PIXELS-TO-ABUNDANCES TRANSLATION WITH SPATIAL-SPECTRAL CONDITIONAL GENERATIVE ADVERSARIAL NETWORKS FOR HYPERSPECTRAL UNMIXING

Li Wang, Xiaohua Zhang, Shengyuan Zheng, Tianrui Li, Jing Wang

Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education,
International Research Center for Intelligent Perception and Computation,
Joint International Research Laboratory of Intelligent Perception and Computation,
School of Artificial Intelligence, Xidian University, Xi'an, Shaanxi Province 710071, China

ABSTRACT

As the interaction of light with the Earth surface is very complex, spectral pixels are composed of sophisticated mixtures of distinct substances. The parameters of the estimated models are difficult to set. In this paper, we first propose a spatial-spectral conditional generative adversarial networks (scGANs) method to solve this problem, based on the following assumptions: the unmixing process from pixels to abundance can be viewed as a transformation of two modalities with an intrinsic specific relationship. The method learns the manifold structure of the hyperspectral data using an adversarial strategy, inputting pixels blocks with spatial information to generate the abundance of the central pixels. Then, we proposed superpixel segmentation and random splitting method to synthesize data with spatial structure. Finally, the proposed scGANs method is evaluated using synthetic data and real hyperspectral data, and compared to several state-of-the-art methods. The proposed method outperforms all the comparison methods in the experiments.

***Index Terms*—**Hyperspectral unmixing, spatial-spectral information, conditional generative adversarial networks, superpixel segmentation and random splitting

1. INTRODUCTION

Hyperspectral unmixing (HSU) is an important and challenging task in the field of hyperspectral remote sensing, aiming to estimate the pure spectral signatures of the mixed pixels in a hyperspectral image (HSI) along with their fractional abundances.

In order to estimate the fractional abundances of the different substances within a hyperspectral pixel, the error between the pixel's true reflectance spectra and the reflectance spectra that is generated by a particular mixing model is minimized. Traditionally, the linear mixing model (LMM) is assumed, which is valid when the incoming rays of light interact with a single pure material in the pixels instantaneous field of view (IFOV) before reaching the sensor. Sparse unmixing via variable splitting augmented lagrangian (SUNSAL) [1] is a representative method that reconstructs HSI by a sparse linear combination of selected endmembers from a spectral dictionary.

To model multiple interactions of the light, nonlinear mixing models have been developed. Some examples are the bilinear mixing models like generalized bilinear model (GBM) [2] that describes the interaction of an incident ray of light with two substances, and extensions of these, e.g. the multilinear mixing model (MLM) [3], and the intimate mixture models such as the Hapke model [4].

One of the problems remaining to be solved is that not all hyperspectral dataset follows the same particular mixing model. Another problem is that the parameters of the estimated models are difficult to set. So as to overcome these difficulties, we proposed a spatial-spectral scGANs method that takes a new perspective on the hyperspectral unmixing process, in which the process of obtaining the estimated abundance from the hyperspectral pixels with the same endmember composition can be regarded as a transformation of two modalities, and the abundance is a manifold representation in the low-dimensional space of the hyperspectral pixels. Many problems in image processing, computer graphics, and computer vision can be posed as "translating" an input image into a corresponding output image. In analogy to image-to-image translation, we define automatic pixels-to-abundances translation as the task of translating one possible representation of an image into another, using limited amounts of abundances labels. The proposed method in this paper is to develop a common framework for the unmixing of the HSI composed of the same endmembers. The method learns the manifold structure

This work was supported in part by the National Natural Science Foundation of China under Grant 61877066 and Aero-Science Fund under Grant 61471274

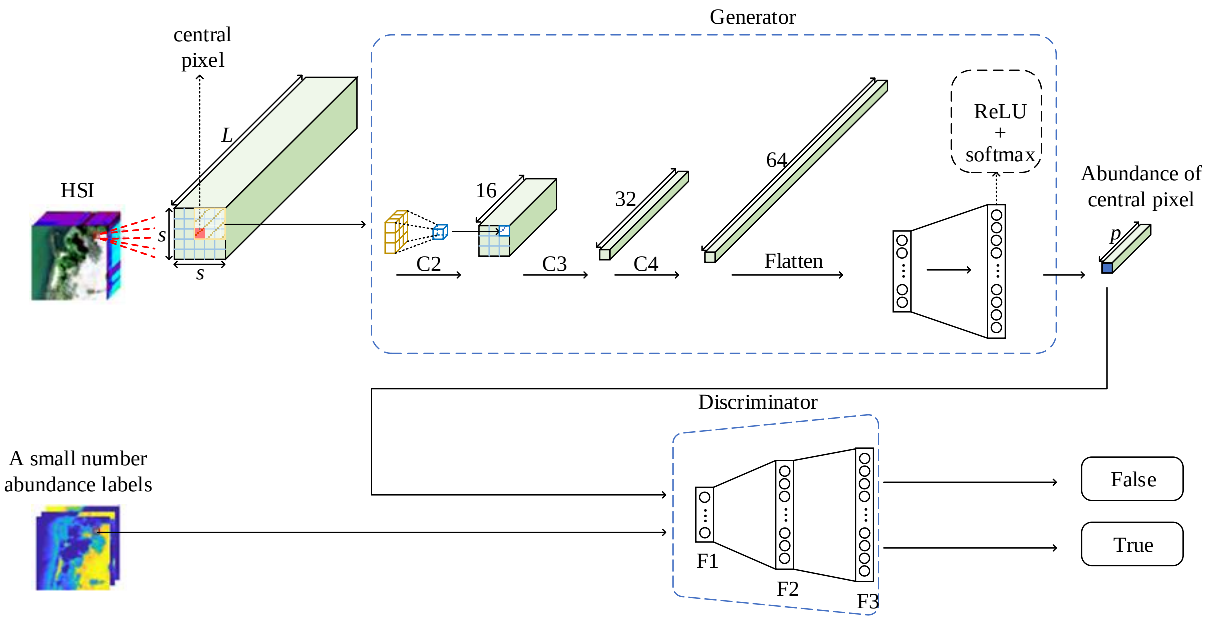

Fig. 1: A schematic of the proposed method.

of the hyperspectral data using an adversarial strategy [5], inputting pixel blocks to generate the abundance of the central pixels, and operating on the patches instead of single pixels enables the method to take advantage of spatial information in HSI under the assumption that all pixels from a small neighborhood are more likely to have similar abundances. Then, we propose a data synthesis method based on superpixel segmentation and random splitting to get synthetic data with spatial structure, while the usual data synthesis method is for a single pixel that without considering the spatial structure of HSI.

The rest of this paper is organized as follows. Section 2 describes the proposed method. The experiments on real HSI data and synthetic data are given in Section 3. Finally, conclusions are given in Section 4.

2. THE METHOD

2.1. Spatial-spectral cGANs

Let $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_N] \in \mathbb{R}^{L\times N}$ be the pixels of HSIs, $\boldsymbol{M} = [\boldsymbol{m}_1, \boldsymbol{m}_2, ..., \boldsymbol{m}_p] \in \mathbb{R}^{L\times p}$ represent the endmember matrix with each column denoting one of the p pure spectral signatures, and $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, ..., \boldsymbol{a}_N] \in \mathbb{R}^{p\times N}$ is the corresponding abundance map. Considering the general mixing mechanism [6], the ith pixel can be expressed as:

$$\boldsymbol{x} = \Phi(\boldsymbol{M}, \boldsymbol{a}) + \boldsymbol{n} \tag{1}$$

where Φ is an implicit function that defines complex interactions between the endmembers in matrix $\boldsymbol{M}$ parameterized by $\boldsymbol{a}$, and $\boldsymbol{n}$ is the natural noise of $\boldsymbol{x}$. We propose a spatial-spectral scGANs method to learn the complex relationship between $\boldsymbol{A}$ and $\boldsymbol{X}$ to complete hyperspectral umixing under the supervised scenes. That is, for HSI($\boldsymbol{X}$) composed of the same endmembers, our method can generate the abundance after being trained with a fraction of the corresponding abundance ($\boldsymbol{A}$). As illustrated in Fig. 1, our model includes mapping G: $\boldsymbol{X}\rightarrow\boldsymbol{A}$, and the corresponding adversarial discriminator D_A aims to distinguish between true abundance $\{a\}$ and generative abundance $\{G(x,n)\}$. Pixels with their natural random noise are used as conditional input, which can alleviate the pattern collapse problem [7] of scGANs, and make scGANs generate a good effect of unmixing abundance. Benefitting from the operation on image patches instead of single pixels, our method is able to take advantage of spatial information in HSI. In addition, ReLU and softmax layer is designed to meet abundance non-negativity constraint (ANC), and abundance sum-to-one constraint (ASC), which are two physical constraints of the abundance maps.

Loss function. The *adversarial loss* [5] applied in this model is expressed as:

$$\min_G \max_D V(D,G) \tag{2}$$

$$V(D,G) = E_{\boldsymbol{a}}[\log D_A(\boldsymbol{a})] + E_{\boldsymbol{x}}[\log(1 - D_A(G(\boldsymbol{x},\boldsymbol{n})))] \tag{3}$$

In addition, the true abundances of HSI are sparse. In order to make the generative abundances and the true abundances as close as possible and make the generative abundances sparse, the L_1 sparsity regularization is adopted to obtain sparse solutions. Thus, the full objective is:

$$L(D,G) = V(D,G) + \lambda_{sps} E_{x}[\| G(\boldsymbol{x},\boldsymbol{n}) - \boldsymbol{a}^{*} \|_1] \\ = E_{\boldsymbol{a}}[\log D_A(\boldsymbol{a})] \\ + E_{x}[\log(1 - D_A(G(\boldsymbol{x},\boldsymbol{n})))] \\ + \lambda_{sps} E_{x}[\| G(\boldsymbol{x},\boldsymbol{n}) - \boldsymbol{a}^{*} \|_1] \tag{4}$$

where $\boldsymbol{a}^{*}$ is the true abundances, and λ_{sps} is a parameter for balancing the authenticity of generating abundance and the antagonism of the generator. The choice for λ_{sps} depends on the performance of the discriminator.

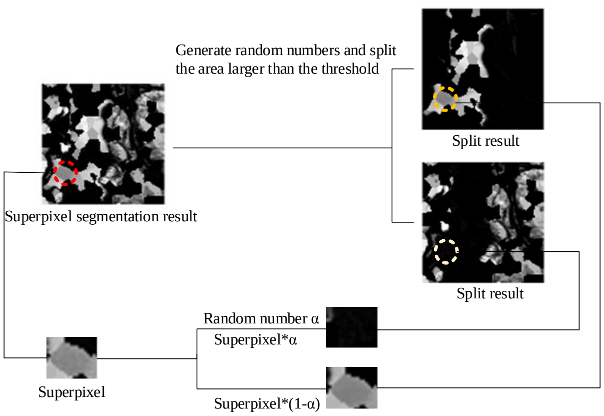

Fig. 2: Superpixel segmentation and random splitting for synthetic data.

2.2. Superpixel Segmentation and Random Splitting

The usual data synthesis method is for a single pixel that ignores the spatial structure of HSI, which makes the synthetic data not real enough. In this paper, we propose a data synthesis method based on superpixel segmentation and random splitting, and the synthesized data obtained by this method have certain spatial structure information. We select a hyperspectral dataset as the base dataset, and unmix the base dataset by the traditional method to obtain the approximate abundance solution, and the approximate abundance solution is segmented by superpixels. With each superpixel as the basic unit, a threshold is defined, and random numbers are generated to decide whether to split and the split ratio, as shown in Fig. 2. The synthetic data are generated by the linear mixing models with the splitting abundance and eight endmembers randomly chosen from the United States Geological Survey (USGS) [8]. Finally, gaussian noise is added with SNR = 30 dB.

3. EXPERIMENTS AND ANALYSIS

To evaluate the performance of the proposed method, simulated and real hyperspectral images were used to illustrate the different unmixing performance. The proposed

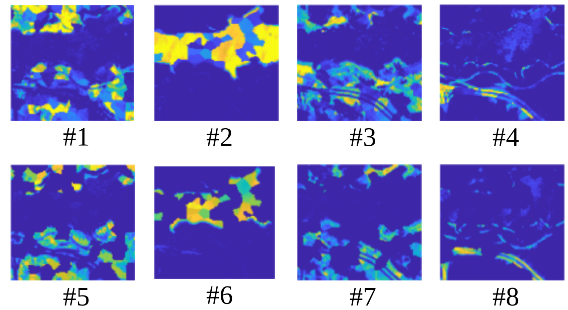

Fig. 3: The generative abundance maps by superpixel segmentation and random splitting.

method was compared with two state-of-the-art unmixing algorithms: SUnSAL, DACE [9], and our cGANs method that without considering spatial information.

The accuracy assessment of all the experiments in this paper was made by the abundance overall root mean square error (aRMSE) [10] as well as abundance angle distance (rmsAAD) [11]. aRMSE measures the numerical similarity between true abundances and the generative ones, while rmsAAD measures dimensional similarity.

A. Implementation Details

Only 5% of the true abundance labels were used to train the network for 200 epochs. The spatial window size was set to 5×5. λ_{sps} was empirically set to 10.

B. Experiments with Simulated Datasets

In the simulated experiments, the synthetic dataset was synthesized using the superpixel segmentation and random splitting method mentioned above. Superpixel segmentation can ensure that the generative abundances have spatial neighborhood structure, which in turn can ensure that the synthetic hyperspectral data have a certain spatial structure between neighboring pixels. Japser Ridge [12] is used as the base data, the size of which is 100×100 pixels and after removing low SNR and water vapor absorption bands (1-3, 108-112, 154-166, and 220-224), there are remaining 198 spectral bands. After splitting, the number of endmembers becomes eight, and the generative abundance maps are shown in Fig. 3.

Table I displays aRMSE and rmsAAD values of estimated abundance maps for simulated datasets. The aRMSE and AAD values of the proposed method is 0.0271 and 0.096 respectively, which validate the superiority of our network architecture.

Table I aRMSE/rmsAAD values of Estimated Abundance Maps for Simulated Datasets. ($\times 10^{-2}$)

Simulated Datasets	SUnSAL	DCAE	cGANs	scGANs
aRMSE	13.87	4.89	9.60	**2.71**
rmsAAD	77.61	16.76	59.62	**9.60**

C. Experiments with Real Hyperspectral Imagery

The Urban dataset [13] used in the experiments is a

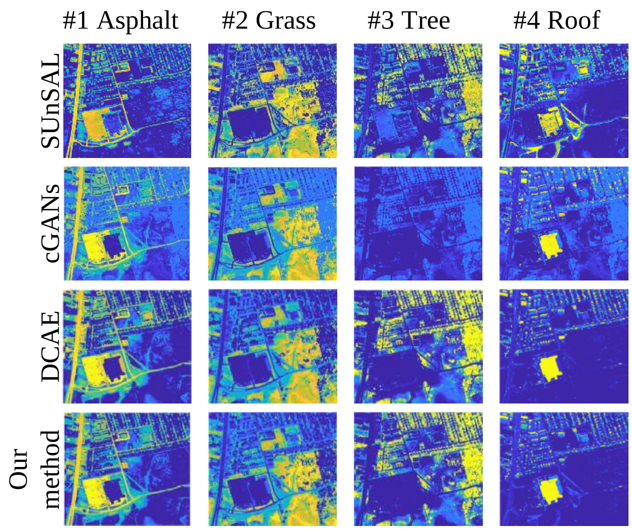

Fig. 4: Estimated abundances for Urban dataset.

widely used HSI for HSU research. It contains 307×307 pixels and 210 spectral bands ranging from 400 to 2500 nm. This data also possess bad bands (1-4, 78, 87, 101-111, 136-153, 198-210) and after removing them, 162 bands remain to be unmixed. There are four endmembers for data: "#1 Asphalt", "#2 Grass", "#3 Tree", and "#4 Roof".

The results obtained are shown in Fig. 4, using SUnSAL, DCAE, cGANs method that without considering spatial information and the proposed scGANs. Compared with cGANs that ignores spatial information, the results obtained by the proposed method are significantly better. As for the advanced traditional unmixing method SUnSAL, our method has better performance in removing background noise. DCAE and the method proposed in this paper are both deep learning unmixing methods, while the latter obtains a smoother abundance distribution.

The quantitative assessment for Urban dataset is given in Table II. The proposed method gains the lowest aRMSE and rmsAAD, which has the best performance.

Table II aRMSE/rmsAAD values of Estimated Abundance Maps for Urban dataset. ($\times 10^{-2}$)

Urban dataset	SUnSAL	DCAE	cGANS	scGANs
aRMSE	16.82	5.32	12.34	**2.29**
rmsAAD	59.26	19.09	49.78	**10.52**

4. CONCLUSIONS

In this paper, we have proposed pixels-to-abundances translation with spatial-spectral conditional generative adversarial networks for hyperspectral unmixing. The method utilizes an adversarial strategy to learn the manifold structure of hyperspectral data and generates the abundances of the central pixels from pixel blocks with spatial information to complete the unmixing. We evaluated the proposed method using synthetic data as well as the real hyperspectral data and compared it to several state-of-the-art methods. The spatial-spectral cGANs was proved to outperform all the comparison methods in the experiments.

5. REFERENCES

[1] M. Iordache, J.M.B. Dias and A. Plaza, "Collaborative sparse regression for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 1, pp. 341-354, 2014.
[2] A. Halimi, Y. Altmann, N. Dobigeon, and J. Tourneret, "Nonlinear unmixing of hyperspectral images using a generalized bilinear model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4153-4162, 2011.
[3] R. Heylen and P. Scheunders, "A multilinear mixing model for nonlinear spectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 1, pp. 240-251, 2016.
[4] B. Koirala, R. Heylen, and P. Scheunders, "A neural network method for nonlinear hyperspectral unmixing," in *2018 IEEE International Geoscience and Remote Sensing Symposium*, 2018, pp. 4233-4236.
[5] I. Goodfellow, J.P. Abadie, M. Mirza, B. Xu, D.W. Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, 2014, vol. 27, pp. 2672–2680.
[6] M. Wang, M. Zhao, J. Chen, and S. Rahardja, "Nonlinear unmixing of hyperspectral data via deep autoencoder networks," in *2019 IEEE International Geoscience and Remote Sensing Symposium*, 2019, pp. 1467-1471.
[7] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," in *Advances in Neural Information Processing Systems*, 2016, vol. 29, pp. 2234–2242.
[8] R.N. Clark, G.A. Swayze, T.V.V. King, A.J. Gallagher, and W.M. Calvin, "The U.S. geological survey digital spectral library: Version 1: 0.2 to 3.0 microns," 1993.
[9] F. Khajehrayeni and H. Ghassemian, "Hyperspectral unmixing using deep convolutional autoencoders in a supervised scenario," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 567–576, 2020.
[10] Y. Zhong, X. Wang, L. Zhao, R. Feng, L. Zhang, and Y. Xu, "Blind spectral unmixing based on sparse component analysis for hyperspectral remote sensing imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 119, pp. 49-63, 2016.
[11] L. Miao, H. Qi, and H. Szu, "A maximum entropy approach to unsupervised mixed-pixel decomposition," *IEEE Transactions on Image Processing*, vol. 16, no. 4, pp. 1008-1021, 2007.
[12] B. Palsson, J.R. Sveinsson, and M.O. Ulfarsson, "Multitask learning for spatial-spectral hyperspectral unmixing," in *2019 IEEE International Geoscience and Remote Sensing Symposium*, 2019, pp. 564-567.
[13] W. He, H. Zhang, and L. Zhang, "Hyperspectral unmixing using total variation regularized reweighted sparse non-negative matrix factorization," in *2016 IEEE International Geoscience and Remote Sensing Symposium*, 2016, pp. 7034-7037.