where $\mathcal{N}$ stands for an implicit function that defines non-linear interactions between the endmembers and the abundances [30].

The mixture models impose some physical constraints over components $\mathbf{E}$ and $\bar{\mathbf{a}}$. Precisely, the non-negativity considering the nature of reflected light in spectral signatures, so that the entries of the endmember matrix $\mathbf{E}$ have to be non-negative values, and the sum-to-one constraint considering the fractional proportions of each endmember at each pixel has to be one, i.e., $\mathbf{a}_i^T \mathbf{1} = 1, \forall i \in N^2$, where $\mathbf{1}$ is a vector with all values in 1.

A. Spectral Image Recovery Problems

The SI acquisition process usually produces degraded and noisy measurements, described mathematically through the general vector forward linear model given by

$$\mathbf{y} = \mathbf{\Phi}\mathbf{f} + \boldsymbol{\omega}, \tag{3}$$

where $\mathbf{y} \in \mathbb{R}^m$ stands for the measurements, $\mathbf{\Phi} \in \mathbb{R}^{m \times n}$ stands for the sensing matrix, and $\boldsymbol{\omega} \in \mathbb{R}^m$ stands for added noise. This paper addresses the following problems,

1) Denoising: when $m = n$, $\mathbf{\Phi} = \mathbf{I}_n \in \mathbb{R}^{n \times n}$ denotes the identity matrix, and $\mathbf{y}$ models a noisy SI. This problem appears by external factors of the optical system and environmental conditions during the acquisition process.

2) Single Hyperspectral Super-Resolution: when $m = n/d$, $\mathbf{\Phi} = \mathbf{DB} \in \mathbb{R}^{m \times n}$, with $\mathbf{B} \in \mathbb{R}^{n \times n}$ being a blurring matrix, and $\mathbf{D} \in \mathbb{R}^{m \times n}$ being a spatial downsampling matrix with downsampling factor $d \in \mathbb{Z}_{++}$, and $\mathbf{y}$ models a spatially degraded SI, referred to as low-resolution HSI (LR-HSI). This problem occurs by the limited incident light affecting the spatial resolution when acquiring several spectral bands.

3) Compressive Spectral Imaging Reconstruction: when $m \ll n$, $\mathbf{\Phi} = \mathbf{H} \in \mathbb{R}^{m \times n}$ denotes a compressive sensing matrix, and $\mathbf{y} \in \mathbb{R}^m$ models the SI compressed measurements. This problem is related to the compressive sensing theory that simultaneously acquire and compress a signal.

The SI can be recovered from the noisy, LR-HSI or compressed measurements by solving the inverse problem

$$\hat{\mathbf{f}} \in \underset{\mathbf{f} \in \mathbb{R}^n}{\operatorname{argmin}} \ F(\mathbf{f}|\mathbf{y}) + \lambda R(\mathbf{f}), \tag{4}$$

where $F(\cdot) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ denotes a data fidelity term, $R(\cdot) : \mathbb{R}^n \to \mathbb{R}$ denotes a regularizer that promotes SI prior information, and $\lambda > 0$ denotes the regularization parameter.

III. PROPOSED SPECTRAL IMAGE RECOVERY METHOD

The proposed SI recovery method aims to include the prior information implicitly in the architecture of a deep model that generates the SI measurements by minimizing

$$\hat{\theta} \in \underset{\theta}{\operatorname{argmin}} \ \mathcal{L}\left(\mathbf{y}, \mathbf{\Phi}\mathcal{M}_\theta(\mathbf{f}^0)\right), \tag{5}$$

where $\hat{\mathbf{f}} := \mathcal{M}_{\hat{\theta}}(\mathbf{f}^0)$ denotes the recovered SI, $\mathcal{M}_\theta(\cdot) : \mathbb{R}^n \to \mathbb{R}^n$ denotes the deep generative network with θ adjustable weights, $\mathcal{L}(\cdot) : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ denotes a customized loss function, and $\mathbf{f}^0 \in \mathbb{R}^n$ denotes the model input. Notice that (5) only requires the measurements $\mathbf{y}$, the acquisition operator $\mathbf{\Phi}$, and the input $\mathbf{f}^0$, i.e., the proposed method is non-data-driven.

Figure 1 schematizes the three components of the proposed SI recovery method. (i) The input structure is based on tensor decomposition. (ii) The Mixture-Net interpretable architecture generates the recovered SI, the learned features interpreted as the abundances, and the adjusted weights interpreted as the endmembers. (iii) The customized loss function includes the forward model, the losses of each deep-block, and the mixture model physical constraints.

Fig. 1: Proposed SI recovery method scheme. (i) The input $\mathbf{f}_0$ is learned as a tensor decomposition to impose a low-rank structure. (ii) Mixture-Net is composed by a sequence of multiple interpretable deep-blocks $\mathcal{M}_{\theta^k}(\cdot)$. The kth interpretable deep-block contains an abundance block-layer $\mathcal{A}_{\theta}^k(\mathbf{f}^{k-1})$, consisting of a CNN to learn the spatial correlations, an endmember block-layer $\mathcal{E}_{\mathbf{E}}(\bar{\mathbf{a}}^k)$ performing the matrix product according to the LMM, and a non-linearity block-layer $\mathcal{N}_{\theta^k}(\mathcal{E}_{\mathbf{E}^k}(\bar{\mathbf{a}}^k))$ containing a CNN to learn the NLMM operator. (iii) The customized loss function $\mathcal{L}(\mathbf{f}|\,\mathbf{y}, \mathbf{\Phi})$ is formulated as a weighted sum of the single losses employed at each interpretable deep-block $\mathcal{L}^k\left(\mathcal{M}_{\theta^k}(\mathbf{f}^{k-1})|\,\mathbf{y}, \mathbf{\Phi}\right)$, whose first term corresponds to the data fidelity term $\left\|\mathbf{y} - \mathbf{\Phi}\mathbf{f}^k\right\|_2^2$, and second term relates the abundance sum-to-one constraint.

A. Input structure component

The input component determines the input structure for the network $\mathcal{M}_\theta(\cdot)$. The proposed method computes the input as an adjustable variable from a blind representation by solving

$$\{\hat{\theta}, \hat{\mathbf{f}}^0\} \in \underset{\theta, \mathbf{f}^0 \in \mathbb{R}^n}{\operatorname{argmin}} \mathcal{L}\left(\mathbf{y}, \mathbf{\Phi}\mathcal{M}_\theta(\mathbf{f}^0)\right). \quad (6)$$

Authors in [21] suggested that imposing a low-dimensional input $\mathbf{f}^0$ will force obtaining a low-dimensional output, helpful for capturing the spectral data structure even in the first layer of the model [31]. Therefore, the input is learned according to the Tucker decomposition as $\mathbf{f}^0 = \operatorname{vec}(\mathcal{Z})$, $\mathcal{Z} = \mathcal{Z}_0 \times_1 \mathbf{U} \times_2 \mathbf{V} \times_3 \mathbf{W}$, where the variables $\mathcal{Z}_0, \mathbf{U}, \mathbf{V}, \mathbf{W}$ are fitted by minimizing (6). The Tucker decomposition maintains the SI 3D structure and guarantees low-dimensionality given that $\mathcal{Z}_o \in \mathbb{R}^{N_\rho \times N_\rho \times L_\rho}$ stands for a 3D low-rank tensor, where $N_\rho < N$ and $L_\rho < L$, and ρ is a scale factor.

B. Mixture-Net: Network architecture component

To implicitly capture the SI prior information, Mixture-Net is composed of a sequence of K interpretable deep-blocks inspired by low-rank mixture models, containing an abundance, an endmember, and a non-linearity block-layer.

1) Abundance block-layer: The first block-layer consists of a CNN that filters the input with the size of the reference SI to obtain an output whose structure and dimensions should match for being interpreted as the abundances. The abundance block-layer of the k-th interpretable deep-block can be expressed as

$$\bar{\mathbf{a}}^k = \mathcal{A}_\theta^k(\mathbf{f}^{k-1}), \quad (7)$$

where $\mathcal{A}_\theta^k(\cdot) : \mathbb{R}^{N^2L} \rightarrow \mathbb{R}^{N^2r}$ connotes the CNN with r being a tunable hyper-parameter related to the SI rank. The block-layer also includes the non-negativity and sum-to-one physical constraints described in Section II. Precisely, the *sigmoid* function is used as the activation of the last layer, and the following regularization term is included in the loss function

$$R(\bar{\mathbf{a}}^k) = \sum_{i=1}^{N^2} \left(\left(\mathbf{a}_i^k\right)^T \mathbf{1} - 1\right)^2. \quad (8)$$

2) Endemember block-layer: The second block-layer consists of an operator that performs the matrix multiplication between the learned features in the abundance block-layer $\bar{\mathbf{a}}^k$ and the adjusted model weights in the endmember block-layer $\mathbf{E}^k$, whose dimensions should match to be interpreted as the endmembers. The endmember block-layer can be expressed as the linear component of the low-rank mixture model as

$$\mathbf{L}^k = \mathcal{E}_{\mathbf{E}^k}\left(\bar{\mathbf{a}}^k\right) = (\mathbf{E}^k \otimes \mathbf{I}_{N^2})\bar{\mathbf{a}}^k, \quad (9)$$

where $\mathcal{E}_{\mathbf{E}^k}(\cdot) : \mathbb{R}^{N^2r} \rightarrow \mathbb{R}^{N^2L}$ models the fully connected endmember block-layer following the LMM. The adjusted weights $\mathbf{E}^k$ are meant to be interpreted as the endmembers, so that their entries must be non-negative (Section II). An activation function will not guarantee non-negativity since the activations affect the layers' outputs, not the weights. Therefore, a projection to the positive real numbers is imposed over the adjusted weights at each gradient step to guarantee the weights interpretability, i.e., to project $\mathbf{E}^k$ into $\mathbb{R}_+$.

3) Non-Linearity block-layer: The third block-layer consists of a convolutional operator determining the non-linear transformation applied to the linear component obtained in the

endmember block-layer $\mathbf{L}^k$. The block-layer learns the non-linear interactions providing an output whose structure and dimension should match to be interpreted as the recovered SI. The third block-layer can be expressed as the non-linear component of the low-rank mixture model as

$$\mathbf{NL}^k = \mathcal{N}_{\theta^k}(\mathcal{E}_{\mathbf{E}^k}(\bar{\mathbf{a}}^k)), \quad (10)$$

where $\mathcal{N}_\theta^k(\cdot) : \mathbb{R}^{N^2 r} \to \mathbb{R}^{N^2 L}$ connotes a CNN that determines the non-linear transformation following the NLMM.

In summarizing, the kth interpretable deep-block estimates the recovered SI, generating three outputs given by

$$\begin{aligned} \mathcal{M}_{\theta^k}(\mathbf{f}^{k-1}) = \quad \mathbf{f}^k &= (1-\lambda)\mathbf{L}^k + \lambda\mathbf{NL}^k, \\ \mathbf{L}^k &= \mathcal{E}_{\mathbf{E}^k}(\mathcal{A}_\theta^k(\mathbf{f}^{k-1})), \quad (11) \\ \mathbf{NL}^k &= \mathcal{N}_{\theta^k}(\mathcal{E}_{\mathbf{E}^k}(\mathcal{A}_\theta^k(\mathbf{f}^{k-1}))), \end{aligned}$$

for $k = 1, \ldots, K$ interpretable deep-blocks. Note that each $\mathbf{f}^k$ follows a low-rank representation since it results from an affine combination between $\mathbf{L}^k$ in (9) and $\mathbf{NL}^k$ in (10), interpreted as the linear and non-linear components of the low-rank mixture model balanced by the parameter $0 \leq \lambda \leq 1$.

C. Customized loss function component

The loss function is decisive for the learning effectiveness in each SI recovery problem. Therefore, the proposed method admits a customized loss function considering the forward model, the independent deep-blocks, and the physical constraints. First, the forward operator $\mathbf{\Phi}$ is applied at each deep-block output to predict the measurements. Subsequently, the weights $\boldsymbol{\theta}$ are adjusted by minimizing a loss that measures the predicted measurements fidelity to the observed measurements through a chosen data-fidelity term, $F(\mathbf{f}^k|\mathbf{y}, \mathbf{\Phi})$, and contains the mixture model physical constraint, $R(\mathcal{A}_\theta^k(\mathbf{f}^k)) = \sum_{i=1}^{N^2}\left(\left(\mathbf{a}_i^k\right)^T \mathbf{1} - 1\right)^2$, where the sum of the fractional proportions at each spatial location must sum one. Lastly, the loss function is composed of the sum of the single losses at each deep-block as follows

$$\begin{aligned} \{\boldsymbol{\theta}^*\} &\in \arg\min_{\boldsymbol{\theta}} \sum_k \tau_k \mathcal{L}^k\left(\mathbf{f}^k \mid \mathbf{y}, \mathbf{\Phi}\right), \\ &\qquad\qquad\qquad\qquad\qquad\qquad (12) \\ \mathcal{L}^k(\mathbf{f}^k \mid \mathbf{y}, \mathbf{\Phi}) &= F(\mathbf{f}^k|\mathbf{y}, \mathbf{\Phi}) + \gamma^k R(\mathcal{A}_\theta^k(\mathbf{f}^k)), \end{aligned}$$

where, $\mathbf{f}^k := \mathcal{M}_{\theta^k}(\mathbf{f}^{k-1})$, $\mathcal{M}_{\theta^k}(\cdot) : \mathbb{R}^{N^2 L} \to \mathbb{R}^{N^2 L}$ stands for the kth interpretable deep-block, $\tau_k > 0$ denotes the kth loss function relative weight, and $\gamma_k > 0$ denotes the kth regularization parameter that controls the trade-off between the data-fidelity term $F(\mathbf{f}^k|\mathbf{y}, \mathbf{\Phi})$ and the abundance sum-to-one constraint defined in (8). The output at each deep-block could be interpreted as the recovered SI; therefore, this paper studies two options (i) using the last deep-block output, i.e., $\hat{\mathbf{f}} := \mathbf{f}^K$ and (ii) using the average between the last two deep-blocks outputs, i.e., $\hat{\mathbf{f}} := (\mathbf{f}^K + \mathbf{f}^{K-1})/2$.

IV. SIMULATIONS AND RESULTS

The experiments to evaluate the Mixture-Net performance were conducted over six publicly available spectral datasets.

The Pavia University dataset[1], acquired with the ROSIS sensor spanning the $(0.43-0.86)$ μm spectral range. It contains 103 spectral bands, and 610×610 spatial pixels. A sub-region of 400×200 pixels is used following the setup in [32].

The hyperspectral Pavia Center dataset[1], acquired with the ROSIS sensor. It contains 102 spectral bands and 1096×1096 spatial pixels. **Remark:** Since only one image is adopted, the training process follows the splitting scheme in [33]. Precisely, the image's central region (1096×715) is cropped and divided into training and testing data. The left sub-region is extracted to form the testing data, with four non-overlapped HSIs with 224×224 pixels. The remaining region is extracted to form the training data with overlapped HSIs. Finally, the ten percent of the training data is included as a validation set.

The multispectral Stuff-Toys and Fake images taken from the CAVE dataset[2] that contains 32 images of everyday objects with 512×512 spatial pixels, and 31 spectral bands ranging from 400nm to 700nm at 10nm steps [34].

The KAIST[3] and the ARAD datasets[4] with 512×512 spatial pixels and 31 spectral bands [35], [36]. ARAD contains 480 images ranging from 400nm to 700nm at 10nm steps.

The Jasper Ridge dataset[1] with 512×614 spatial pixels and 224 spectral bands, ranging from 380nm to 2500nm at 9.46nm steps. Jasper Ridge contains four land classes: Road, Soil, Water, and Tree, [37]. A sub-region of $100 \times 100 \times 66$ spatial-spectral dimension aligned at the $(105, 269)^{th}$ location in the original image is used following the setup in [37].

Mixture-Net was tested in three SI recovery tasks, SI denoising, HSI-SR, and CSI reconstruction. For HSI-SR and CSI reconstruction, the ℓ_2-norm was used as the data fidelity term, given by $F(\mathbf{f}^k|\mathbf{y}, \mathbf{\Phi}) = \left\|\mathbf{y} - \mathbf{\Phi}\mathbf{f}^k\right\|_2^2$. For SI denoising, the state-of-the-art SURE loss was chosen to avoid over-fitting for high levels of noise, described in [32] as

$$F(\mathbf{f}^k|\mathbf{y}) = \quad \left\|\mathbf{y} - \mathbf{\Phi}\mathbf{f}^k\right\|_2^2 - \sigma^2 + \frac{2\sigma}{N^2 L}\mathrm{div}_{\mathbf{y}}(\mathbf{\Phi}\mathbf{f}^k), \quad (13)$$

where $\mathrm{div}_{\mathbf{y}}(\mathbf{\Phi}\mathbf{f}^k)$ is the divergence of $\mathbf{\Phi}\mathbf{f}^k := \mathbf{\Phi}\mathcal{M}_{\theta^k}(\mathbf{f}^0)$ computed with the Monte-Carlo SURE strategy in [38] as

$$\mathrm{div}_{\mathbf{y}}(\mathbf{\Phi}\mathbf{f}^k)) \approx \mathbf{b}^T\left(\frac{\mathbf{\Phi}(\mathcal{M}_{\theta^k}(\mathbf{f}^0 + \epsilon)) - \mathbf{\Phi}\mathcal{M}_{\theta^k}(\mathbf{f}^0)}{\epsilon}\right). \quad (14)$$

$\mathbf{b} \in \mathbb{R}^{N^2 L}$ is an i.i.d. Gaussian distribution with zero mean and unit variance. The SURE loss entails two hyperparameters ϵ and σ. ϵ is a small value with order 1×10^{-5}, and σ is the noise level calculated by the band-wise mean of the median absolute deviation estimator of the 2D wavelet transform [38].

The hyper-parameters τ_k and γ_k are tuned for each evaluated scenario following the sweeping methodology described in Supplementary Material, Section I.

The quality improvement is quantified through the spectral angle mapper (SAM), the root mean squared error (RMSE), the dimensionless global relative error of synthesis (ERGAS), the peak signal-to-noise ratio (PSNR), and the structural similarity (SSIM) metrics calculated as in [11], [28]. Simulations

---

[1]Available in Remote Sensing Datasets. Accessed: 09-Sep-2022.
[2]Available in Multispectral Image Dataset. Accessed: 09-Sep-2022.
[3]Available in KAIST Dataset. Accessed: 09-Sep-2022.
[4]Available in ARAD dataSet. Accessed: 09-Sep-2022.

were run on an Intel Xeon W-3223 with 64GB of memory, and an NVIDIA RTX 3090 GPU with 24GB of memory[5].

A. Characterization of Mixture-Net

This section studies the components and hyper-parameters tuning affecting Mixture-Net. (i) The input strategy, varying the network input structure. (ii) The abundance block-layer scheme, varying the number of layers and filters. (iii) The non-linearity block-layer, varying the rank. (iv) The number of deep-blocks K with single and multiple losses. The SI-SR task is carried out over Pavia Center to visualize the characterization results.

1) Input structure: This experiment evaluates the influence of learning the structure of the input $\mathbf{f}^0 \in \mathbb{R}^n$ against using a fixed input, i.e., the entries do not change during the learning process. Precisely, four fixed input strategies imposing a different structure were considered: A *Constant* input, using a tensor with all values equal to 0.5, i.e., $\boldsymbol{\mathcal{Z}} \in \{0.5\}^{N \times N \times L}$; A *Random* input, generating a random tensor with a normal Gaussian distribution; A *Mesh-grid* input, initializing the tensor as in [20]. And, an *Estimated* input, roughly estimating the image as $vec(\boldsymbol{\mathcal{Z}}) = \boldsymbol{\Phi}^T \mathbf{y}$. The proposed *Learned* strategy learns a low-dimensional Tucker Decomposition from random noise using $\rho = 0.4$ as described in Section III-A.

The input $\mathbf{f}^0$ is perturbed at each iteration of the learning process to emulate external noise, improving the results quality as shown in [20]. The perturbation is given by

$$\mathbf{f}^{(0)} = \text{vec}(\boldsymbol{\mathcal{Z}}) + \beta \boldsymbol{\eta}, \tag{15}$$

[5]The source code is publicly available in Mixture-Net code

Fig. 2: Quantitative performance box plot measured in *(left)* PSNR (purple) - SAM (green) and *(right)* SSIM (yellow) - RMSE (red) metrics for the five strategies imposing a structure in the input of Mixture-Net. The level of perturbation was varied across the values $[0, 1e^{-2}, 3e^{-2}, 5e^{-2}, 8e^{-2}, 1e^{-1}]$. The Abundance block-layer is a CNN with six layers using thirty two filters, and the hyper-parameters were set to $lr = 5e^{-3}$, $\lambda = 0$ (only the LMM), and $r = 8$.

where vec($\cdot$) denotes a vectorization operator, $\beta \in \mathbb{R}_+$ is a parameter controlling the perturbation level, and $\boldsymbol{\eta}_{i,j} \sim \mathcal{N}(0, 1)$ stands for Gaussian additive noise.

Figure 2 shows the Mixture-Net quality results varying the input structures. The *Constant* and *Random* inputs are far away from being good inputs, where even the *Mesh-grid* or *Estimated* strategies obtain a better quality. The proposed *Learned* input emerges as the best strategy, with a variance no greater than 0.47 [dBs] of PSNR and 0.27 degrees of SAM.

2) Abundance-block-layer: The abundance block-layer tuning evaluates three architectures to learn the spatial features leading to the estimated abundance maps. The *Convolutional* refers to a sequence of 2D convolutional layers with padding as in [39]. The *Auto-encoder* refers to a sequence of 2D convolutional layers, where the first half increase at the double at each feature, i.e., $[\ell, 2\ell, \ldots, 2\upsilon\ell]$, and the remaining half decrease $[2\upsilon\ell, (2\upsilon-1)\ell, \ldots, \ell]$ as in [35]. The *ResNet* refers to a sequence of residual neural layers, where the first and the last convolutional layers are concatenated as in [21]. Figure 3 shows the PSNR and SAM metrics across the architectures, varying the number of layers and the learned features per layer. It can be observed that the *Convolutional* architecture provides the highest quality, showing the most stable behavior with minimal variances when using thirty two features or more.

3) Non-linearity block-layer varying the rank: This experiment evaluates the effect of introducing the non-linearity block-layer and varying the rank. The rank can be addressed as a hyper-parameter related to the number of different materials in the SI. Figure 4 shows a quantitative comparison between Mixture-Net with just the LMM ($\lambda = 0$ in (11)) and Mixture-Net when the non-linearity block-layer is included, i.e., $\lambda > 0$. A clear improvement in the quality can be observed, especially for small values of the rank. This result indicates that using many endmembers compensates for the non-linear relations in the underlying scene. However, applications such as material identification should be aware of just a few ground truth endmembers that interact in non-linear ways to form the scene instead of using a large number of endmembers that do not match the spectral response of any material [29].

4) Number of interpretable deep-blocks with single and multiple losses: This experiment evaluates the influence of concatenating multiple interpretable deep-blocks and of using a single or multiple losses scheme. The single scheme includes the non-linearity block-layer and one loss only at the end of the whole network; meanwhile, the multiple scheme includes a loss at the end of each intermediate deep-block.

Mixture-Net is evaluated in two scenarios: (i) varying the number of interpretable deep-blocks and (ii) fixing five deep-blocks, calculating the quality of each intermediate deep-block output $\mathbf{f}^k$, interpreted as a recovered image.

Figure 5 *(left)* plots the resulting quality in PSNR as a function of the number of interpretable deep-blocks for the single and multiple schemes. An increasing relationship can be observed, where using multiple losses consistently achieves higher performance than using a single loss. Figure 5 *(right)* plots the quality at the intermediate deep-blocks when fixing a total number of five deep-blocks. The multiple scheme obtains

a remarkably higher performance than the single loss scheme from the first intermediate deep-block.

In both scenarios, the quality obtained with two deep-blocks is comparable to the quality when using three, four, and five deep-blocks. Therefore, the experiments in the rest of the manuscript employ two deep-blocks, a trade-off between the recovered image quality and the method complexity.

B. Mixture-Net Performance for Spectral Image Recovery

This section compares the Mixture-Met performance against state-of-the-art methods for denoising, HSI-SR, and CSI.

1) Denoising: A SI can be corrupted by different types of noise such as additive Gaussian, Poisson, and pepper [41]. This experiment evaluates the case of additive Gaussian noise over Pavia University. Mixture-Net uses the SURE loss to avoid over-fitting in the learning process. The comparison methods cover the block-matching and 3D filtering denoiser (BM3D) [42], the first order spectral roughness penalty

Fig. 3: Quantitative performance in terms of PSNR *(left)* and SAM *(right)* across three architectures. The hyper-parameters are varied as follows: the number of layers varies in the range $[1-5]$ for the *Convolutional* and *Res-Net* architectures, and in the range $[3, 5, 7]$ for the *Auto-encoder*; and the number of features per layer varies in the range $[8, 16, 32, 64]$. The best learning-rate was found to be $lr = 1e^{-3}$, and the noise perturbation of the input was set to be $\beta = 0.7$.

for denoising (FORPDN) [12], the hyperspectral restoration (HyRes) [14], the hyperspectral-DIP [23], the SURE-CNN [35], and the unsupervised disentangled spatio-spectral deep priors method (DS2DP) [25].

Table I summarizes the quantitative results regarding PSNR and SSIM metrics, where the Input column refers to the noisy image quality. Mixture-Net obtains a significant improvement, especially in the SSIM metric, indicating that the recovered SI intrinsic structures with Mixture-Net are improved over the state-of-the-art methods as observed in Fig. 6. Further, the zoomed version visualizes a noise reduction in smooth regions.

2) Single Hyperspectral Super-Resolution: This experiment employs the Pavia Center and Stuff-Toys datasets for two downsampling factors, $d = [4, 8]$. The comparison methods cover the bicubic interpolation; three gray/RGB image SR methods, EDSR [43], RCAN [44], and SAN [45]; three single HSI-SR methods, 3DCNN [46], GDRRN [47], and SSPSR [33]; and the non-data-driven Hyperspectral-DIP method [23]. The optimization uses the Adam algorithm with $lr = 1e^{-3}$ and $\gamma^k = 0.5$. The rank r and the number of deep-blocks K were set to $r = 6$ and $K = 3$, and $r = 12$ and $K = 4$, for Pavia Center and Stuff-Toys, respectively.

Tables II and III compare the quantitative results for the single HSI-SR over the evaluated methods. Our non-data-driven Mixture-Net outperforms or achieves competitive quality even against the data-driven methods such as the SSPSR for both datasets, both downsampling factors, and all evaluated metrics.

Fig. 4: Quantitative performance in terms of PSNR *(left)* and SAM *(right)* when using the linear block-layer, i.e., $\lambda = 0$ in (11), and when including the non-linearity block-layer with $\lambda = 0.7$. The rank value varies in the range $[3, 4, 5, 6, 7, 8, 9, 10, 15, 20]$. The non-linearity block-layer is composed of two consecutive spatial-spectral networks presented in [40], with the learning rate set to $lr = 1e^{-3}$.

TABLE I: Denoising Quantitative Results for Pavia University

Noise σ	Metric	Input	BM3D	FORPDN	HyRes	DIP	SURE CNN	DS2DP	Mixture -Net
$\frac{100}{255}$	↑PSNR	8.130	29.14	26.03	28.41	26.47	<u>29.62</u>	27.54	**30.95**
	↓SSIM	0.025	0.754	0.597	0.738	0.683	<u>0.802</u>	0.718	**0.869**
$\frac{50}{255}$	↑PSNR	14.15	32.97	30.44	31.78	30.69	<u>33.29</u>	32.12	**34.47**
	↓SSIM	0.102	0.881	0.799	0.855	0.846	<u>0.905</u>	0.896	**0.942**
$\frac{25}{255}$	↑PSNR	20.17	<u>36.48</u>	34.34	35.35	34.48	36.09	35.55	**36.99**
	↓SSIM	0.286	0.942	0.906	0.927	0.917	0.945	<u>0.951</u>	**0.965**

Fig. 5: (*left*) Mixture-Net performance, varying the number of interpretable deep-blocks in the range $[1-5]$ for single and multiple losses. (*right*) Recovered image quality at each intermediate deep-block when fixing five deep-blocks.

Fig. 6: Visual RGB mapping of denoised Pavia University with noise level $\sigma = 50/255$. The white box at the bottom of each image shows the quantitative quality in PSNR. The zoomed sub-region shows that Mixture-Net suppresses the Gaussian noise in smooth regions while maintaining the shapes and structures. Unlike, the noise is still visible in HSI-DIP, SURE, and DS2DP methods, and the edges and structure are blurred in the BM3D case. A spectral quality comparison is shown at the bottom for two spatial pixels **P1** and **P2**.

The spectral quality measured with the SAM metric results remarkably improved for the Pavia Center dataset. Thus, the intrinsic low-rank prior is more substantial when a higher number of correlated spectral bands are considered. Figure 7 shows a visual comparison of the RGB mapping of some super-resolved images for the Pavia Center dataset, where Mixture-Net improves the spatial quality, particularly for high down-sampling factors.

3) Compressive Spectral Imaging Reconstruction for KAIST and ARAD datasets: This experiment compares the Mixture-Net effectiveness for CSI reconstruction against data-driven and non-data-driven state-of-the-art methods, including the

TABLE II: Single Hyperspectral Super-Resolution Quantitative Results for Pavia Center

Method	d	SAM↓	RMSE↓	ERGAS↓	PSNR↑	SSIM ↑
Bicubic	4	6.1399	0.0437	6.8814	27.5874	0.6961
EDSR	4	5.8657	0.0379	6.0199	28.7981	0.7722
RCAN	4	5.9785	0.0376	6.0485	28.8165	0.7719
SAN	4	5.9590	0.0374	5.9903	28.8554	0.7740
3DCNN	4	5.8669	0.0396	6.2665	28.4114	0.7501
GDRRN	4	5.4750	0.0393	6.2264	28.4726	0.7530
SSPSR	4	<u>5.4612</u>	<u>0.0362</u>	**5.8014**	<u>29.1581</u>	<u>0.7903</u>
DIP	4	6.2665	0.0410	6.4845	28.1061	0.7365
Mixture-Net	4	**4.2120**	**0.0352**	<u>5.8084</u>	**29.914**	**0.8396**
Bicubic	8	7.8478	0.0630	4.8280	24.5972	0.4725
EDSR	8	7.8594	0.05983	4.6359	25.0041	0.5130
RCAN	8	7.9992	0.0604	4.6930	24.9183	0.5086
SAN	8	8.0371	0.0609	4.7646	24.8485	0.5054
3DCNN	8	7.6878	0.0605	4.6469	24.9336	0.5038
GDRRN	8	7.3531	0.0607	4.6220	24.8648	0.5014
SSPSR	8	<u>7.3312</u>	<u>0.0586</u>	<u>4.5266</u>	<u>25.1985</u>	<u>0.5365</u>
DIP	8	7.9281	0.0618	4.7366	24.7252	0.4963
Mixture-Net	8	**6.7855**	**0.0485**	**4.0015**	**26.9041**	**0.7148**

TABLE III: Single Hyperspectral Super-Resolution Quantitative Results for Stuff-Toys

Method	d	SAM↓	RMSE↓	ERGAS↓	PSNR↑	SSIM ↑
Bicubic	4	4.1759	0.0212	5.2719	34.7214	0.9277
EDSR	4	3.5499	0.0149	3.5921	38.1575	0.9522
RCAN	4	3.6050	0.0142	3.4178	38.7585	0.9530
SAN	4	3.5951	0.0143	3.4200	38.7188	0.9531
3DCNN	4	3.3463	0.0154	3.7042	37.9759	0.9522
GDRRN	4	<u>3.4143</u>	0.0145	3.5086	38.4507	0.9538
SSPSR	4	**3.1846**	0.0138	3.3384	<u>39.0892</u>	0.9553
DIP	4	8.4935	<u>0.0124</u>	<u>2.5358</u>	38.1329	<u>0.9631</u>
Mixture-Net	4	5.3285	**0.0110**	**2.1997**	**39.1640**	**0.9821**
Bicubic	8	5.8962	0.0346	4.2175	30.2056	0.8526
EDSR	8	5.6865	0.0279	3.3903	32.4072	0.8842
RCAN	8	5.9771	0.0268	3.1781	32.9544	0.8884
SAN	8	5.8683	0.0267	3.1437	33.0012	0.8888
3DCNN	8	<u>5.0948</u>	0.0292	3.5536	31.9691	0.8863
GDRRN	8	5.3597	0.0280	3.3460	32.5763	0.8890
SSPSR	8	**4.4874**	0.0257	3.0419	<u>33.4340</u>	0.9010
DIP	8	8.3342	<u>0.0231</u>	<u>2.3697</u>	32.7324	<u>0.9291</u>
Mixture-Net	8	5.5027	**0.0208**	**2.1061**	**33.6270**	**0.9432**

non-data-driven Plug-and-Play (PnP) [48] and TL-DIP [21], and the data-driven Deep Non-local Unrolling (DNU) [17], Autoencoder (AE) [35], and Joint non-linear Representation and Recovery Network (JR2Net) [49]. The CSI reconstruction is carried out over the KAIST and ARAD datasets, where the DNU and AE methods were trained with the publicly available codes using a training dataset of 27 and 450 images for the

Fig. 7: RGB representation of the reconstructed composite images of Pavia Center datatest for spatial downsampling factor $d = 4$ (*top*) and $d = 8$ (*bottom*). Notice that, the spatial quality is especially improved for a factor $d = 8$.

Fig. 8: Compressive sensing reconstruction quality comparison using the non-data-driven PnP and DIP methods, and the training-data dependent DNU, and AE methods against the interpretable method Mixture-Net for the KAIST and ARAD datasets. The quality is measured in terms of PSNR / SSIM.

TABLE IV: Compressive Spectral Imaging Reconstruction Quantitative Results

	KAIST					
Metric	PnP	DNU	AE	TL-DIP	JR2net	Mixture-Net
↑PSNR	29.8 ± 2.6	32.6 ± 1.8	37.9 ± 1.1	32.8 ± 0.9	$\underline{39.1 \pm 2.5}$	$\mathbf{40.3 \pm 2.6}$
↑SSIM	0.92 ± 0.02	0.91 ± 0.01	0.96 ± 0.01	0.92 ± 0.01	$\underline{0.97 \pm 0.02}$	$\mathbf{0.99 \pm 0.01}$
↓SAM	10.9 ± 4.2	12.9 ± 1.8	9.20 ± 0.4	12.4 ± 2.7	$\underline{6.00 \pm 2.3}$	$\mathbf{3.26 \pm 1.1}$
	ARAD					
↑PSNR	31.5 ± 4.5	29.5 ± 4.5	33.5 ± 4.4	35.8 ± 3.8	$\underline{36.5 \pm 4.6}$	$\mathbf{39.1 \pm 4.1}$
↑SSIM	0.83 ± 0.08	0.82 ± 0.09	0.90 ± 0.05	0.95 ± 0.03	$\underline{0.97 \pm 0.01}$	$\mathbf{0.98 \pm 0.01}$
↓SAM	6.94 ± 2.7	8.26 ± 3.1	5.50 ± 1.7	4.45 ± 2.5	$\underline{2.85 \pm 0.95}$	$\mathbf{2.04 \pm 0.53}$

KAIST and ARAD datasets, respectively.

Table IV summarizes the average reconstruction quality across all testing images for each database with the standard deviation. For all metrics, Mixture-Net outperforms state-of-the-art methods, showing the effectiveness to recover the spatial and spectral details. Figure 8 visualizes the recovered images with the reconstruction quality measured in PSNR and SSIM. Mixture-Net improves the quality of reconstruction by up to 7 [dB] and obtains the highest SSIM, outperforming even the methods employing data during the training step. Figure 9 shows the recovered spectral signatures at a random spatial location. The absolute errors at right confirm that Mixture-Net obtains the best response estimation for both databases.

C. Remote Sensing Tasks Experiments

This experiment aims to evaluate the potential application of Mixture-Net to perform remote sensing tasks such as unmixing and material identification without running complementary routines. For this, Mixture-Net takes advantage of its architecture's interpretability, providing two outputs interpreted as the abundances and endmembers.

The unmixing experiment is carried out using one single shot of the Dual Disperser Coded Aperture Snapshot Spectral

Fig. 9: Spectral signatures comparison at P1 for each dataset. The absolute error plots confirm that the spectral signatures obtained by the Mixture-Net method are more accurate than those obtained by the comparison methods.

Fig. 11: *(left)* RGB Ground truth and estimated Fake image. *(Top right)* Two obtained features interpreted as the abundances. *(Bottom right)* Binary maps after applying a threshold, identifying the differences in the materials.

Imaging (DD-CASSI) [50] over the Jasper Dataset. The hyperparameters are the same as the experiment in Section IV-A2, with $\lambda = 0$, i.e., considering only the linear component. Figure 10 depicts a false RGB representation of Jasper's ground-truth, the Mixture-Net estimation, and the learned features and adjusted weights. The learned features in the abundance block-layer and the adjusted weights in the endmember block-layer converge to a rough estimation of the abundance maps and endmembers, demonstrating that Mixture-Net could be used for the unmixing problem at any additional cost.

The material identification experiment is carried out using a single shot of the DD-CASSI and the Fake image containing a real and a fake plastic fruit with similar shape and color. The image is spatially sub-sampled to 256×256 pixels. Figure 11 shows the ground truth and a false RGB mapping of the Mixture-Net estimation with $r = 15$. At the right top, the figure shows two obtained features that can be interpreted as abundances. Then, thresholding is applied over the abundances obtaining the binary maps at the right bottom to determine if the materials of both objects are the same. The color checker shows that the obtained abundance can determine that the materials are different, identifying the fake object without running additional routines.

D. Real-Data Experiment

This experiment evaluates the proposed Mixture-Net for the super-resolution task with real-data, acquired in the Optics

Fig. 10: Unmixing remote sensing task application resulting from the interpretable learned features and adjusted weights of Mixture-Net. RGB representation of the ground truth and the estimated recovered image, endmember, and abundance matrices.

Laboratory at Universidad Industrial de Santander with a test-bed composed of a SI and an RGB arm.

The SI arm is composed of an adjustable mechanical slit (Thorlabs VA100C-30 mm, 8-32 Tap) in which the 4f system focuses the scene. The entrance slit has a height of 13.59mm and a width of $20\mu m$. Then, a relay lens at 100 mm from the slit forms a parallel beam that reaches the 600 grooves/mm transmission grating that diffracts the light rays onto the sensor. The push-broom imagery spectrogram illuminates the slit with 1032 wavelengths in the spectral range $420 - 700$nm with steps of 2.7nm. The SI acquisition performs 50 horizontal steps with a resolution of $48\mu m$, so that the acquired low-resolution spectral image, referred to as LR-SI, has 100×50 spatial pixels and 1032 spectral bands.

The RGB arm is composed of an RGB camera, where the 4f system focuses the scene onto its detector array, so we employ a sub-region of 1200×600 spatial pixels as a reference RGB image with high spatial resolution, referred to as HR-RGB.

Figure 12 (left) compares the super-resolved image, referred to as HR-SI, against the LR-SI and the HR-RGB. The spatial structure of the HR-SI results comparable to the observed in the HR-RGB, and the HR-SI spectral signatures are estimated accordingly to the LR-SI. In particular, the HR-SI contains

smoother spectral signatures in comparison to the LR-SI.

Figure 12 (right) illustrates the Mixture-Net interpretability, where the depicted learned features match the features of a set of abundance maps, and the plotted adjusted weights follow the behavior of a set of endmembers. Therefore, besides super-resolving the image, Mixture-Net provides a set of abundance and endmembers maps that can be used for unmixing or material identification purposes.

V. CONCLUSIONS

This paper proposes a non-data-driven spectral image recovery method based on deep image prior, where the low-rank mixture models inspire the network dubbed Mixture-Net. Beyond previous methods, Mixture-Net includes prior knowledge implicitly in the architecture, addressing the black-box nature of standard deep learning. Mixture-Net is divided into three components, the input, imposing a low-dimensional structure through the learning of a Tucker decomposition; the interpretable Mixture-Net, following a sequence of multiple deep-blocks to estimate the abundances, the endmembers, and the spectral image with non-linear relations; and the custom loss function that considers a regularization related to physical constraints. The non-data-driven Mixture-Net effectiveness was demonstrated over three spectral image recovery problems: spectral image denoising, single hyperspectral super-resolution, and compressive spectral imaging reconstruction in terms of different metrics, outperforming even data-driven methods requiring the training of a vast amount of data. Along with the experiments, we found that imposing a low-dimensional structure over the input improves the quality of the recovered image. On the other hand, the non-linearity block-layer drastically improved the obtained quality by considering the non-linear relationships between the endmembers. We remark that Mixture-Net can be extended to any other spectral image recovery task, even using different loss functions, where the interpretable advantage of Mixture-Net allows to execute further remote sensing tasks as linear unmixing and material identification without using additional routines.

Fig. 12: Real-data super-resolution.*(Left-top)* Spatial comparison of the estimated super-resolved image (HR-SI) against the reference RGB (HR-RGB) and the acquired SI (LR-SI).*(Left-bottom)* LR-SI spectral signatures at two spatial locations $P_1 = (42, 32)$ and $P_2 = (62, 8)$ compared against the estimated HR-SI spectral signatures at the corresponding locations $P_1 = (493, 373)$ and $P_2 = (733, 85)$.*(Right)* Mixture-Net interpretability analysis. The learned features and adjusted weights appear to be the abundance maps and the endmembers spanning the image. In particular, notice that the pairs $(\mathbf{a}_1, \mathbf{e}_1)$, $(\mathbf{a}_6, \mathbf{e}_6)$, $(\mathbf{a}_8, \mathbf{e}_8)$ can be interpreted as the maps and spectral signatures of the red, yellow, and green pixels.

REFERENCES

[1] H. Vargas, A. Camacho, and H. Arguello, "Spectral unmixing approach in hyperspectral remote sensing: a tool for oil palm mapping," *TecnoLógicas*, vol. 22, no. 45, pp. 131–145, 2019.

[2] J. M. Amigo and C. Santos, "Preprocessing of hyperspectral and multispectral images," in *Data Handling in Science and Technology*. Elsevier, 2020, vol. 32, pp. 37–53.

[3] C. Cao, J. Yu, C. Zhou, K. Hu, F. Xiao, and X. Gao, "Hyperspectral image denoising via subspace-based nonlocal low-rank and sparse factorization," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 3, pp. 973–988, 2019.

[4] J. Xue, Y. Zhao, W. Liao, and J. C.-W. Chan, "Nonlocal tensor sparse representation and low-rank regularization for hyperspectral image compressive sensing reconstruction," *Remote Sensing*, vol. 11, no. 2, p. 193, 2019.

[5] C. Guilloteau, T. Oberlin, O. Berné, and N. Dobigeon, "Hyperspectral and multispectral image fusion under spectrally varying spatial blurs–application to high dimensional infrared astronomical imaging," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1362–1374, 2020.

[6] L. Zhang, J. Nie, W. Wei, Y. Li, and Y. Zhang, "Deep blind hyperspectral image super-resolution," *IEEE Transactions on Neural Networks and Learning Systems*, 2020.

[7] J. Bacca, T. Gelvez, and H. Arguello, "Deep coded aperture design: An end-to-end approach for computational imaging tasks," *IEEE Transactions on Computational Imaging*, pp. 1–1, 2021.

[8] O. F. Kar and F. S. Oktem, "Compressive spectral imaging with diffractive lenses," *Optics letters*, vol. 44, no. 18, pp. 4582–4585, 2019.

[9] T. Blumensath, M. E. Davies, G. Rilling, Y. Eldar, and G. Kutyniok, "Greedy algorithms for compressed sensing," 2012.

[10] H. U. Mullah, B. Deka, and A. Prasad, "Fast multi-spectral image super-resolution via sparse representation," *IET Image Processing*, vol. 14, no. 12, pp. 2833–2844, 2020.

[11] T. Gelvez, H. Rueda, and H. Arguello, "Joint sparse and low rank recovery algorithm for compressive hyperspectral imaging," *Applied optics*, vol. 56, no. 24, pp. 6785–6795, 2017.

[12] B. Rasti, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "Hyperspectral image denoising using first order spectral roughness penalty in wavelet domain," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2458–2467, 2013.

[13] T. Gelvez-Barrera, H. Arguello, and A. Foi, "Joint nonlocal, spectral, and similarity low-rank priors for hyperspectral-multispectral image fusion," *IEEE Transactions on Geoscience and Remote Sensing*, 2022.

[14] B. Rasti, M. O. Ulfarsson, and P. Ghamisi, "Automatic hyperspectral image restoration using sparse and low-rank modeling," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 12, pp. 2335–2339, 2017.

[15] J. Bacca, C. V. Correa, and H. Arguello, "Noniterative hyperspectral image reconstruction from compressive fused measurements," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 4, pp. 1231–1239, 2019.

[16] R. Dian, S. Li, L. Fang, and J. Bioucas-Dias, "Hyperspectral image super-resolution via local low-rank and sparse representations," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2018, pp. 4003–4006.

[17] L. Wang, C. Sun, M. Zhang, Y. Fu, and H. Huang, "Dnu: Deep non-local unrolling for computational spectral imaging," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 1661–1671.

[18] J. M. Ramírez, J. I. M. Torre, and H. A. Fuentes, "Ladmm-net: An unrolled deep network for spectral image fusion from compressive data," *arXiv preprint arXiv:2103.00940*, 2021.

[19] H. Vu, G. Cheung, and Y. C. Eldar, "Unrolling of deep graph total variation for image denoising," *arXiv preprint arXiv:2010.11290*, 2020.

[20] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9446–9454.

[21] J. Bacca, Y. Fonseca, and H. Arguello, "Compressive spectral image reconstruction using deep prior and low-rank tensor representation," *Applied optics*, vol. 60, no. 14, pp. 4197–4207, 2021.

[22] H. Zhang, L. Liu, W. He, and L. Zhang, "Hyperspectral image denoising with total variation regularization and nonlocal low-rank tensor decomposition," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 5, pp. 3071–3084, 2019.

[23] O. Sidorov and J. Yngve Hardeberg, "Deep hyperspectral prior: Single-image denoising, inpainting, super-resolution," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2019, pp. 0–0.

[24] B. Rasti, B. Koirala, P. Scheunders, and P. Ghamisi, "Undip: Hyperspectral unmixing using deep image prior," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2021.

[25] Y.-C. Miao, X.-L. Zhao, X. Fu, J.-L. Wang, and Y.-B. Zheng, "Hyperspectral denoising using unsupervised disentangled spatio-spectral deep priors," *arXiv preprint arXiv:2102.12310*, 2021.

[26] T. Gelvez, J. Bacca, and H. Arguello, "Interpretable deep image prior method inspired in linear mixture model for compressed spectral image recovery," in *2021 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2021, pp. 1934–1938.

[27] B. Rasti, B. Koirala, P. Scheunders, and J. Chanussot, "Misicnet: Minimum simplex convolutional network for deep hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, 2022.

[28] T. Gelvez and H. Arguello, "Nonlocal low-rank abundance prior for compressive spectral image fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 1, pp. 415–425, 2020.

[29] 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 journal of selected topics in applied earth observations and remote sensing*, vol. 5, no. 2, pp. 354–379, 2012.

[30] M. Wang, M. Zhao, J. Chen, and S. Rahardja, "Nonlinear unmixing of hyperspectral data via deep autoencoder networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 9, pp. 1467–1471, 2019.

[31] L. Huang, R. Luo, X. Liu, and X. Hao, "Spectral imaging with deep learning," *Light: Science & Applications*, vol. 11, no. 1, pp. 1–19, 2022.

[32] H. V. Nguyen, M. O. Ulfarsson, and J. R. Sveinsson, "Sure based convolutional neural networks for hyperspectral image denoising," in *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2020, pp. 1784–1787.

[33] J. Jiang, H. Sun, X. Liu, and J. Ma, "Learning spatial-spectral prior for super-resolution of hyperspectral imagery," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1082–1096, 2020.

[34] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum," *IEEE transactions on image processing*, vol. 19, no. 9, pp. 2241–2253, 2010.

[35] I. Choi, D. S. Jeon, G. Nam, D. Gutierrez, and M. H. Kim, "High-quality hyperspectral reconstruction using a spectral prior," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 6, pp. 1–13, 2017.

[36] B. Arad, R. Timofte, O. Ben-Shahar, Y.-T. Lin, and G. D. Finlayson, "Ntire 2020 challenge on spectral reconstruction from an rgb image," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 446–447.

[37] F. Zhu, Y. Wang, S. Xiang, B. Fan, and C. Pan, "Structured sparse method for hyperspectral unmixing," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 88, pp. 101–118, 2014.

[38] S. Ramani, T. Blu, and M. Unser, "Monte-carlo sure: A black-box optimization of regularization parameters for general denoising algorithms," *IEEE Transactions on image processing*, vol. 17, no. 9, pp. 1540–1554, 2008.

[39] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[40] L. Wang, C. Sun, Y. Fu, M. H. Kim, and H. Huang, "Hyperspectral image reconstruction using a deep spatial-spectral prior," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8032–8041.

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

[42] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on image processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[43] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 136–144.

[44] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 286–301.

[45] T. Dai, J. Cai, Y. Zhang, S.-T. Xia, and L. Zhang, "Second-order attention network for single image super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11 065–11 074.

[46] S. Mei, X. Yuan, J. Ji, Y. Zhang, S. Wan, and Q. Du, "Hyperspectral image spatial super-resolution via 3d full convolutional neural network," *Remote Sensing*, vol. 9, no. 11, p. 1139, 2017.

[47] Y. Li, L. Zhang, C. Dingl, W. Wei, and Y. Zhang, "Single hyperspectral image super-resolution with grouped deep recursive residual network," in *2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM)*. IEEE, 2018, pp. 1–4.

[48] X. Yuan, Y. Liu, J. Suo, and Q. Dai, "Plug-and-play algorithms for large-scale snapshot compressive imaging," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 1447–1457.

[49] B. Monroy, J. Bacca, and H. Arguello, "Jr2net: a joint non-linear representation and recovery network for compressive spectral imaging," *Appl. Opt.*, vol. 61, no. 26, pp. 7757–7766, Sep 2022. [Online]. Available: https://opg.optica.org/ao/abstract.cfm?URI=ao-61-26-7757

[50] M. Gehm, R. John, D. Brady, R. Willett, and T. Schulz, "Single-shot compressive spectral imaging with a dual-disperser architecture," *Optics express*, vol. 15, no. 21, pp. 14 013–14 027, 2007.

Tatiana Gelvez (S'17) received the B.S. degree in industrial engineering and systems engineering from the Universidad Industrial de Santander, Bucaramanga, Colombia, in 2016, where she is currently pursuing the Ph.D. degree with the Department of Electrical Engineering. During the second semester of 2019 and 2020 she was an intern at Tampere University, Tampere, Finland. Her research interests include numerical optimization, high-dimensional signal processing, spectral imaging, and compressive sensing.

Jorge Bacca (S'17-M'17) received the B.S. degree in computer science and the Ph.D. degree in Computer Science from the Universidad Industrial de Santander, Bucaramanga, Colombia in 2017 and 2021, respectively. His current research interests include inverse problem, deep learning methods, optical imagining, compressive sensing, phase retrieval, hyperspectral imaging and numerical optimization.

Henry Arguello (S'11–M'13–SM'17) received the B.Sc. Eng. degree in electrical engineering and the M.Sc. degree in electrical power from the Universidad Industrial de Santander, Bucaramanga, Colombia, in 2000 and 2003, respectively, and the Ph.D. degree in electrical engineering from the University of Delaware, Newark, DE, USA, in 2013. He is currently an Associate Professor with the Department of Systems Engineering, Universidad Industrial de Santander. In first semester 2020, he was a Visiting Professor with Stanford University, Stanford, CA, USA, funded by Fulbright. His research interests include high-dimensional signal processing, optical imaging, compressed sensing, hyperspectral imaging, and computational imaging.