

Autoencodeurs pour la visualisation d'images hyperspectrales

Alexandre BOULCH¹, Nicolas AUDEBERT^{1,2} Dominique DUBUCQ³

¹ONERA - *The French Aerospace Lab*, FR-91761 Palaiseau, France

²Univ. Bretagne-Sud, UMR 6074, IRISA, F-56000 Vannes, France

³Total, CSTJF, 6 Avenue Larribau, 64018 Pau, France

Résumé – La visualisation de données de grandes dimensions telles que les données hyperspectrales est un élément clé en vue de leur analyse. La représentation des données sous la forme d'une image RGB, facilement utilisable, est un problème récurrent. Une telle compression est habituellement effectuée en utilisant des a priori ou des méthodes non supervisées pouvant parfois conduire à une perte d'information. Nous proposons dans ce travail de compresser les données à l'aide d'autoencodeurs permettant de produire un signal codé sur trois canaux. Nous montrons qu'il est possible d'étendre cette compression avec l'ajout d'annotations pour faire ressortir les éléments désirés de la scène.

Abstract – Visualization of high dimensional data such as hyperspectral data is a key step for analysis. A recurrent issue is to represent the data as a RGB image for easy display. This compression is usually obtained using priors on the data or unsupervised meaningful information extraction which may result in information loss. Rather than making any assumption on the type of data, we propose to compress the data into a three channel image such that it is possible to reconstruct the original. For this, we use deep convolutional autoencoders in the spectrum dimension. We also show that our approach can be extended to take external knowledge into account such as visual hints for color display.

1 Introduction

Dans le cadre de la collaboration Naomi entre Total et l'Onera, nous tentons de détecter et identifier la présence d'hydrocarbures naturels ou accidentels sur des images hyperspectrales. Ces hydrocarbures présents en surface ou proche surface peuvent être détectables directement ou s'exprimer par l'impact qu'ils ont sur la végétation ou la minéralogie de surface. Mais même lorsque l'hydrocarbure est présent, la réponse d'un mélange intime entre l'hydrocarbure et le sol de surface est largement non-linéaire; l'expression de l'hydrocarbure dans les spectres enregistrés est ténue, rendant largement inopérant l'utilisation d'indices tels que le NDVI [12] ou le HI [7]. Il est donc important dans l'exploitation de la donnée hyperspectrale de pouvoir extraire l'information structurante de l'image tout en conservant les signaux faibles qui nous intéressent.

Les images hyperspectrales (HI) contiennent jusqu'à plusieurs centaines d'échantillons spectraux par pixel du visible à l'infrarouge. Ce grand nombre d'échantillons en fait un outil utile pour l'identification, la classification [10] ou la caractérisation des éléments de la scène observée comme le stress de la végétation [6]. Cependant, le grand nombre de bandes spectrales mène à des problèmes de visualisation, les outils d'affichage, tels que les écrans d'ordinateur, ne permettant de visualiser que trois bandes. Le problème est de transférer autant d'information que possible de la matrice 3D originale à une image RGB, de manière structurée de sorte que l'utilisateur comprenne cette information. Lorsque des informations additionnelles sont disponibles (lieu, objectifs de visualisation)

il est possible de sélectionner trois bandes ou combinaisons de bandes a priori intéressantes pour, par exemple, mettre en exergue les caractéristiques d'un matériau (par exemple, le NDVI pour la végétation ou l'indice d'hydrocarbures), mais ce choix fixe des paramètres peut mener à une perte d'information (seuils mal réglés, mauvais choix de bande) et donc un défaut de visualisation. Les méthodes non supervisées, visent à créer une représentation de l'espace adaptée à la scène observée et à sélectionner trois directions contenant le plus d'information utile. Parmi ces méthodes, on compte le *spectral clustering* [9]; l'analyse en composantes principales (PCA) [11]; la *Minimum noise fraction* [2] opérant deux PCAs sur les données. L'analyse en composantes indépendantes (ICA) [13]; ou encore la *1bit transform* [5] sélectionnant les bandes les plus significatives.

Notre objectif est de compresser le signal spectral pour produire une image interprétable. Notre méthode utilise des réseaux de neurones profonds avec une architecture d'autoencodeur. Le réseau est entraîné à reconstruire le signal original, assurant ainsi que le signal codé contient autant d'information que possible, les phénomènes peu énergétiques n'étant ainsi pas oubliés. De plus, nous proposons une extension pour exploiter des annotations expertes disponibles, comme des zones d'intérêt ou des indications de couleur.

2 Autoencodeurs pour la visualisation

Les autoencodeurs sont des réseaux de neurones utilisés en apprentissage non supervisé [1]. Ceux-ci ont déjà été utilisés

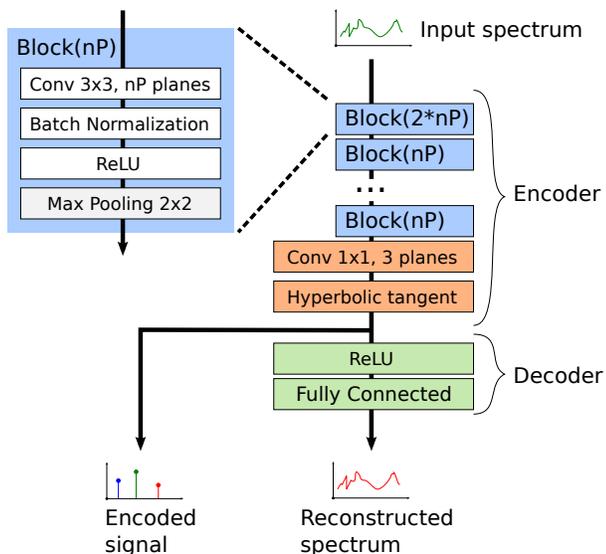


FIGURE 1 – Réseau pour la visualisation.

pour la classification hyperspectrale [3], les modèles convolutifs (CNN) obtenant d'excellents résultats [14]. Afin d'en améliorer les performances, plusieurs travaux ont proposé d'ajouter des contraintes sur les poids du réseaux comme la non négativité [4], ou encore une contrainte de parcimonie [8].

Dans ce travail, nous ne contraignons que les signaux codés et décodés mais il est envisageable de l'étendre en ajoutant de telles contraintes. Notre fonction objectif \mathcal{L} est la suivante :

$$\mathcal{L}_B = E_{\text{rec}} + \lambda E_{\text{code}} \quad (1)$$

où E_{rec} est le terme de reconstruction imposant une sortie similaire à l'entrée, ici une erreur quadratique moyenne :

$$E_{\text{rec},B} = \text{MSE}_B(x, \hat{x}) = \frac{1}{n|B|} \sum_{\mathbf{x} \in B} \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 \quad (2)$$

avec B l'ensemble des spectres d'entrée, \mathbf{x} un spectre de dimension n et $\hat{\mathbf{x}}$ le signal reconstruit. Le terme de contrainte E_{code} exprime la volonté d'utiliser l'ensemble des couleurs disponibles à l'affichage et ainsi interdire la production d'une image uniforme excepté pour certains pixels de grande énergie (par exemple figure 2, PCA). Notons $c(\mathbf{x})$ le signal codé correspondant à l'entrée \mathbf{x} :

$$E_{\text{code},B} = \exp \left(\frac{-\alpha}{|B|^2} \sum_{(\mathbf{x}_1, \mathbf{x}_2) \in B^2} \|c(\mathbf{x}_1) - c(\mathbf{x}_2)\|_2^2 \right) \quad (3)$$

L'architecture du réseau est présentée figure 1. L'encodeur est composé d'une pile de blocs (convolution + normalisation + ReLU + Max Pooling) réduisant de moitié la dimension du signal. La taille du voisinage spatial détermine le nombre de blocs nécessaires dans l'encodeur pour obtenir un signal 1D le long de l'axe spectral. Puis, le signal est filtré par une tangente hyperbolique pour éliminer les valeurs extrêmes. Le décodeur est, quant à lui, composé d'une couche de neurones entièrement connectés permettant de reconstruire le signal d'entrée.

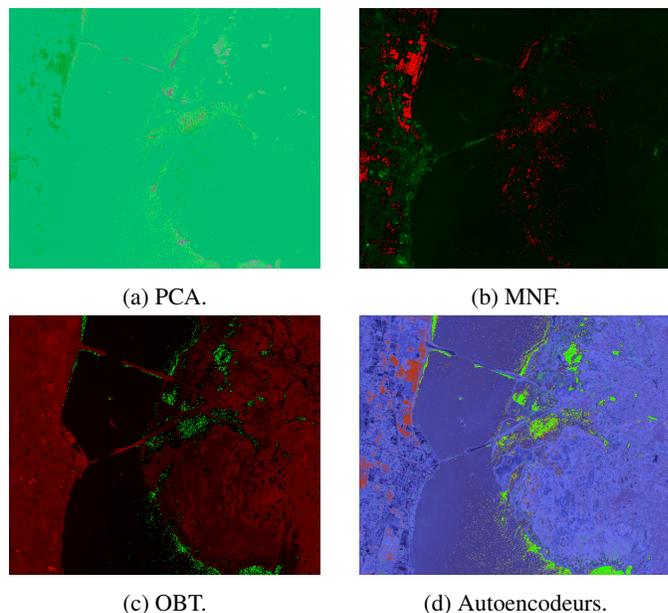


FIGURE 2 – Méthodes de visualisation, Kennedy Space center.

Expériences : Dans nos expériences nous utilisons $\lambda = 0.01$. La figure 2 présente une comparaison entre notre méthode et les algorithmes de la littérature : PCA, MNF et OBT. Les autoencodeurs sont moins sujets à l'influence des pixels très énergétiques, le résultat étant plus contrasté et les détails mieux visibles. De plus, notre approche n'a pas de direction privilégiée et répartit le signal équitablement sur les trois canaux colorés, contrairement à la PCA ou la MNF. Le tableau 1 récapitule des temps de calcul pour différentes images hyperspectrales et méthodes. La PCA et la MNF utilisent la librairie Spectral Python, tandis que la OBT a été recodée en Python, sans optimisation particulière, expliquant sa lenteur relative. Notre méthode est compétitive pour ce qui est du temps de prédiction. Dans le cas où le réseau n'a pas été pré-entraîné, il faut ajouter un temps d'entraînement (2 à 3 minutes avec une carte graphique NVIDIA K2200). Si de nouvelles images issues du même capteur (même nombre de bandes et même comportement dynamique) sont acquises, il est possible de ré-utiliser un réseau pré-entraîné et d'obtenir un résultat cohérent sur les toutes les données. La figure 3 illustre l'utilisation d'un réseau entraîné en premier lieu sur Pavia University puis utilisé pour visualiser Pavia Centre, sans ré-entraînement.

Utilisation de données additionnelles

Des données additionnelles comme des annotations ou une vérité terrain partielle peuvent être disponibles. Dans ce cas, il est intéressant de pouvoir inclure ces informations supplémentaires pour pouvoir améliorer la sortie en guidant l'entraînement. La flexibilité des réseaux de neurones permet l'inclusion d'information externe. Nous illustrons cette possibilité en guidant l'optimisation avec des informations colorimétriques. L'utilisateur annote l'image et nous imposons au signal codé

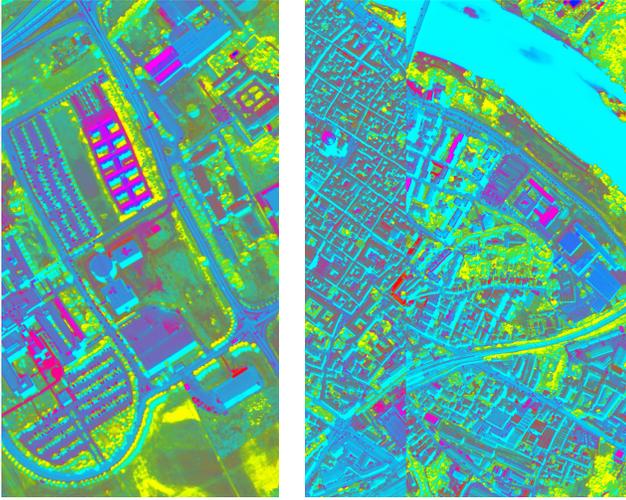


FIGURE 3 – Utilisation d’un modèle pré-entraîné. Données d’entraînement (gauche) et prédictions (droite).

Scène	Pixels	PCA	MNF	OBT	AutoEnc.
Indian Pines	21k	0.28	0.27	4.7	0.35
Kennedy SC.	314k	2.2	2.6	90	3.8
Pavia Univ.	207	0.24	0.26	4.9	1.6
Botswana	378	1.9	2.2	82.4	3.7
Pavia Centre	784	2.6	2.7	97	5.9

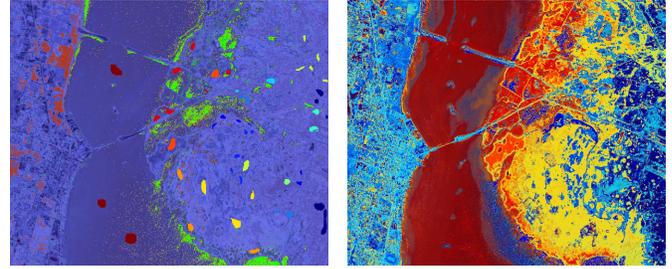
TABLE 1 – Temps de calculs (secondes)

d’être similaire à ces annotations, comme illustré sur la figure 4b. Contrairement à l’équation (3) définie par une fonction décroissante afin d’utiliser l’ensemble de l’espace des couleurs, nous définissons ici un nouveau terme de régression sur la couleur pénalisant les fortes déviations vis-à-vis de l’annotation. Soit e le signal codé à un pixel donné et a la couleur souhaitée pour ce pixel. Le terme de contrainte de l’équation (3) devient un terme d’erreur quadratique sur la couleur :

$$E_{\text{coded}, B_a} = \frac{1}{|B_a|} \sum_{B_a} \frac{1}{n} \sum_j (e_j - a_j)^2 \quad (4)$$

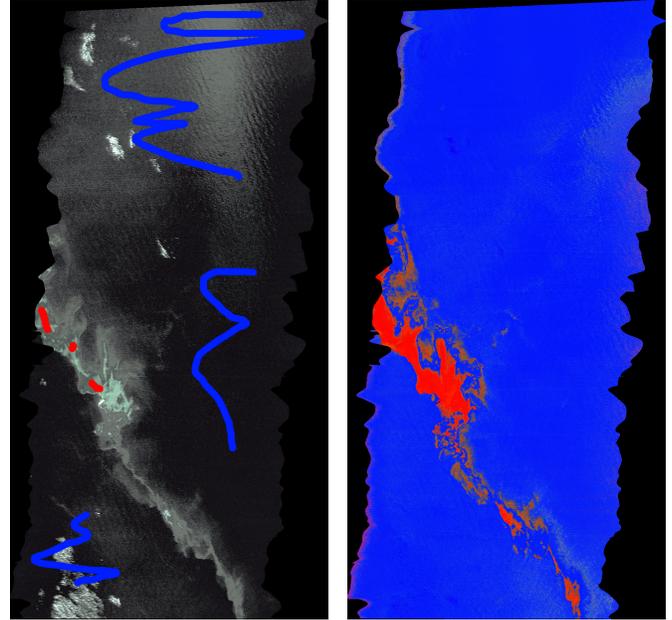
où B_a est l’ensemble des spectres d’entrée pour lesquels une annotation est donnée et e_j (resp. a_j) sont les coordonnées de e (resp. a).

Expériences : Des résultats de visualisation avec annotations partielles sont donnés dans les figures 4 et 5. La première utilise les numéros de classe fournis avec les données, tandis que la seconde illustre une annotation manuelle approximative. Ces indications colorées ne tiennent donc pas compte d’une possible proximité spectrale entre deux éléments (deux types de végétation peuvent avoir des couleurs totalement différentes). Toutes les expériences ont été effectuées avec $\lambda = 100$. Comme attendu, les annotations sont prises en compte dans la carte finale et les résultats sont visuellement cohérents, y compris pour des annotations rares, comme pour la zone d’intérêt (rouge) de la figure 4b, sous-représentée par rapport au reste (bleu).



Autoencodeur + annotations (gauche), résultat guidé (droite).

(a) Kennedy Space Center



Les annotations manuelles définissent une zone d’intérêt (rouge) et le reste (bleu).

(b) Deepwater oil spill.

FIGURE 4 – Annotations et résultats de l’entraînement guidé.

3 Limitations et perspectives

La principale limitation est la non reproductibilité des résultats : les poids du réseaux de neurones sont initialisés aléatoirement et la descente de gradient stochastique induit une optimisation différente à chaque entraînement. Ce phénomène est gommé lors de l’utilisation d’annotations : les couleurs étant partiellement imposées, la variabilité de colorimétrie est réduite. L’une des perspectives liées à ce travail est l’inclusion d’a priori physiques. En effet, rien ne garantit que le réseau va apprendre et faire ressortir des propriétés physiques pertinentes. Ainsi, il est envisageable d’imposer la forme des convolutions afin de rechercher les bandes d’absorption d’un matériau particulier. En outre, l’utilisation de convolutions à petits noyaux et non de couches entièrement connectées favorise la recherche de relations locales mais n’exploite pas la pleine dimension spectrale des données hyperspectrales. Une autre piste d’évolution de ce travail consisterait à appliquer cette méthode pour la clas-

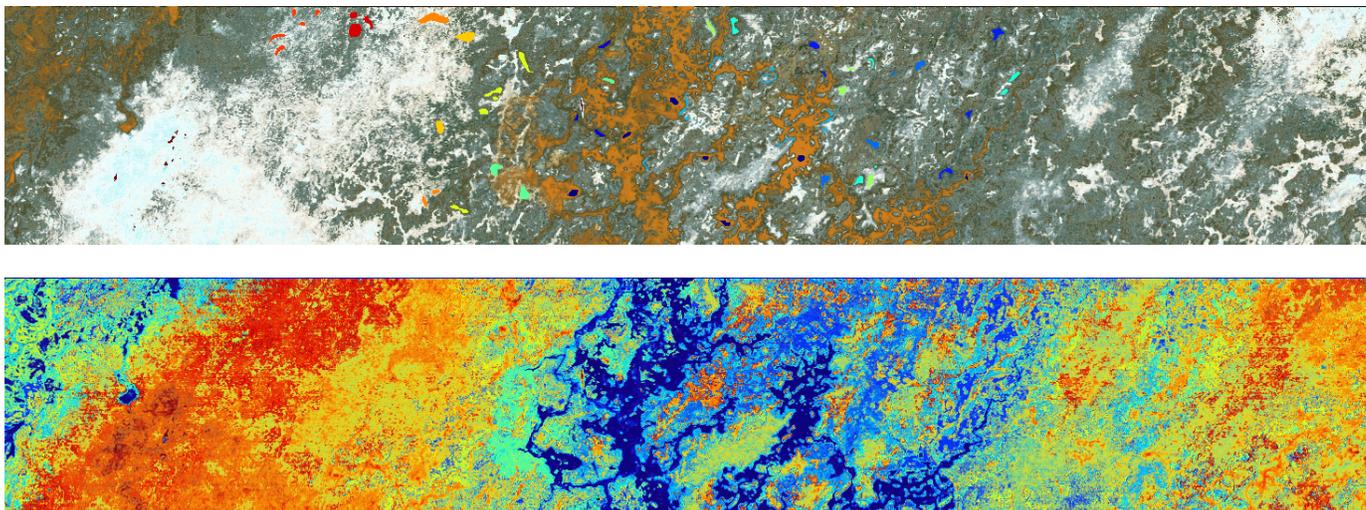


FIGURE 5 – Botswana.

sification semi-supervisée et ainsi utiliser l'espace de représentation appris pour la propagation des labels colorés.

4 Conclusion

Nous avons présenté une méthode pour la visualisation de données hyperspectrales utilisant les réseaux de neurones convolutifs autoencodeurs. La flexibilité des réseaux de neurones a permis de minimiser une fonction objectif comprenant un terme d'attache aux données, pénalisant la perte d'information lors de la compression, et un terme de contrainte. Ce dernier impose soit un étalement des couleurs utilisées pour la représentation des données dans le cas non supervisé, soit le respect des annotations expertes fournies par l'utilisateur. La même architecture de réseau a été utilisée avec succès dans les deux cas d'étude présentés dans cet article.

Remerciements Ce travail s'inscrit au sein du projet MEDUSA (w3.onera.fr/medusa/) pour le Big Data en télédétection pour l'Observation de la Terre. Les travaux de Nicolas Audebert sont financés par le projet NAOMI, projet de recherche ONERA-TOTAL.

Références

- [1] Y. Bengio. Learning deep architectures for ai. *Foundations and trends in Machine Learning*, 2(1) :1–127, 2009.
- [2] J. Boardman, F. Kruse, and R. Green. Mapping target signatures via partial unmixing of aviris data. *5th JPL Airborne, Earth Science Workshop*, 1 :23–26, 1995.
- [3] Yushi Chen, Zhouhan Lin, Xing Zhao, Gang Wang, and Yanfeng Gu. Deep learning-based classification of hyperspectral data. *IEEE JSTARS*, 7(6) :2094–2107, 2014.
- [4] Jan Chorowski and Jacek M Zurada. Learning understandable neural networks with nonnegative weight constraints. *IEEE transactions on neural networks and learning systems*, 26(1) :62–69, 2015.
- [5] J. Feng, K-T. Lo, H. Mehrpour, and AE Karbowskiak. Adaptive block matching motion estimation algorithm using bit-plane matching. In *IEEE ICIP*, 1995.
- [6] R. Hong-Yan, Z. Da-Fang, P. Jian-Jun, S. Xue-Zheng, and W. Hong-Jie. Hyper-spectral remote sensing to monitor vegetation stress. *Journal of Soils and Sediments*, 8(5) :323–326, 2008.
- [7] F Kühn, K Oppermann, and B Hörig. Hydrocarbon index—an algorithm for hyperspectral detection of hydrocarbons. *IJRS*, 25(12) :2467–2473, 2004.
- [8] Andre Lemme, René Felix Reinhart, and Jochen Jakob Steil. Efficient online learning of a non-negative sparse autoencoder. In *ESANN*, 2010.
- [9] A. Martínez-Usó, Martínez-Usó, F. Pla, J.M. Sotoca, and P. García-Sevilla. Clustering-based hyperspectral band selection using information measures. *IEEE TGRS*, 45(12) :4158–4171, 2007.
- [10] F. Melgani and L. Bruzzone. Classification of hyperspectral remote sensing images with support vector machines. *IEEE TGRS*, 42(8) :1778–1790, 2004.
- [11] J. A Richards and JA Richards. *Remote sensing digital image analysis*, volume 3. Springer, 1999.
- [12] J. W. Rouse, R. H. Haas, J. A. Schell, and D. W. Deering. Monitoring vegetation systems in the Great Plains with ERTS. *NASA 3d ERTS-1 Symposium*, 1974.
- [13] J. Wang and C-I. Chang. Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis. *IEEE TGRS*, 44(6) :1586–1600, 2006.
- [14] Jun Yue, Wenzhi Zhao, Shanjun Mao, and Hui Liu. Spectral-spatial classification of hyperspectral images using deep convolutional neural networks. *Remote Sensing Letters*, 6(6) :468–477, 2015.