AN ITERATIVE METHOD FOR HYPERSPECTRAL PIXEL UNMIXING LEVERAGING LATENT DIRICHLET VARIATIONAL AUTOENCODER

Kiran Mantripragada[*], Paul R. Adler[†], Peder A. Olsen[‡], Faisal Z. Qureshi[*]

[*]Ontario Tech University

[†]USDA - Agricultural Research Service

[‡]Microsoft Research

ABSTRACT

We develop a hyperspectral pixel unmixing method that uses a Latent Variational Autoencoder within an analysis-synthesis loop to (1) construct pure spectra of the materials present in an image and (2) infer the mixing ratios of these materials in hyperspectral pixels without the need of labelled data. On OnTech-HSI-Syn-6em synthetic dataset that contains pixel unmixing groundtruth, the proposed method achieves $acc = 100\%$, $SAD = 0.0582$ and $RMSE = 0.0695$ for segmentation, endmember extraction and abundance estimation, respectively. On HYDICE Urban benchmark, the proposed method achieves $acc = 72.4\%$, $SAD = 0.1669$ and $RMSE = 0.1984$ for segmentation, endmember extraction and abundance estimation, respectively. Additionally, we applied this technique for crop analysis on hyperspectral data collected by the United States Department of Agriculture and achieved a coefficient of determination $R^2 = 0.7$ with respect to the ground truth. These results confirm that the proposed method is able to perform pixel unmixing without using labelled data.

Index Terms— Iterative Method, Latent Dirichlet Variational Autoencoder, Hyperspectral Unmixing.

I. INTRODUCTION

Pixel Unmixing is an important hyperspectral analysis technique that seeks to extract the materials and their mixing ratios responsible for the spectra observed at each pixel location [1, 2]. In recent years, machine learning-based approaches have achieved good results on this problem; however, these techniques rely upon large amounts of pixel-level labelled data (containing materials and their mixing ratios for hyperspectral pixels), which is often infeasible and always costly to obtain in practice. Here we develop iLDVAE, a method that relaxes the necessity of labelled data. The proposed method uses a Latent Dirichlet Variational Autoencoder (LDVAE) in an analysis-synthesis loop to solve pixel unmixing in the absence of labelled data [3]. Given a hyperspectral image, iLDVAE iteratively searches for pure pixels, i.e., pixels that contain spectra of a single material, and uses this information to extract the endmembers and their abundances for each pixel.

We have evaluated iLDVAE on three datasets: 1) OnTech-HSI-Syn-6em, a synthetic HSI image constructed using six endmembers from the United States Geological Survey Spectral Library [4], 2) HYDICE Urban benchmark [5], and 3) Cover-Crop-HSI dataset from the United States Department of Agriculture. The results obtained confirm the validity of the iLDVAE for pixel unmixing in the absence of labelled data. The main contribution of this work is a deep learning-based self-supervised method for hyperspectral pixel unmixing that does not require access to pixel-level labelled data, which is almost always costly, and often infeasible to collect.

II. RELATED WORK

Hyperspectral pixel unmixing is an active area of research, and it is not possible to provide a comprehensive overview of various approaches available in the literature in the space allotted for this paper. Instead, we refer the kind reader to [6, 7, 8] that provide a good summary of the existing techniques for pixel unmixing.

Broadly speaking, we can classify pixel unmixing techniques into two categories. Methods that belong to the first category employ the so-called traditional optimization techniques, such as non-negative matrix factorization, blind source separation, etc. Examples of methods that belong to this category are [9, 10, 11, 12, 13]. Methods in the second category use neural networks, and these methods require a data-intensive "training" phase that requires access to pixel-level labelled data [14, 15]. DeepGUn, one example of a deep learning-based approach for hyperspectral pixel unmixing is [16], uses the learned latent representations for endmembers to generate their spectra. Unlike [16], LDVAE [3] encodes endmembers and their mixing ratios as a Dirichlet distribution constructed using the encoder stage of a variational autoencoder. LDVAE can subsequently use the decoder stage to construct the pixel spectra of pure or mixed pixels.

iLDVAE, the method presented in this paper, is an extension of LDVAE. Unlike LDVAE, which requires labelled data for training the variational autoencoder responsible for analyzing hyperspectral pixels and constructing pixel spectra, iLDVAE uses LDVAE within an analysis-synthesis loop that allows iLDVAE to "learn" endmember spectra without any labelled data. Over multiple iterations, iLDVAE learns endmember spectra and then uses this information to infer abundances in hyperspectral pixels. iLDVAE, consequently, unlike other deep learning-based approaches, does not require access to pixel-level labelled "training" data.

III. METHOD

iLDVAE is an iterative algorithm that uses LDVAE within an analysis-synthesis loop [3] to estimate endmembers and their abundances over successive iterations without using any labelled data (See Algorithm 1). iLDVAE implements the idea that pixels with high purity-index can serve as proxy for endmembers. Given a target image $\mathbf{I}$ that needs to be unmixed, at each iteration k, N randomly selected pixels with purity-index above a certain threshold are chosen as endmembers $\{\mathbf{e}_i^{(k)}\}_{i=1}^N$ that are used to synthesize a hyperspectral image $\mathbf{I}^{(k)}$. Each pixel in $\mathbf{I}^{(k)}$ contains $\{\mathbf{e}_i^{(k)}\}_{i=1}^N$ in random proprotions. LDVAE is trained on $\mathbf{I}^{(k)}$ and subsequently used to estimate endmembers $\{\mathbf{e}_i^{(k+1)}\}_{i=1}^N$ and per-pixel abundances $\mathbf{a}_{(x,y)}$ in $\mathbf{I}$, where subscript (x, y) denote pixel at that location and $\mathbf{a}_{(x,y)} \in \mathbb{R}^N$. The estimated endmembers along with their per-pixel abundances serve to compute the purity-index for pixels in $\mathbf{I}$ for the next step. The process stops when loop termination conditions are met. We discuss loop termination in the following section.

III-A. Loop Termination

iLDVAE loop is terminated when one of the two following conditions are met: 1) the number of allowed iterations is reached and 2) the disagreement err between endmembers estimations for two consecutive iterations is less than or equal to a pre-defined threshold ϵ, where

$$\text{err} = \frac{1}{\rho N} \sum_{i=1}^{N} \sqrt{\frac{\|\mathbf{e}_i^{(k+1)} - \mathbf{e}_i^{(k)}\|^2}{L_i}}. \tag{1}$$

We define L_i and ρ as follows. Let $\mathbf{S}$ denote the segmentation image such that $\mathbf{S}_{(x,y)} = \text{argmax}_i\, \mathbf{a}_{(x,y)}$. $\mathbf{S}$ is a single-channel image with the same spatial dimensions as $\mathbf{I}$. Each pixel in $\mathbf{S}$ contains the index of the most abundant endmember at that pixel. Then $L_i = \sum_{(x,y)} \mathbb{1}_{\mathbf{S}(x,y)=i}$, where $\mathbb{1}$ is the indicator function. Let

$$p_i = \max_{\forall \mathbf{S}_{(x,y)}=i} \mathbf{a}_{(x,y)}$$

denotes the maximum abundance value for endmember i then $\rho = \min\{p_i\}_{i=1}^N$.

Algorithm 1 iLDVAE algorithm for hyperspectral pixel unmixing

Require: Target image $\mathbf{I}$
Require: The number of endmembers N present in the target image
Ensure: Estimated endmembers $\{\mathbf{e}_i\}_{i=1}^N$
Ensure: Estimated per-pixel abundances $\{a_i\}_{i=1}^N$ where $a_i \geq 0$ and $\sum_i a_i = 1$
1: $k = 0$ ▷ iteration count
2: Randomly select N pixels from $\mathbf{I}$ ▷ initial endmembers
3: Use $\{\mathbf{e}_i^{(k)}\}_{i=1}^N$ to synthesize $\mathbf{I}^{(k)}$ hyperspectral image where each pixel has random, but known, abundances
4: Train LDVAE on $\mathbf{k}^{(t)}$
5: Increment k
6: Use the LDVAE trained in the Step to estimate endmembers $\{\mathbf{e}_i^{(k)}\}_{i=1}^N$ and per-pixel abundances in $\mathbf{I}$
7: if err $\leq \epsilon$ then ▷ see Sec. III-A for details about err
8: Terminate loop
9: end if
10: Collect pixels for which $a_i >$ pure-pixel-threshold ▷ a_i denotes the abundance value for endmember i
11: Randomly select N pixels from the set of pixels in the previous step
12: if Maximum numbers of iterations reached then
13: Terminate loop
14: end if
15: Go to Step 3

IV. EXPERIMENTAL SETUP AND RESULTS

We evaluate iLDVAE on three datasets:

• OnTech-HSI-Syn-6em: is a synthetic dataset containing pixels containing random proportions of six species of vegetation taken from the Aviris 2014, Chapter V (vegetation), USGS Spectral Library [4]. This dataset allows us (and others) to evaluate unmixing methods since per-pixel endmember abundances are available. Note that ground truth information is only used to evaluate iLDVAE.

• HYDICE Urban: a 307×307 hyperspectral image that is widely used as a benchmark for unmixing methods [5]. Here we only use 162 channels: channels 1–4, 76, 87, 101–111, 136–153 and 198–210 are removed due to dense water vapor and atmospheric effects [5].

• Cover Crop USDA: A cover crop experiment with 60 plots was established on September 5, 2019, at the USDA-ARS Research Farm, located in Pennsylvania Furnace, PA. Four species consisting of two grasses and two legumes were planted in a simplex design. Individual plots were 3.5m by 15m with a row spacing of 17.8cm. The HSIs were collected from the site on April 28th, 2020, using a DJI Matrice 600 Pro, which collects data in 270 spectral bands over the range of 400–1000 nm with a spectral resolution of 2.2 nm. We evaluate a total of 120

Table I: SAD and RMSE metric, respectively for endmember extraction abundances estimation (OnTech-HSI-Syn-6em dataset).

	SAD	RMSE
sagebrush	0.0447 ±0.0005	0.0664 ±0.0001
tumbleweed	0.0853 ±0.0005	0.0712 ±0.0002
grass	0.0233 ±0.0009	0.0342 ±0.0001
cattail	0.0721 ±0.0021	0.0899 ±0.0000
juniper	0.0429 ±0.0014	0.0796 ±0.0001
cactus	0.0810 ±0.0023	0.0758 ±0.0001
average	0.0582 ±0.0248	0.0695 ±0.0191

Table II: SAD and RMSE metric, respectively for endmember extraction abundances estimation (HYDICE Urban dataset).

	SAD	RMSE
asphalt	0.0626 ±0.0029	0.1438 ±0.000 02
grass	0.1251 ±0.0010	0.1571 ±0.000 04
tree	0.1142 ±0.0016	0.2547 ±0.000 04
roof	0.0841 ±0.0017	0.1497 ±0.000 02
metal	0.5362 ±0.0047	0.3360 ±0.000 02
dirt	0.0793 ±0.0042	0.1490 ±0.000 01
average	0.1669 ±0.1824	0.1984 ±0.0795

datacubes from the 60 plots (2 datacubes per plot). Each datacube is called a quadrat. Thirty-two quadrats are monocultures, and 88 correspond to multicultures with up to 4 species plus soil. All the quadrats are presented with a spatial size of 55×55 pixels and 270 bands.

IV-A. Results: OnTech-HSI-Syn-6em & HYDICE Urban

Ground truth is only available for OnTech-HSI-Syn-6em and HYDICE Urban dataset. We use iLDVAE to estimate endmembers and per-pixel abundances for these datasets and use Spectral Angle Distance (SAD) and Root Mean Squared Error (RMSE) metrics to capture the performance of iLDVAE [10, 17]. Tables I and II show SAD and RMSE results for OnTech-HSI-Syn-6em and HYDICE Urban datasets. SAD scores capture the quality of the estimated endmembers, whereas RMSE scores record the agreement between ground truth and estimated per-pixel abundances.

In addition, we also use segmentation accuracy

$$\text{acc}_{\text{seg}} = \frac{1}{W \cdot H} \mathbb{1}_{\mathbf{S}(x,y)=\mathbf{S}^{\text{gt}}(x,y)},$$

where W and H represent the width and height of the $\mathbf{I}$, respectively, $\mathbf{S}^{\text{gt}}$ is the ground truth segmentation, and $\mathbb{1}$ is the indicator function. Figures 1 and 2 show segmentation results for OnTech-HSI-Syn-6em and HYDICE Urban datasets.

IV-B. Results: Cover Crop USDA

Cover Crop USDA dataset does not contain per-pixel abundances. Rather it only contains species plus soil proportions in each quadrat. For this dataset, we

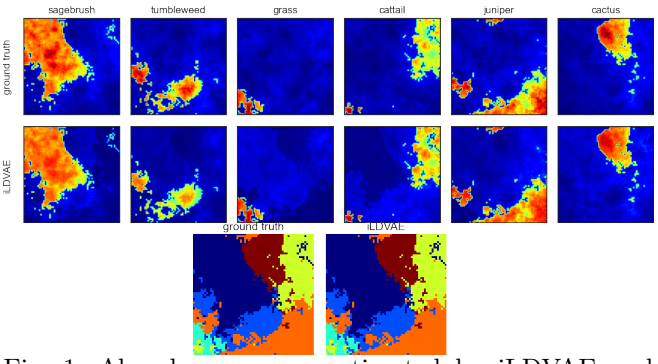

Fig. 1: Abundances maps estimated by iLDVAE and Segmentation results (accuracy = 1.00)

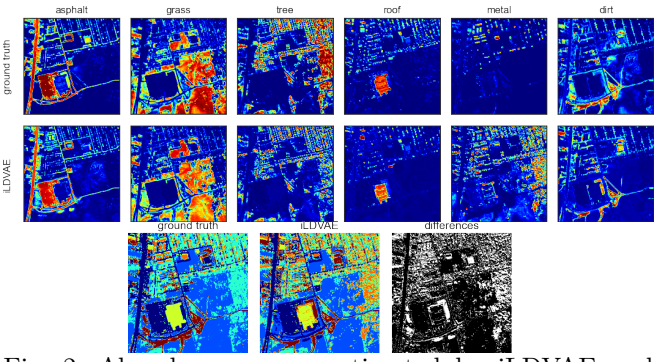

Fig. 2: Abundances maps estimated by iLDVAE and Segmentation results (accuracy = 0.7238)

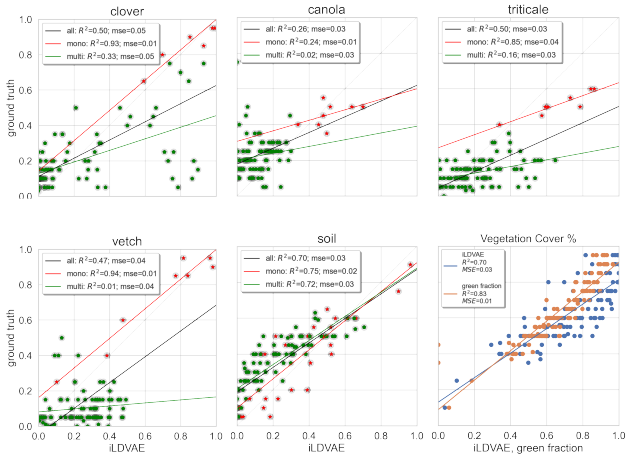

Fig. 3: Abundances of each species estimated by iLDVAE (x-axis) vs. ground truth (y-axis); bottom-right: Percentage of vegetation cover: iLDVAE-estimated and Canopeo Green Fraction index [18, 19] vs. ground truth (y-axis). Each data point corresponds to one quadrat (total number of quadrats=120).

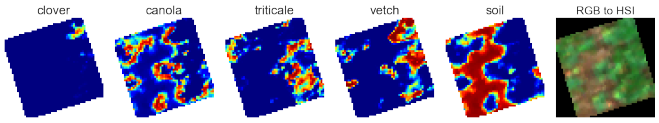

Fig. 4: iLDVAE-estimated abundance maps and RGB image composite from HSI.

aggregate per-pixel abundances for each quadrat to estimate the proportion of the four species plus soil in that quadrat. We use per-species (plus soil) Pearson's Coefficient of Determination R^2 [20] to see how well our method is able to estimate the proportions. Results are shown in Fig. 3. Additionally, we compare iLDVAE against Canopeo Green Fraction method [18, 19], which estimates the percentage of vegetation plus soil in each quadrate using RGB information only. Fig. 4 shows estimated abundances for one quadrat. Most noteworthy is the abundance map of soil, which closely matches the RGB image on the right. Soil is visible in this RGB image that is constructed from the HSI image.

IV-C. Discussion

The experiments on OnTech-HSI-Syn-6em resulted in an accuracy of 100% on the segmentation task, with an average $SAD = 0.0582$ and $RMSE = 0.0695$, respectively for endmembers extraction abundances estimation (Table I and Figure 1). The results on HYDICE Urban showed an accuracy of 72.38% on the segmentation task, average $SAD = 0.1699$ and $RMSE = 0.1984$ respectively for endmembers extraction and abundances estimation (Table II and Figure 2). The main reason for the lower accuracy on HYDICE Urban dataset is the lack of pixels of higher purity for some of the materials. For Cover Crop USDA, iLDVAE demonstrated a high correlation with the "Percentage Vegetation Cover" metric compared to Canopeo Green Fraction algorithm and ground truth (Coefficient of Determination $R^2 = 0.7$).

V. CONCLUSIONS

We presented iLDVAE, a self-supervised method for hyperspectral pixel unmixing. Results on ONTech-Syn-HSI-6em and HYDICE Urban suggest that iLDVAE is able to extract endmembers and per-pixel abundances without the need of ground truth "training" data. We also show how the proposed method is able to estimate proportions of various species plus soil in a given area by leveraging hyperpixel unmixing. The main limitation of the proposed method is that it requires pixels with high-purity index to bootstrap endmember extraction and per-pixel abundance estimation. In a practical setting, this limitation can be addressed by either providing endmembers present in the image or by ensuring that atleast a few pixels have a high-purity index, i.e., the image contains pixels that are "close" to endmembers. In the future we plan to write an extended version of this paper, including a comparison with other unmixing methods.

References

[1] E. Maggiori et al., "Models for hyperspectral image analysis: From unmixing to object-based classification," Math Models for Remote Sensing Image Processing, 2018.

[2] J. M. Bioucas-Dias et al., "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 2012.

[3] K. Mantripragada et al., "Hyperspectral pixel unmixing with latent dirichlet variational autoencoder," arXiv preprint arXiv:2203.01327, 2022.

[4] U. S. G. Survey et al., "Usgs spectral library version 7," U. S. Geological Survey, Reston, VA, Tech. Rep., 2017. [Online]. Available: http://pubs.er.usgs.gov/publication/ds1035

[5] "Hydice urban hsi dataset," https://rslab.ut.ac.ir/data (last accessed on November 16, 2021).

[6] J. S. Bhatt, Regularization in hyperspectral unmixing, ser. SPIE. Spotlight ; SL25. Bellingham, Washington (1000 20th St. Bellingham WA 98225-6705 USA): SPIE, 2016.

[7] R. Pu, Hyperspectral remote sensing : fundamentals and practices, ser. Remote sensing applications. Boca Raton, Florida: CRC Press, 2017.

[8] Advances in hyperspectral image processing techniques, ser. IEEE Press Ser. Hoboken, New Jersey: Wiley, 2023.

[9] M. E. Winter, "N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data," vol. 3753.

[10] E. Ibarrola-Ulzurrun et al., "Hyperspectral classification through unmixing abundance maps addressing spectral variability," IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 7, pp. 4775–4788, 2019.

[11] R. Mahabir et al., "A critical review of high and very high-resolution remote sensing approaches for detecting and mapping slums: Trends, challenges and emerging opportunities," Urban Science, 2018.

[12] J. M. Nascimento et al., "Does independent component analysis play a role in unmixing hyperspectral data?" IEEE Transactions on Geoscience and Remote Sensing, vol. 43, no. 1, pp. 175–187, 2005.

[13] J. Ball et al., "Hyperspectral pixel unmixing via spectral band selection and dc-insensitive singular value decomposition," IEEE geoscience and remote sensing letters, vol. 4, no. 3, pp. 382–386, 2007.

[14] Y. Bazi, Advanced Deep Learning Strategies for the Analysis of Remote Sensing Images. MDPI - Multidisciplinary Digital Publishing Institute, 2021.

[15] T. Fang et al., "Nonlinear hyperspectral unmixing based on multilinear mixing model using convolutional autoencoders," arXiv.org, 2023.

[16] R. A. Borsoi et al., "Deep generative endmember modeling: An application to unsupervised spectral unmixing," IEEE Transactions on Computational Imaging, vol. 6, pp. 374–384, 2019.

[17] H. Deborah et al., "A comprehensive evaluation of spectral distance functions and metrics for hyperspectral image processing," 2015.

[18] Y. S. Chung et al., "Case study: Estimation of sorghum biomass using digital image analysis with canopeo," Biomass and Bioenergy, 2017.

[19] A. Patrignani et al., "Canopeo: A powerful new tool for measuring fractional green canopy cover," Agronomy Journal.

[20] D. Chicco et al., "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," PeerJ Computer Science, 2021.