



HAL
open science

Hyperspectral Classification Based on Spectral Indices Learned Through Soft Attention Units

Romain Thoreau, Veronique Achard, Xavier Briottet

► **To cite this version:**

Romain Thoreau, Veronique Achard, Xavier Briottet. Hyperspectral Classification Based on Spectral Indices Learned Through Soft Attention Units. IGARSS 2021 - 2021 IEEE International Geoscience and Remote Sensing Symposium, Jul 2021, Brussels, Belgium. pp.2544-2547, 10.1109/IGARSS47720.2021.9554256 . hal-03417120

HAL Id: hal-03417120

<https://hal.science/hal-03417120v1>

Submitted on 5 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

HYPERSPECTRAL CLASSIFICATION BASED ON SPECTRAL INDICES LEARNED THROUGH SOFT ATTENTION UNITS

Romain Thoreau^{*,1,2}, Véronique Achard¹ and Xavier Briottet¹

¹ ONERA, The French Aerospace Lab, 2 Av. Edouard Belin, 31055 Toulouse, France

² Magellium, 1 rue Ariane, 31520 Ramonville Saint-Agne, France

ABSTRACT

Classification of hyperspectral scenes has been dominated in recent years by Convolutional Neural Networks (CNNs). Spectral and spatial convolutions have proven to be very effective in learning discriminative representations. The power of CNNs however is limited by their small receptive fields. Training deeper CNNs with wider receptive fields is a very difficult task with regard to the small amount of available training samples. Furthermore, CNNs seem to fail in capturing very local spectral features such as absorption peaks. In the present paper, we introduce a new paradigm inspired by physics-based hyperspectral indices and attention mechanisms. Our model learns spectral indices by focusing on specific spectral bands through soft attention units. It achieves high or better overall accuracy and kappa score than state-of-the-art CNNs on Pavia University, Kennedy Space Center and our own real-world dataset while dramatically reducing the number of parameters needed and increasing interpretability.

Index Terms— Hyperspectral classification, spectral indices, attention mechanism

1. INTRODUCTION

In recent years, machine learning methods have been performed successfully in the field of remote sensing. Hyperspectral imagery especially has raised a growing interest in computer vision methods to monitor land use and land cover. A recent comparative review on hyperspectral (HS) classification has shown that CNNs outperform conventional machine learning models [1]. They can indeed deal very well with the high dimensionality of the data and highlight discriminative spectral characteristics. CNNs can also embed simultaneously spectral and spatial features in a common representation space with higher class separability.

However, the weakness of CNNs holds in their narrow receptive fields [2]. In order to learn spectral long-range dependencies, CNNs must either go deeper or use fully-connected layers that are driving up the number of parameters. Both solutions are not optimal with regard to the availability of labeled

training samples in remote sensing. The high ratio of the number of the models parameters over the number of training sample raises a high concern for generalization capabilities.

To mitigate this issue, regularization techniques are widely used even though they rely on the tuning of penalization coefficients. Furthermore, attention mechanisms have recently generated a lot of interest in computer vision as an alternative to CNNs [2]. Some attempts to introduce spectral attention in hyperspectral classifiers have been made. Liu et al. introduced a model that computes attention vectors on spectral groups through fully-connected layers with a sigmoid activation [3]. If the use of a softmax unit would have led to more parsimonious attention, it also causes vanishing gradients due to the high spectral dimension of the data. Haut et al. include spatial-spectral attention masks within a residual CNN. Those masks are computed by residual blocks with downsampling and upsampling operations to increase the size of the attention receptive field [4].

Meanwhile, hyperspectral indices based on chemical knowledge of materials are well suited to focus on specific bands in order to highlight some spectral characteristics. For instance, they can easily discriminate plastics based on specific absorption peaks [5]. As far as spectral indices rely on the chemical composition of materials, they can be used for various sensors and geographic areas.

In order to circumvent the problems raised by CNNs, we introduce a new model that statistically learns hyperspectral indices through an attention mechanism. Unlike the previous work on spectral attention, our model is more straightforward, needs much less parameters and leads to more parsimonious attention, at the expense, however, of a lower model capacity.

2. METHOD

In this section, we present the architecture of our model. We argue that spectral indices, ie. linear combination of bands are well suited to provide discriminative and easily interpretable features for classification. This is actually what fully-connected neural networks (FNNs) do and one could interpret the first hidden layer of FNNs as spectral indices. FNNs although do not perform as well as CNNs on classification tasks [1] and their physics-based interpretation is

* Corresponding author: romain.thoreau@onera.fr

hard even with additional parsimony constraints. This is why we introduce a naturally parsimonious model with very few parameters that leads to clear spectral attention.

We denote our model by f and the spectrum of a given pixel by a $B \times 1$ vector \mathbf{s} where B is the number of spectral bands. The output of the model $f(\mathbf{s})$ is a probability distribution over the C classes:

$$f(\mathbf{s}) = \text{Softmax}(W^T I(\mathbf{s}) + \mathbf{b}) \quad (1)$$

where W is a $N \times C$ matrix of weights, \mathbf{b} is a $C \times 1$ bias and $I(\mathbf{s})$ is a $N \times 1$ matrix of N hyperspectral indice (HSI) unit defined as:

$$I_n(\mathbf{s}) = \text{ReLU}\left(\sum_{i=1}^{k_n} w_i S_i(\mathbf{s})\right) \quad \forall n \in \{1, \dots, N\} \quad (2)$$

where w_i are learnable coefficients and $S_i(\mathbf{s})$ are soft-band attention (SBA) units defined as the dot product of a spectrum \mathbf{s} and a $B \times 1$ attention vector \mathbf{a}_i :

$$S_i(\mathbf{s}) = \mathbf{a}_i^T \mathbf{s} \quad (3)$$

The attention vector \mathbf{a}_i is parametrized by a learnable parameter μ_i which defines where to focus and by a hyperparameter σ which defines the size of the spectral domain of attention:

$$\mathbf{a}_i = \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(b-\mu_i)^2}{2\sigma^2}} \right)_{b \in \{1, \dots, B\}} \quad (4)$$

The complete architecture of the model is illustrated in figure 1.

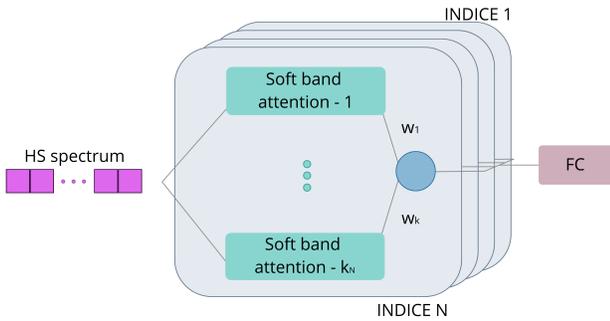


Fig. 1. Schematic view of our HSI learning model

3. NUMERICAL EXPERIMENTS

The model is implemented in PyTorch and released on GitHub upon publication¹. We have added our model to the DeepHyperX toolbox [1].

¹<https://github.com/Romain3Ch216/DeepHyperX>

3.1. Datasets

We conduct experiments on three hyperspectral datasets: Pavia University, Kennedy Space Center and our own dataset denoted as Mauzac. The dataset will be available on request upon publication. Sensors spectral range, ground sampling distance (GSD) and available labeled data are summarized in table 2. False color RGB images are shown on table 1.

Pavia University is one of the main reference hyperspectral dataset acquired by ROSIS sensor. We use the radiance hypercube and the standard train/test split provided by the IEEE GRSS DASE website [6]. Those train and test sets are more independent than a random train/test split as far as one geo-entity does not belong to both sets. Thus, this ground truth is more likely to evaluate the generalization capability of the models to new geographic areas.

The **Kennedy space center** (KSC) dataset is acquired by the sensor AVIRIS. The training, validation and test sets are selected from disjoint parts of the image. About 20% of the available samples are used for training, 5% for validation and 75% for evaluation.

The **Mauzac dataset** was acquired by the sensor AISA-FENIX. Atmospheric correction using COCHISE software [7] is applied on the image in order to convert the radiance hypercube to reflectance. Water absorption bands as well as noisy bands are removed, resulting in 257 reliable bands. We build the ground truth of Mauzac with the same degree of rigour than the DASE benchmarking initiative. For the learning and the test sets, we select geo-entities with different illuminations, if possible, and of a slightly different natures.

3.2. Comparison with state-of-the-art models

We compare our model with the state-of-the-art hyperspectral classifiers identified by Audebert et al. in their 2019 comparative review [1].

Spectral CNN - Hu et al. [8] The CNN of Hu et al. learns a collection of 1D convolutions, operates a max-pooling operation and feeds the resulting vector in a 2-layer fully-connected neural network. The kernel size is set according to an empirical rule from the original paper to 12 for Pavia University, to 25 for Kennedy Space Center and to 29 for Mauzac. Thus, the convolutions are only able to learn local middle-range dependencies.

Spectral-spatial CNN - Li et al. [9] The CNN of Li et al. has two layers of 3D convolutions. The kernel size of the first and second convolutions are respectively $7 \times 3 \times 3$ and $3 \times 3 \times 3$ for every datasets (as in the original paper). The resulting flattened vector then feeds one fully-connected layer. As well as Hu et al. model, those 3D convolutions cannot learn spectral long-range dependencies.



Table 1: False color RGB images. From left to right: Pavia University, Kennedy Space Center and Mauzac.

Pavia University		Kennedy Space Center		Mauzac	
430 - 860nm — 1.3m GSD		400 - 2500nm — 18m GSD		420 - 2500nm — 0.55m GSD	
Classes	Samples	Classes	Samples	Classes	Samples
Self-blocking bricks	514	Scrub	144	Vegetation shadows	455
Meadows	540	Willow swamp	48	High vegetation	402
Gravel	392	Cabbage palm hammoc	57	Ground Vegetation	1293
Shadows	231	Cabbage palm/oak hammock	47	Dry vegetation	786
Bitumen	375	Slash pine	39	Bare Soil	229
Bare Soil	532	Oak / broadleaf hammock	46	Water body	452
Painted metal sheets	265	Hardwood swamp	26	Pool	150
Asphalt	548	Graminoid marsh	82	Pool cover	105
Trees	524	Spartina marsh	101	Curbstone	176
		Cattail marsh	79	Tile	1448
		Salt marsh	77	Asphalt	474
		Mud flats	96	Other shadows	331
		Water	184		

Table 2: Spectral range, GSD and number of training samples per class

3.3. Model hyperparameters and training

The models are trained on a machine with a single Nvidia GPU and on an 8-core Intel i7 CPU.

For every datasets, metrics on the validation sets show that best standard deviation for every SBA units is 0.5, leading to sharp spectral attention. For the Mauzac dataset, 20 HSI units with 3 SBA units are used. For the Kennedy Space Center dataset, 20 HSI units with 3 SBA units, 10 HSI units with 10 SBA units and 10 HSI units with 20 SBA units are used. For the Pavia University dataset, 80 HSI units with 10 SBA units are used. Moreover, for materials that are difficult to discriminate, indices are learned from 1 VS 1 classifications, transfered and freed during the training. 10 HSI units with 10 SBA units are specialized in discriminating meadows from trees and asphalt from bitumen. 10 HSI units with 20 SBA units are specialized in discriminating self-blocking bricks from gravel and meadows from bare soil. Those classes were chosen since they are often confused. Every models are trained with standards stochastic gradient descent algorithms (SGD) according to the cross-entropy loss. Both state-of-the-art CNNs are trained by the standard SGD with momentum.

For our model, we use the Adam optimizer with a 0.1 learning rate for the attention means and a 0.01 learning rate for the other parameters. Hyperparameters are tuned on a validation set.

3.4. Results

As shown on table 3, our model outperforms the 1D CNN of Hu et al. on the Mauzac and the KSC datasets and the 3D CNN of Li et al. on the KSC dataset with up to 198 times less parameters than the state-of-the-art models.

Our model succeeds in learning long-range spectral correlations between bands in the VNIR (visible and near infrared) and in the SWIR (short-wave infrared). An interesting property illustrated by fig. 2 is that only few indices are activated for a given class which facilitates the interpretation of learned indices and their comparison with state-of-the-art spectral indices such as NDVI.

Lower classification performances on the Pavia University dataset may be explained by a too low model capacity and too few non-linearities. The main confusions are between meadows / trees and between meadows / bare soil. In con-

trary, the 3D CNN takes advantage of the textural information to discriminate spectrally close classes. On the KSC dataset however, the spatial resolution is too coarse to extract textural information, leading to poor results with the 3D CNN.

Mauzac			
Model	OA	Kappa	Nbr of parameters
Hu et al.	72.8 ±1.0	0.70 ±0.012	78 013
Li et al.	81.6 ±2.4	0.80 ±0.026	120 141
HSI (ours)	78.9 ±0.97	0.76 ±0.011	393

Pavia University			
Model	OA	Kappa	Nbr of parameters
Hu et al.	81.2 ±2.0	0.76 ±0.023	61 370
Li et al.	84.3 ±0.72	0.80 ±0.0090	46 570
HSI (ours)	78.6 ±2.7	0.74 ±0.028	3210

Kennedy Space Center			
Model	OA	Kappa	Nbr of parameters
Hu et al.	90.3 ±0	0.89 ±1.4e-5	79 934
Li et al.	77.4 ±8.8	0.75 ±0.099	91 950
HSI (ours)	90.5 ±0.47	0.90 ±0.0052	1294

Table 3: Experimental results on the three datasets. Mean and standard deviation of three experiments are shown. Experimental results of Hu et al. and Li et al. on IEEE GRSS DASE Pavia University dataset are taken from Audebert et al.

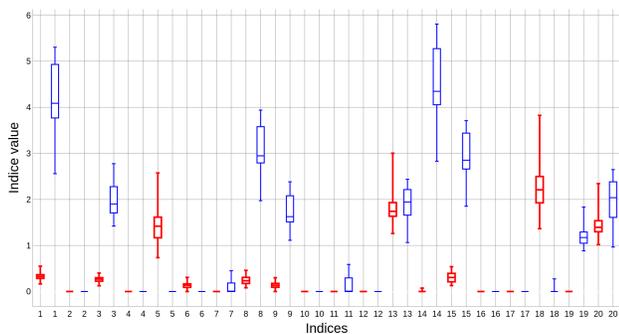


Fig. 2. Box plots for high vegetation spectra (in bold red) and for tile spectra (in thin blue) of Mauzac dataset. Extreme values are shown by the horizontal lines.

4. CONCLUSION

In this paper, we have introduced a statistical model that can learn long-range spectral dependencies with very few parameters by focusing on specific spectral bands. Experiments show that this new model can outperform state-of-the-art convolutional neural networks. In future work, we would like to assess the benefits of the low number of trainable parameters against the model low capacity on wider geographic areas. Besides, we would like to assess the capability of the learned

indices to be transferred to other sensors with different spectral or spatial resolutions.

5. ACKNOWLEDGMENT

We thank IEEE GRSS Image Analysis and Data Fusion Technical Committee for providing the DASE benchmarking website used to test our model performance on the Pavia University dataset. We thank Prof. Paolo Gamba from the Telecommunications and Remote Sensing Laboratory, Pavia University (Italy) for providing the Pavia dataset.

6. REFERENCES

- [1] N. Audebert, B. Le Saux, and S. Lefèvre, “Deep learning for classification of hyperspectral data: A comparative review,” *IEEE Geoscience and Remote Sensing Magazine*, vol. 7, no. 2, pp. 159–173, June 2019.
- [2] Niki Parmar, Ashish Vaswani, Jakob Uszkoreit, Lukasz Kaiser, Noam Shazeer, Alexander Ku, and Dustin Tran, “Image transformer,” 2018.
- [3] Q. Liu, Z. Li, S. Shuai, and Q. Sun, “Spectral group attention networks for hyperspectral image classification with spectral separability analysis,” *Infrared Physics and Technology*, vol. 108, pp. 103340, August 2020.
- [4] J. M. Haut, M. E. Paoletti, J. Plaza, A. Plaza, and J. Li, “Visual attention-driven hyperspectral image classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 10, pp. 8065–8080, 2019.
- [5] Alexandre Alakian and Véronique Achard, “Classification of hyperspectral reflectance images with physical and statistical criteria,” *Remote Sensing*, vol. 12, pp. 2335, 07 2020.
- [6] “Ieee grss dase website,” <http://dase.grss-ieee.org/>.
- [7] C. Miesch, L. Poutier, V. Achard, X. Briottet, X. Lenot, and Y. Boucher, “Direct and inverse radiative transfer solutions for visible and near-infrared hyperspectral imagery,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 7, pp. 1552–1562, 2005.
- [8] Wei Hu, Yangyu Huang, Li Wei, Fan Zhang, and Hengchao Li, “Deep convolutional neural networks for hyperspectral image classification,” *Journal of Sensors*, vol. 2015, pp. 1–12, 07 2015.
- [9] Ying Li, Haokui Zhang, and Qiang Shen, “Spectral–spatial classification of hyperspectral imagery with 3d convolutional neural network,” *Remote Sensing*, vol. 9, no. 1, pp. 67, Jan 2017.