

SPACNN-LDVAE: SPATIAL ATTENTION CONVOLUTIONAL LATENT DIRICHLET VARIATIONAL AUTOENCODER FOR HYPERSPECTRAL PIXEL UNMIXING

Soham Chitnis*

Kiran Mantripragada, Faisal Z. Qureshi

Birla Institute of Technology and Science Pilani
Department of CS & IS
Goa, India

Ontario Tech University
Faculty of Science
Oshawa, Canada

ABSTRACT

The hyperspectral pixel unmixing aims to find the underlying materials (endmembers) and their proportions (abundances) in pixels of a hyperspectral image. This work extends the Latent Dirichlet Variational Autoencoder (LDVAE) pixel unmixing scheme by taking into account local spatial context while performing pixel unmixing. The proposed method uses an isotropic convolutional neural network with spatial attention to encode pixels as a dirichlet distribution over endmembers. We have evaluated our model on Samson, Hydice Urban, Cuprite, and OnTech-HSI-Syn-21 datasets. Our model also leverages the transfer learning paradigm for Cuprite Dataset, where we train the model on synthetic data and evaluate it on the real-world data. The results suggest that incorporating spatial context improves both endmember extraction and abundance estimation.

Index Terms— Hyperspectral image analysis, Unmixing, Endmember Extraction, Abundance Estimation, Variational Autoencoder, Deep Learning, Spatial Attention Convolution

1. INTRODUCTION

Hyperspectral images can be applied in a variety of applications and are very prominent in the field of remote sensing. They are known for having lower spatial resolution but very high spectral resolution. The lower spatial resolution implies that a single pixel covers a large region of space [1]. Such a phenomenon is mostly attributed to the images collected from high altitudes. The pixel unmixing task is to infer the materials with their mixing ratios. This problem is of particular interest to the hyperspectral image analysis community. In literature, it is sometimes referred to as *spectral unmixing* [2]. It involves recovering the “pure” spectra of materials known as endmembers with the ratios of mixing known as abundances.

Existing approaches can usually be divided into two classes: physics-based or data-driven methods. Physics-based methods use phenomenological models for the radiance response of the materials [1, 3]. These are costly to

implement, since it is often tedious to develop accurate radiance response models. On the other hand, data-driven methods are easier to use in practice; however, these require access to data for model setup and training. It is not always straightforward to collect this data. Some examples of data-driven techniques are Blind Source Separation (BSS), Non-Negative Matrix Factorization (NMF), Principle Component Analysis (PCA), and Linear Discriminant Analysis (LDA) [4, 5, 6, 7]. A number of commonly used approaches, such as N-FINDR, Pixel Purity Index (PPI), and Vector Component Analysis (VCA), first extract endmembers and then estimate abundances [8, 9, 10] by leveraging abundances-sum-to-one (ASC) and abundances-non-negative (ANC) constraints [11].

Variants of (Linear) Mixing Models (LMM) include Perturbed LMM, Extended LMM, Non-linear LMM, and Data-driven LMM [12, 13, 14, 15, 16]. Many NMF methods—SSWNMF, SGSNMF, TV-RSNMF, GLNMF, $L_{1/2}$ NMF—use regularization achieve better performance [17, 18, 19, 20, 21]. Recently, a number of deep learning approaches have been proposed to address the problem of hyperspectral pixel unmixing—EACNN, DAEN, DeepGUn, and TANet [22, 23, 24, 25].

Data-driven methods can further be divided into supervised and unsupervised schemes. Unsupervised schemes, e.g., BSS, do not require access to labeled data for model setup and training. Supervised schemes, such as EACNN, on the other hand, require access to labeled data for training. The method proposed here requires access to labeled data. It is, however, possible to apply this method in scenarios where labeled data is not available by training the model on synthetic data and leveraging *transfer learning*.

This paper builds on the work of the Latent Dirichlet Variation Autoencoder (LDVAE), which assumes that the endmember spectrum can be represented using a multivariate Normal Distribution and the mixing ratios (abundances) can be represented using Dirichlet distribution [26]. LDVAE is a Multilayer Perceptron model (MLP), and it does not incorporate spatial information. The approach presented in this paper aims to exploit spatial coherence, i.e., nearby pix-

*Work done in MITACS Globalink Research Internship

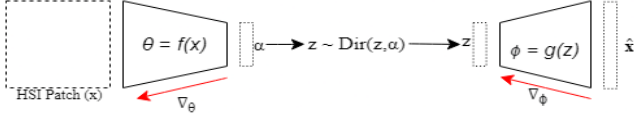


Fig. 1: CNN Latent Dirichlet Variational Autoencoder

els often contain the “same” endmembers and have “similar” abundances. The encoder stage uses Isotropic CNN layers plus a spatial attention layer to capture the local spatial structure around a pixel as its latent representation is constructed. Subsequently, similar to LDVAE, the decoder uses this latent representation for endmember extraction and abundance estimation. In the interest of clarity, hereafter, we will refer to the original LDVAE approach as MLP-LDVAE [26].

2. METHODOLOGY

Latent Dirichlet Allocation (LDA) [27], a popular scheme in natural language processing, inspired MLP-LDVAE. LDA assumes a collection of documents, where each document is a mixture of a set of topics. Both the topics and their mixing ratios within each document is not known *a priori*. This is similar to the problem of hyperspectral pixel unmixing. A hyperspectral image is composed of pixels, and each pixel is a mixture of endmembers. Both the endmembers and the abundances, i.e., the mixing ratios of endmembers, in any pixel are not known. In the parlance of LDA, endmember extraction is akin to topic discovery and abundance estimation is similar to computing the mixing ratio of topics in a particular document.

Our model uses the VAE architecture depicted by Figure 1. LDVAE is a Variational Autoencoder where the latent representation follows a Dirichlet distribution. The encoder is parameterized by θ , which outputs the Dirichlet distribution parameter α . LDVAE takes a single signal of the pixel. We now try to leverage spatial information in the encoder using a CNN.

2.1. Spatial Attention Convolutional Neural Network Encoder

Our model uses a CNN encoder (Figure 2), which receives a rectangular patch as input and returns the Dirichlet distribution parameter α corresponding to the center pixel. The abundances \mathbf{z} are sampled from the Dirichlet distribution and fed into the decoder, which reconstructs the spectral signal of the middle pixel $\hat{\mathbf{x}}$. The decoder follows the model used in [26].

The encoder employs an isotropic CNN model; therefore, the spatial resolution is maintained. The CNN encoder comprises of three modules: (1) stem, (2) body and (3) the spatial attention branch. The stem consists of a 2D convolution

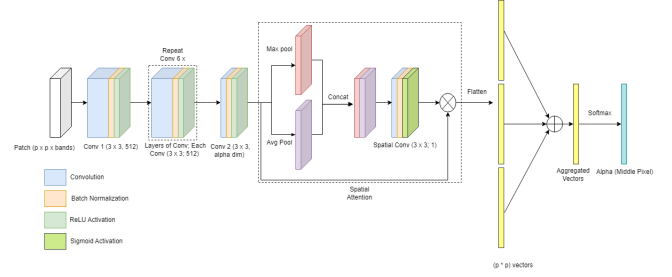


Fig. 2: Spatial Attention Convolutional Neural Network Encoder

layer (kernel size: 3 and padding: 1) with Batch Normalization (BN) and Rectified Linear Unit (ReLU) activation. The body consists of six blocks of convolution layer followed by BN and ReLU. The spatial attention block follows the model introduced in [28]. The output features are aggregated to generate the Dirichlet parameter α .

Given the intermediate feature map $\mathbf{F} \in \mathbb{R}^{H \times W \times C}$, apply average and max pooling along the channel dimension. Concatenate the results and perform 2D convolution with kernel size of 3 and apply sigmoid activation to obtain the 2D spatial attention map

$$\mathbf{A} = \sigma(f^{3 \times 3}(\text{AvgPool}_C(\mathbf{F}) \oplus \text{MaxPool}_C(\mathbf{F}))), \quad (1)$$

where σ denotes the sigmoid function, $f^{3 \times 3}$ denotes the convolution operation with a 3×3 filter, \oplus denotes the concatenation operation, and $\mathbf{A} \in \mathbb{R}^{H \times W}$. The latent representation $\mathbf{z}' \in \mathbb{R}^C$ is computed as follows:

$$\mathbf{z}' = \sum_{i=1}^H \sum_{j=1}^W \mathbf{A}_{i,j} \mathbf{F}_{i,j}. \quad (2)$$

Softmax activation is applied to obtain α in the final layer of encoder as follows:

$$\alpha_k = \frac{e^{z_k}}{\sum_{k=1}^C e^{z_k}}. \quad (3)$$

This ensures that the model satisfies ASC and ANC.

2.2. Spectral Reconstruction With Multivariate Normal Distribution

As stated previously, we employ the decoder used in MLP-LDVAE, which uses an MLP to reconstructs the spectrum given abundances \mathbf{z} . The decoder serves two roles: (1) it is able to construct spectrum for previously unseen abundances and (2) it is able to perform endmember extraction by setting up the abundances appropriately. The model assumes that spectra follow a multivariate Normal Distribution as be-

low:

$$\begin{aligned} \mathbf{x} &\sim \text{Normal}(\mathbf{x}; \mu, \Sigma) \text{ where} \\ \mathbf{x} &= \{x_1, x_2, x_3, \dots, x_k\} \\ \mu &= \{\mu_1, \mu_2, \mu_3, \dots, \mu_k\} \\ \Sigma &= \text{diag}(\sigma_1, \sigma_2, \sigma_3, \dots, \sigma_k) \end{aligned} \quad (4)$$

2.3. Loss function

The loss function has two components: (1) the reconstruction loss

$$\mathcal{L}(\mathbf{z}, \hat{\mathbf{z}}) = (\mathbf{z} - \hat{\mathbf{z}})^2, \quad (5)$$

where \mathbf{z} and $\hat{\mathbf{z}}$ are ground truth and estimated abundances, respectively, and (2) the ELBO Loss

$$\mathcal{L}(\mathbf{x}; \theta, \phi) = \mathbb{E}_{q_\theta} [\log p_\phi - \text{KL}(q_\theta(\mathbf{z}|\mathbf{x})||p(\mathbf{z}))]. \quad (6)$$

The first term on the left side is the reconstruction loss for the spectrum. The second term represents KL divergence that forces the latent representation towards a Dirichlet distribution. For further details, please refer [26, 29, 30, 31, 32].

3. EMPIRICAL EVALUATION

3.1. Datasets

All the experiments are conducted with four datasets. Samson is 95×95 , 156-channel hyperspectral image [33]. This dataset contains three endmembers: soil, tree, and water. UHYDICE Urban is 307×307 , 162-channel hyperspectral image covering a $2 \times 2 \text{ m}^2$ [34]. This dataset has three versions, containing four, five, and six endmembers. In this work, six endmembers are used. 80 : 20 training/testing split is used for both datasets. Cuprite dataset is a 512×614 , 188-channel hyperspectral image [35]. It contains twelve minerals (endmembers). This dataset does not provide ground truth abundances; therefore, we use Cuprite-synthetic dataset for training [26]. This showcases transfer learning, where a model trained on synthetic dataset is subsequently used to perform inference on a real world dataset. Lastly, OnTech-HSI-Syn-21 is a synthetic dataset containing nine endmembers. It is 128×128 , 224-channel hyperspectral image.

3.2. Metrics

Root Mean Square Error (RMSE) is used to evaluate abundance estimation accuracy. It is computed as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{n=1}^N (z_n - \hat{z}_n)^2},$$

where \mathbf{z} is the ground truth abundances, and $\hat{\mathbf{z}}$ is generated abundances. N denotes the number of pixels used in the computation.

Spectral Angle Distance (SAD) is a distance measurement between two spectral signals

$$\text{SAD} = \arccos \left(\frac{\hat{\mathbf{x}}_e^T \mathbf{x}_e}{\|\hat{\mathbf{x}}_e\| \|\mathbf{x}_e\|} \right),$$

where $\hat{\mathbf{x}}_e$ represents the estimated endmember and \mathbf{x}_e denotes the ground truth endmember. SAD is sometimes referred to as SAM [36]. In this work, we use SAD to capture endmember extraction accuracy.

3.3. Experimental Settings

SpACNN-LDVAE was implemented using Pyro Python library [37]. All models were trained on NVidia V100-SXM2 gpus. Adam optimizer was used with learning rate 0.001 [38].

4. RESULTS

Table 1 provides RMSE and SAD metrics (both endmember specific and average values) for Samson dataset. Similarly Table 2 lists these metrics for HYDICE Urban dataset. Cuprite lacks abundance ground truth, so only endmember extraction results are provided for this dataset (see Table 3). Tables 4 and 5 provide endmember extraction (SAD) and abundance estimation (RMSE) results for OnTech-Syn-HSI-21 dataset under various SNR settings, respectively. The proposed method achieves lower RMSE and SAD numbers over MLP-LDVAE method. However, the SAD scores on Cuprite are comparable to those achieved by MLP-LDVAE. Recall that the model was trained on Cuprite Synthetic dataset, which lacks spatial coherence. This is perhaps why the proposed method did not achieve better numbers than MLP-LDVAE. This merits further discussion. In all datasets, the proposed method achieves lower standard deviations.

5. CONCLUSION

This work extends MLP-LDVAE by replacing its encoder stage with a CNN with spatial attention. This allows the encoder to attend to the neighbouring pixels when constructing the latent representation for a given pixel. Our intuition is that hyperspectral images, similar to RGB images, exhibit spatial coherence, i.e. neighbouring pixels may contain similar endmembers and may exhibit similar abundances. The proposed model is evaluated on four datasets and the results suggest that incorporating local spatial information improves both endmember extraction and abundance estimation. Similar to MLP-LDVAE, our model is also capable of generating hyperspectral pixels given their abundances.

Table 1: Abundance Estimation and Endmember Extraction Results on Samson Dataset

		SpACNN-LDVAE	MLP-LDVAE [26]	VCA+FCLS [10]	PLMM [14]	ELMM [12, 13]	GLMM [39]	DeepGUn [24]	EACNN [22]
Soil	RMSE	0.2522 ± 0.00	0.2609 ± 0.00	-	-	-	-	-	-
	SAD	0.2097 ± 0.01	0.0959 ± 0.10	-	-	-	-	-	0.0328
Tree	RMSE	0.2614 ± 0.00	0.3431 ± 0.00	-	-	-	-	-	-
	SAD	0.5347 ± 0.03	1.2788 ± 1.28	-	-	-	-	-	0.0519
Water	RMSE	0.2098 ± 0.00	0.3165 ± 0.00	-	-	-	-	-	-
	SAD	0.8233 ± 0.04	0.4022 ± 0.40	-	-	-	-	-	0.1026
Average	RMSE	0.2412 ± 0.00	0.3078 ± 0.00	0.0545	0.0239	0.0119	0.0006	0.0862	0.0171
	SAD	0.5525 ± 0.03	0.5923 ± 0.59	-	-	-	-	-	0.0624

Table 2: Abundance Estimation and Endmember Extraction Results on HYDICE Urban Dataset

		SpACNN-LDVAE	MLP-LDVAE [26]	SSWNMF [17]	SGSNMF [18]	TV-RSNMF [19]	RSNMF [19]	GLNMF [20]	$L_{1/2}$ NMF [21]	VCA+FCLS [10]
Asphalt road	RMSE	0.1566 ± 0.00	0.2889 ± 0.00	-	-	-	-	-	-	-
	SAD	0.2786 ± 0.02	0.4262 ± 0.43	0.0782 ± 3.29	0.0841 ± 4.01	0.0770 ± 2.97	0.0869 ± 3.81	0.1008 ± 3.19	0.0889 ± 2.88	0.2246 ± 3.44
Grass	RMSE	0.1977 ± 0.00	0.1832 ± 0.00	-	-	-	-	-	-	-
	SAD	0.1936 ± 0.01	0.3323 ± 0.33	0.1490 ± 3.58	0.1516 ± 3.25	0.1495 ± 3.54	0.1594 ± 3.62	0.1531 ± 3.06	0.1452 ± 3.57	0.1981 ± 3.39
Tree	RMSE	0.1632 ± 0.00	0.1737 ± 0.00	-	-	-	-	-	-	-
	SAD	0.4411 ± 0.04	0.3177 ± 0.32	0.1173 ± 3.46	0.1199 ± 3.36	0.1269 ± 4.02	0.1457 ± 4.29	0.1424 ± 3.79	0.1509 ± 3.18	0.2137 ± 2.41
Roof	RMSE	0.1283 ± 0.00	0.125 ± 0.00	-	-	-	-	-	-	-
	SAD	0.4502 ± 0.03	0.4393 ± 0.44	0.0713 ± 3.61	0.0731 ± 3.54	0.0746 ± 4.09	0.0849 ± 3.90	0.0986 ± 4.62	0.0863 ± 4.06	0.2673 ± 3.77
Metal	RMSE	0.0992 ± 0.00	0.2599 ± 0.00	-	-	-	-	-	-	-
	SAD	0.3241 ± 0.02	0.7004 ± 0.70	0.1241 ± 2.76	0.1250 ± 3.81	0.1247 ± 3.53	0.1324 ± 4.15	0.1370 ± 4.28	0.1334 ± 3.90	0.1848 ± 3.68
Dirt	RMSE	0.1894 ± 0.00	0.1334 ± 0.00	-	-	-	-	-	-	-
	SAD	0.2026 ± 0.01	0.2806 ± 0.28	0.0802 ± 3.17	0.0859 ± 3.91	0.0849 ± 3.92	0.0798 ± 3.77	0.1059 ± 3.96	0.1063 ± 3.54	0.1992 ± 3.43
Average	RMSE	0.1558 ± 0.00	0.1840 ± 0.00	0.0048 ± 0.72	0.0061 ± 0.67	0.0055 ± 0.81	0.0053 ± 0.98	0.0069 ± 0.85	0.0044 ± 0.76	0.0119 ± 0.66
	SAD	0.3151 ± 0.02	0.4161 ± 0.42	0.1034 ± 3.31	0.1060 ± 3.68	0.1063 ± 3.68	0.1148 ± 3.92	0.1230 ± 3.52	0.1185 ± 3.52	0.2142 ± 3.35

Table 3: Endmember Extraction Results on Cuprite Dataset

		SpACNN-LDVAE	MLP-LDVAE [26]	SSWNMF [17]	SGSNMF [18]	TV-RSNMF [19]	RSNMF [19]	GLNMF [20]	$L_{1/2}$ NMF [21]	VCA+FCLS [10]
alunite		0.0683 ± 0.00	0.0097 ± 0.01	0.1497 ± 3.97	0.1238 ± 4.01	0.1204 ± 4.37	0.1189 ± 4.39	0.1353 ± 3.83	0.1496 ± 3.32	0.1574 ± 3.71
Andradite		0.0462 ± 0.00	0.0381 ± 0.04	-	-	-	-	-	-	-
Buddingtonite		0.0227 ± 0.00	0.0051 ± 0.01	0.0958 ± 4.69	0.1021 ± 3.47	0.0903 ± 5.08	0.1342 ± 4.72	0.1437 ± 3.62	0.1441 ± 4.16	0.1412 ± 3.74
Dumortierite		0.0500 ± 0.00	0.1922 ± 0.19	-	-	-	-	-	-	-
Kaolinite_1		0.0740 ± 0.00	0.0258 ± 0.03	0.0885 ± 2.94	0.0986 ± 3.18	0.1097 ± 3.47	0.0955 ± 3.07	0.0967 ± 4.01	0.0825 ± 4.66	0.0736 ± 4.42
Kaolinite_2		0.0249 ± 0.00	0.0699 ± 0.07	0.1206 ± 3.67	0.1375 ± 3.48	0.1213 ± 3.82	0.1396 ± 4.11	0.1356 ± 3.91	0.1402 ± 4.18	0.1420 ± 4.16
Muscovite		0.0320 ± 0.00	0.0064 ± 0.01	0.1024 ± 4.24	0.1061 ± 3.18	0.1131 ± 2.88	0.0997 ± 3.46	0.0961 ± 3.77	0.0889 ± 3.03	0.1007 ± 3.31
Montmorillonite		0.0214 ± 0.00	0.0496 ± 0.05	0.0651 ± 3.08	0.0705 ± 3.36	0.0783 ± 3.95	0.0744 ± 3.12	0.0838 ± 4.28	0.0876 ± 2.91	0.0974 ± 3.39
Nontronite		0.0639 ± 0.00	0.1048 ± 0.10	0.1138 ± 4.15	0.1046 ± 3.80	0.0911 ± 3.49	0.0832 ± 4.18	0.0953 ± 3.41	0.1038 ± 4.46	0.0772 ± 2.10
Pyrope		0.0342 ± 0.00	0.0156 ± 0.02	0.1106 ± 3.32	0.1208 ± 3.83	0.1253 ± 3.10	0.1469 ± 3.12	0.1318 ± 3.18	0.1123 ± 4.91	0.1437 ± 3.76
Sphene		0.1030 ± 0.00	0.0347 ± 0.03	0.1024 ± 3.79	0.1179 ± 4.02	0.1190 ± 2.97	0.1134 ± 2.54	0.1291 ± 4.21	0.1252 ± 5.18	0.1277 ± 4.08
Chalcedony		0.0281 ± 0.00	0.055 ± 0.01	0.1496 ± 4.12	0.1221 ± 4.02	0.1387 ± 4.01	0.1224 ± 4.19	0.1341 ± 2.98	0.1520 ± 3.43	0.1514 ± 3.83
Average		0.0470 ± 0.00	0.0465 ± 0.05	0.1099 ± 3.80	0.1104 ± 3.63	0.1107 ± 3.71	0.1128 ± 3.69	0.1182 ± 3.72	0.1186 ± 4.02	0.1212 ± 3.65

Table 4: Endmember Extraction Results on OnTech-Syn-HSI-21 Dataset

SNR	SpACNN-LDVAE	MLP-LDVAE [26]	SSWNMF [17]	SGSNMF [18]	TV-RSNMF [19]	RSNMF [19]	GLNMF [20]	$L_{1/2}$ NMF [21]	VCA + FCLS [10]
20 dB	0.0584 ± 0.00	0.0224 ± 0.01	0.0636 ± 0.40	0.0782 ± 0.50	0.0679 ± 0.30	0.0731 ± 0.50	0.0724 ± 0.05	0.0744 ± 0.40	0.1358 ± 0.30
30 dB	0.0616 ± 0.00	0.0138 ± 0.01	0.0122 ± 0.01	0.0176 ± 0.03	0.0131 ± 0.03	0.0138 ± 0.05	0.0144 ± 0.04	0.0142 ± 0.04	0.0350 ± 0.06
40 dB	0.0613 ± 0.00	0.0081 ± 0.00	0.0029 ± 0.02	0.0033 ± 0.03	0.0036 ± 0.02	0.0041 ± 0.04	0.0044 ± 0.05	0.0037 ± 0.04	0.0125 ± 0.05
50 dB	0.0545 ± 0.00	0.0082 ± 0.00	0.0012 ± 0.02	0.0019 ± 0.02	0.0014 ± 0.03	0.0020 ± 0.04	0.0023 ± 0.04	0.0024 ± 0.03	0.0049 ± 0.06
INF	0.0594 ± 0.00	0.0069 ± 0.00	-	-	-	-	-	-	-

Table 5: Abundance Estimation Results on OnTech-Syn-HSI-21 Dataset

SNR	SpACNN-LDVAE	MLP-LDVAE [26]	SSWNMF [17]	SGSNMF [18]	TV-RSNMF [19]	RSNMF [19]	GLNMF [20]	$L_{1/2}$ NMF [21]	VCA + FCLS [10]
20 dB	0.0948 ± 0.00	0.0052 ± 0.00	0.1339 ± 0.20	0.1322 ± 0.40	0.1342 ± 0.30	0.1426 ± 0.40	0.1434 ± 0.60	0.1430 ± 0.50	0.1704 ± 0.03
30 dB	0.3356 ± 0.00	0.0302 ± 0.00	0.0386 ± 0.20	0.0391 ± 0.30	0.0420 ± 0.20	0.0426 ± 0.30	0.0429 ± 0.03	0.0432 ± 0.20	0.0548 ± 0.20
40 dB	0.3343 ± 0.00	0.0303 ± 0.00	0.0122 ± 0.03	0.0148 ± 0.05	0.0142 ± 0.04	0.0147 ± 0.05	0.0150 ± 0.04	0.0153 ± 0.03	0.0164 ± 0.10
50 dB	0.3335 ± 0.00	0.0303 ± 0.00	0.0041 ± 0.02	0.0059 ± 0.05	0.0050 ± 0.03	0.0055 ± 0.03	0.0064 ± 0.04	0.0061 ± 0.04	0.0087 ± 0.08
INF	0.0948 ± 0.00	0.0052 ± 0.00	-	-	-	-	-	-	-

6. REFERENCES

- [1] Rob Heylen, Mario Parente, and Paul Gader, "A review of nonlinear hyperspectral unmixing methods," *Selected Topics in Applied Earth Observations and Remote Sensing*, *IEEE Journal of*, vol. 7, pp. 1844–1868, 06 2014.
- [2] Emmanuel Maggiori, Antonio Plaza, and Yuliya Tarabalka, *Models for Hyperspectral Image Analysis: From Unmixing to Object-Based Classification*, pp. 37–80, 01 2018.
- [3] Javier Plaza, Eligius Hendrix, Inmaculada García Fernández, Gabriel Martín, and Antonio Plaza, "On end-member identification in hyperspectral images without pure pixels: A comparison of algorithms," *Journal of Mathematical Imaging and Vision*, vol. 42, pp. 163–175, 02 2012.
- [4] Shaoquan Zhang, Guorong Zhang, Fan Li, Chengzhi Deng, Shengqian Wang, Antonio Plaza, and Jun Li, "Spectral-spatial hyperspectral unmixing using nonnegative matrix factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.
- [5] Farshid Khajehrayeni and Hassan Ghassemian, "A linear hyperspectral unmixing method by means of autoencoder networks," *International Journal of Remote Sensing*, vol. 42, pp. 2517–2531, 04 2021.
- [6] Min Zhao, Xiuheng Wang, Jie Chen, and Wei Chen, "A plug-and-play priors framework for hyperspectral unmixing," 12 2020.
- [7] Antonio Plaza and Chein-I Chang, "Fast implementation of pixel purity index algorithm," *Proceedings of SPIE - The International Society for Optical Engineering*, 06 2005.
- [8] Michael E. Winter, "N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data," in *Imaging Spectrometry V*, Michael R. Descour and Sylvia S. Shen, Eds., Oct. 1999, vol. 3753 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pp. 266–275.
- [9] Joseph W. Boardman, Fred A. Kruse, and Robert O. Green, "Mapping target signatures via partial unmixing of aviris data: in summaries," 1995.
- [10] J.M.P. Nascimento and J.M.B. Dias, "Vertex component analysis: a fast algorithm to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898–910, 2005.
- [11] Daniel C. Heinz and Chein-I. Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote. Sens.*, vol. 39, pp. 529–545, 2001.
- [12] M.A. Veganzones, L. Drumetz, G. Tochon, M. Dalla Mura, A. Plaza, J. Bioucas-Dias, and J. Chanussot, "A new extended linear mixing model to address spectral variability," in *2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2014, pp. 1–4.
- [13] Lucas Drumetz, Miguel-Angel Veganzones, Simon Henrot, Ronald Phlypo, Jocelyn Chanussot, and Christian Jutten, "Blind hyperspectral unmixing using an extended linear mixing model to address spectral variability," *IEEE Transactions on Image Processing*, vol. 25, no. 8, pp. 3890–3905, 2016.
- [14] Pierre-Antoine Thouvenin, Nicolas Dobigeon, and Jean-Yves Tournet, "Hyperspectral unmixing with spectral variability using a perturbed linear mixing model," *IEEE Transactions on Signal Processing*, vol. 64, no. 2, pp. 525–538, 2016.
- [15] Nicolas Dobigeon, Jean-Yves Tournet, Cedric Richard, Jose Carlos M. Bermudez, Stephen McLaughlin, and Alfred O. Hero, "Nonlinear unmixing of hyperspectral images: Models and algorithms," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 82–94, jan 2014.
- [16] Jignesh S. Bhatt, Manjunath V. Joshi, and Mehul S. Raval, "A data-driven stochastic approach for unmixing hyperspectral imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 1936–1946, 2014.
- [17] Jiming Tang, Wenxing Bao, Bingbing Lei, Kewen Qu, and Wei Feng, "Spatial-spectral weighted sparse multi-layer nonnegative matrix factorization for hyperspectral image unmixing," in *2023 13th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2023, pp. 1–5.
- [18] Xinyu Wang, Yanfei Zhong, Liangpei Zhang, and Yanyan Xu, "Spatial group sparsity regularized nonnegative matrix factorization for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 11, pp. 6287–6304, 2017.
- [19] Wei He, Hongyan Zhang, and Liangpei Zhang, "Total variation regularized reweighted sparse nonnegative matrix factorization for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3909–3921, 2017.
- [20] Xiaoqiang Lu, Hao Wu, Yuan Yuan, Pingkun Yan, and Xuelong Li, "Manifold regularized sparse nmf for hy-

- perspectival unmixing,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 5, pp. 2815–2826, 2013.
- [21] Yuntao Qian, Sen Jia, Jun Zhou, and Antonio Robles-Kelly, “Hyperspectral unmixing via $l_{1/2}$ sparsity-constrained nonnegative matrix factorization,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4282–4297, 2011.
- [22] Baohua Jin, Yunfei Zhu, Wei Huang, Qiqiang Chen, and Sijia Li, “An efficient attention-based convolutional neural network that reduces the effects of spectral variability for hyperspectral unmixing,” *Applied Sciences*, vol. 12, no. 23, 2022.
- [23] Yuanchao Su, Jun Li, Antonio Plaza, Andrea Marinoni, Paolo Gamba, and Somdatta Chakravorty, “Daen: Deep autoencoder networks for hyperspectral unmixing,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 7, pp. 4309–4321, 2019.
- [24] Ricardo Borsoi, Tales Imbiriba, and Jose Bermudez, “Deep generative endmember modeling: An application to unsupervised spectral unmixing,” *IEEE Transactions on Computational Imaging*, vol. PP, pp. 1–1, 10 2019.
- [25] Qiwen Jin, Yong Ma, Xiaoguang Mei, and Jiayi Ma, “Tanet: An unsupervised two-stream autoencoder network for hyperspectral unmixing,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2022.
- [26] Kiran Mantripragada and Faisal Qureshi, “Hyperspectral pixel unmixing with latent dirichlet variational autoencoder,” 03 2022.
- [27] David M. Blei, Andrew Y. Ng, and Michael I. Jordan, “Latent dirichlet allocation,” *J. Mach. Learn. Res.*, vol. 3, no. null, pp. 993–1022, mar 2003.
- [28] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon, “Cbam: Convolutional block attention module,” 2018.
- [29] Lucas Pinheiro Cinelli, Matheus Araújo Marins, Eduardo Ant3nio Barros da Silva, and S3rgio Lima Netto, *Variational Autoencoder*, pp. 111–149, Springer International Publishing, Cham, 2021.
- [30] David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe, “Variational inference: A review for statisticians,” *Journal of the American Statistical Association*, vol. 112, no. 518, pp. 859–877, apr 2017.
- [31] Charles Fox and Stephen Roberts, “A tutorial on variational bayes,” *Artificial Intelligence Review - AIR*, vol. 38, pp. 1–11, 08 2012.
- [32] Diederik P Kingma and Max Welling, “Auto-encoding variational bayes,” 2022.
- [33] “Samson hsi dataset,” .
- [34] “Hydice urban hsi dataset,” .
- [35] “Cuprite dataset,” .
- [36] “Spectral angle mapper algorithm for remote sensing image classification,” 2014.
- [37] Eli Bingham, Jonathan P. Chen, Martin Jankowiak, Fritz Obermeyer, Neeraj Pradhan, Theofanis Karaletsos, Rohit Singh, Paul Szerlip, Paul Horsfall, and Noah D. Goodman, “Pyro: Deep Universal Probabilistic Programming,” *Journal of Machine Learning Research*, 2018.
- [38] Diederik P. Kingma and Jimmy Ba, “Adam: A method for stochastic optimization,” 2017.
- [39] Tales Imbiriba, Ricardo Augusto Borsoi, and Jos3 Carlos Moreira Bermudez, “Generalized linear mixing model accounting for endmember variability,” in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 1862–1866.