



HAL
open science

Algorithme d'apprentissage pour la segmentation d'images hyperspectrales compressées

Maud Biquard, Antoine Rouxel, Simon Lacroix, Hervé Carfantan, Antoine
Monmayrant, Henri Camon

► **To cite this version:**

Maud Biquard, Antoine Rouxel, Simon Lacroix, Hervé Carfantan, Antoine Monmayrant, et al.. Algorithme d'apprentissage pour la segmentation d'images hyperspectrales compressées. ORASIS 2021, Centre National de la Recherche Scientifique [CNRS], Sep 2021, Saint Ferréol, France. hal-03339620

HAL Id: hal-03339620

<https://hal.science/hal-03339620v1>

Submitted on 9 Sep 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.

CSSNet : un algorithme d'apprentissage pour la segmentation d'images hyperspectrales compressées

M. Biquard¹ A. Rouxel² S. Lacroix¹ H. Carfantan³ A. Monmayrant¹ H. Camon¹

¹ LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France.

² Airbus Defence and Space, Toulouse, France.

³ IRAP, Université de Toulouse, Université Toulouse 3 - Paul Sabatier, CNRS, CNES, Toulouse, France.

antoine.rouxel@laas.fr

1 Introduction

Dans l'observation et la surveillance environnementale depuis l'espace, l'imagerie hyperspectrale joue un rôle central par la richesse de l'information qu'elle fournit. Elle permet notamment la segmentation de la scène à partir du cube hyperspectral, étape essentielle pour de nombreuses applications. L'imagerie hyperspectrale pose cependant un double défi d'une part par la quantité de données à mesurer et à retransmettre au sol, et d'autre part par l'importance des ressources nécessaires à leur analyse en post-traitement.

Dans ce contexte, l'acquisition compressée d'image hyperspectrale, où chaque pixel de l'image mesurée combine plusieurs composantes spectrales, permet de réduire la quantité de données à mesurer et à transmettre. L'approche classique est de reconstruire le cube hyperspectral par des étapes de post-traitement exploitant un certain nombre d'hypothèses sur la scène observée : régularité et redondance, connaissance a priori des spectres présents, etc...

Nous proposons ici un algorithme basé sur un réseau de neurones peu profond qui permet la segmentation sémantique de la scène à partir d'images hyperspectrales compressées sans passer par une étape de reconstruction du cube hyperspectral.

Travaux similaires. Un certain nombre de travaux visent à extraire des informations sémantiques à partir du cube hyperspectral complet via des réseaux de neurones. Étant donné le manque de cubes labellisés, cette extraction est faite à partir de petits patchs de la scène. Des approches viennent les *classifier* à l'aide de convolutions 1D et 2D [8] [10] [5], ou de convolutions 3D [9] [6]. D'autres font de la *segmentation sémantique* sur ces patchs [11].

Des travaux récents cherchent, comme nous le faisons, à segmenter la scène directement à partir des images compressées. Des approches inspirées du *compressed sensing* segmentent la scène en s'appuyant sur des hypothèses de parcimonie [12]. D'autres s'appuient sur les probabilités

bayésiennes pour proposer une segmentation itérative [3]. Hao Zhang et al. utilisent des réseaux de neurones pour co-optimiser la segmentation de la scène et la génération de masques [13].

2 Système double-disperseur CASSI

Notre proposition repose sur l'utilisation d'un imageur hyperspectral DD-CASSI similaire à celui décrit dans [7]. Celui-ci est constitué de deux spectro-imageurs symétriques, de part et d'autre d'une matrice de micro-miroirs (DMD). Le premier spectro-imageur sépare spatialement et image les différents plans spectraux de la scène sur le DMD. Chaque plan spectral est donc filtré par un morceau différent du DMD. Le deuxième spectro-imageur recombine les plans spectraux filtrés et les ré-image sur la caméra, en annulant toute dispersion spectrale. La propriété essentielle de ce système et nécessaire à notre approche algorithmique est la *colocalisation* : quel que soit le filtrage spatial introduit par le DMD, les composantes spectrales présentes en un point (x, y) de la caméra proviennent uniquement du spectre présent dans la scène au point correspondant. Une conséquence de la colocalisation est un accès direct sur la caméra à l'image panchromatique en ouvrant totalement le DMD (filtre passant en tout point).

Dans le cas plus général d'un motif de filtrage quelconque sur le DMD, nous mesurons en chaque point de la caméra une somme pondérée des différentes composantes des spectres en ce point (voir équation 1). Les poids sont directement liés au masque programmé sur le DMD et à la dispersion des plans spectraux sur celui-ci. Il est ainsi possible à partir du masque et de la dispersion du système de construire un *cube de filtrage*.

Le modèle discretisé de l'imageur DD-CASSI s'exprime ainsi :

$$I_{r,c} = \sum_{k=0}^{n_\lambda-1} h_{r,c,k} S_{r,c,k} \quad (1)$$

avec $I_{r,c}$ l'intensité mesurée par le pixel (r, c) du détecteur,

$r \in \llbracket 0, n_r - 1 \rrbracket$ et $c \in \llbracket 0, n_c - 1 \rrbracket$ avec n_r et n_c les nombres de lignes et de colonnes sur le détecteur. n_λ représente le nombre de composantes spectrales. $h_{r,c,k}$ correspond à la valeur du cube de filtrage associée à la k -ème composante spectrale du pixel (r, c) et $S_{r,c,k}$ l'intensité spectro-spatiale de la scène observée, associée au voxel (r, c, k) .

3 Architecture de CSSNet

3.1 Présentation de l'entrée

Notations. Le réseau fonctionne avec des patches de taille p . Pour former l'entrée, on a besoin de 3 objets : le cube de filtrage aux dimensions spatiales du patch $H \in [0, 1]^{p \times p \times n_\lambda}$, l'image compressée associée $I \in \mathbb{R}^{p \times p}$ et l'image panchromatique de la scène $P \in \mathbb{R}^{p \times p}$. La sortie est $O \in \mathbb{R}^{p \times p \times q}$ avec q le nombre de classes. On estime $S \in \{0, 1, \dots, q - 1\}^{p \times p}$ avec $\hat{S} = (\operatorname{argmax}_{0 \leq i, j \leq p-1} O_{i,j})_{0 \leq i, j \leq p-1}$.

Normalisation. On veut extraire l'information spectrale de I en retirant P . Or I ne comporte qu'un pourcentage du flux de la lumière de P et n'a pas les mêmes variations. On définit alors $I_{spec} = \bar{I} - \bar{P}$ avec \bar{I} et \bar{P} les versions normalisées de I et P .

Entrée. L'entrée du réseau est constituée de $(\bar{H}, \bar{I}_{spec}, \bar{P})$, \bar{H} et \bar{I}_{spec} étant les versions normalisées de H et I_{spec} .

3.2 Architecture du réseau

Le réseau CSSNet ("Compressed Semantic Segmentation Network") se décompose en deux parties. La première traite l'information spectrale apportée par l'image compressée et le cube de filtrage. La deuxième partie mutualise ces données avec l'information spatiale extraite de l'image panchromatique.

Le réseau est constitué uniquement de couches de convolutions sur des "feature maps" de taille $p \times p$: un "padding" suffisant est appliqué afin de conserver la taille de ces "features maps".

Un schéma détaillé de l'architecture est donné Figure 1.

Traitement spectral. Cette partie comprend 2 blocs de convolutions \mathcal{F}_1 et \mathcal{F}_2 . Le traitement effectué est :

$$H_2 = \mathcal{F}_1(\bar{H}), \quad H_3 = \mathcal{F}_2([H_2, \bar{I}_{spec}]) \quad (2)$$

où l'opération $[\cdot, \cdot]$ représente la concaténation le long des channels. \bar{H} est de dimension $p \times p \times n_\lambda$: on considère pour l'entrée de \mathcal{F}_1 que \bar{H} est de dimension $p \times p$ avec n_λ channels.

Traitement spatial/spectral. L'architecture utilisée est inspirée du réseau *DSSNet* [11] utilisée pour la segmentation du cube hyperspectral complet, à laquelle deux couches de convolutions ont été ajoutées, et le nombre de channels modifié. Cette partie se décompose en 3 blocs de convolutions \mathcal{F}_3 , $\mathcal{F}_{3, sym}$, et \mathcal{F}_4 .

$$H_4 = \mathcal{F}_3(H_3), \quad H_5 = \mathcal{F}_{3, sym}([H_4, \bar{P}]), \quad O = \mathcal{F}_5(H_5) \quad (3)$$

Optimisation. La loss utilisée est la weighted cross entropy loss afin d'évaluer la différence entre \hat{S} et S :

$$\mathcal{L}_{CE} = -\frac{1}{N} \sum_{s=1}^N \sum_{k=0}^{q-1} w_k \sum_{0 \leq i, j \leq p} t_{i,j,k}^s \log(p_{i,j,k}^s) \quad (4)$$

avec N le nombre de patches, $p_{i,j,k}^s$ la probabilité softmax que le pixel (i, j) du patch s appartienne à la classe k , $t_{i,j,k}^s$ le label de la classe k pour le pixel (i, j) et w_k l'inverse de la fréquence médiane, calculée sur tout le dataset d'entraînement, de la classe k .

4 Résultats

4.1 Méthodologie

Datasets utilisés. Le réseau est testé sur deux datasets : Indian Pines [2] et Pavia University [4]. Pour construire la base de données, ces deux cubes hyperspectraux passent dans un simulateur du système DD-CASSI. Les caractéristiques des deux datasets sont présentées Figure 2.

Noms des datasets	Dimensions	Nombre de bandes	Nombre de classes
PaviaU	610 x 340	103	9
Indian Pines	145 x 145	220	16

FIGURE 2 – Caractéristiques des datasets utilisés

Séparation entraînement inférence. L'entraînement et l'inférence se déroulent sur des images compressées provenant du même cube hyperspectral. Les expérimentations sont faites en utilisant une séparation bien définie dans laquelle aucun pixel du cube hyperspectral intervenant dans l'inférence n'est vu pendant la phase d'entraînement. Une zone de 10% est réservée à l'inférence, 80% à l'entraînement (le reste des patches se situant à la frontière).

Réglages expérimentaux. Les patches sont de taille 7×7 . Pour chaque dataset, 40 images compressées ainsi qu'une image panchromatique est acquise. 80% des images compressées sont utilisées pour l'entraînement et 20% pour l'inférence. Pour l'inférence, l'image est parcourue à l'aide d'une fenêtre glissante de stride 3. Le réseau est entraîné sur 60 epochs avec "early stopping". Le "learning rate" est de 0.01, le "batch size" de 100, le "weight decay" de $5e-4$. L'optimiseur utilisé est SGD (Stochastic Gradient Descent) avec un momentum de 0.9.

Métriques. Deux métriques sont utilisées : l'accuracy (OA) et le f-score moyen (moyenne des f-scores par classe).

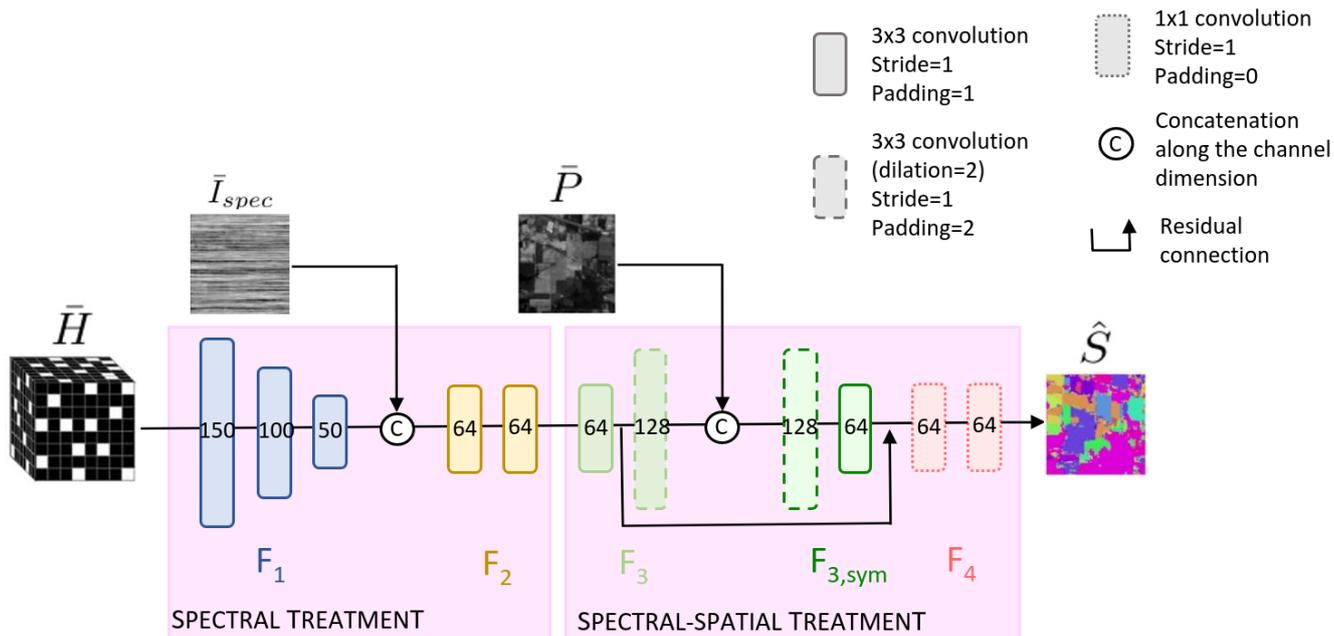


FIGURE 1 – Architecture du réseau CSSNet. Chaque rectangle bleu, jaune vert ou rouge représente une couche de convolutions. Le nombre à l’intérieur de chaque couche de convolution représente le nombre de channels. Les entrées du réseau sont : le patch "image panchromatique" normalisé \bar{P} , le patch "image compressée" normalisé \bar{I}_{spec} et le patch "cube de filtrage associé" normalisé \bar{H} . La sortie du réseau est \hat{S} qui associe à chaque pixel du patch un vecteur de probabilités.

Comparaison avec l’existant. Les détails d’implémentation menant aux résultats présentés dans [13] n’étant pas explicites, nous nous comparons uniquement à des réseaux de segmentation sur cubes hyperspectraux complets : le 3D CNN de [8] implémenté par [1] et DSSNet (notre propre implémentation).

4.2 Résultats de segmentation

	Pavia U		Indian Pines	
	Accuracy	F-score moyen	Accuracy	F-score moyen
3D-CNN	93.7	93.0	67.9	69.4
DSSNet	98.4	98.0	67.2	63.0
CSSNet	84.0	78.6	58.4	39.1
CSSNet (mix)	80.2	75.6	53.7	41.0

FIGURE 3 – Comparaison des résultats des algorithmes de segmentation sur PaviaU et Indian Pines. Les deux lignes grisées correspondent aux algorithmes de référence auxquels nous nous comparons. Contrairement à nos réseaux, ces algorithmes prennent en entrée l’ensemble du cube hyperspectral. Les deux lignes suivantes du tableau contiennent les résultats de deux versions différentes de notre réseau de segmentation. CSSNet (mix) correspond à une version de CSSNet où l’image panchromatique \bar{P} est introduite en même temps que l’image compressée \bar{I}_{spec} .

Chaque réseau a été entraîné 5 fois, c’est la moyenne des scores qui est affichée Figure 3. La contre-performance du 3D-CNN sur Pavia University, représentant l’état de l’art,

est sans doute due à un problème - dont nous n’avons pas trouvé la source - de réglage d’hyperparamètres.

Pour chaque entraînement de CSSNet, les métriques sont calculées pour plusieurs masques DMD et la moyenne est prise. Il est à noter qu’avec les masques qui fonctionnent le mieux, l’accuracy sur Pavia University est en moyenne de 86.5.

Par rapport à la segmentation sur le cube complet, il y a une perte de performance qui est naturelle car nous traitons 50 et 100 moins de volume de données sur Pavia University et Indian Pines. Mais les résultats restent néanmoins acceptables, le détail des erreurs est présenté Figure 4.

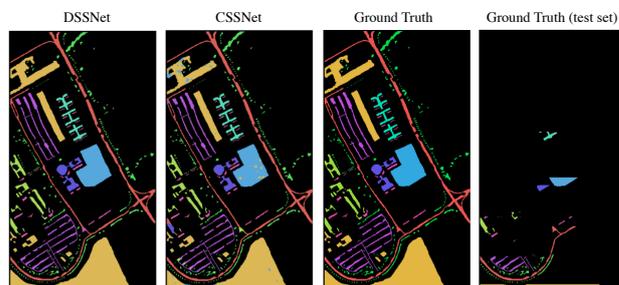


FIGURE 4 – Résultats de la segmentation pour DSSNet et CSSNet. Chaque couleur représente une classe, les zones non labellisées sont en noir. À droite : les zones de test exploitées pour les évaluations numériques présentées Table 3

Sur la séparation spatial-spectral. \bar{P} n'est introduite qu'à la fin du réseau. Une partie conséquente de celui-ci est dédié uniquement à l'extraction d'informations spectrales. On observe, en introduisant \bar{P} en même temps que \bar{I}_{spec} (réseau CSSNet (mix) dans la Figure 3), une moins bonne généralisation du réseau à l'inférence. L'hypothèse faite est qu'extraire des informations spectrales à partir de \bar{H} et \bar{I}_{spec} est une tâche plus complexe qu'extraire les contours à partir de \bar{P} . En incorporant \bar{P} plus tôt, le réseau lui accorde trop d'importance et "sur-apprend" la partie spatiale.

5 Conclusion

Dans cet article, nous avons présenté un algorithme de segmentation sémantique à partir de données hyperspectrales compressées. Il est basé sur une architecture de réseaux de neurones prenant en compte les spécificités du système de mesures. Les résultats de simulation indiquent des performances inférieures à celles de l'état de l'art sur cube hyperspectral complet mais le nombre d'acquisitions nécessaire à l'obtention des mesures est drastiquement réduit. Les prochaines étapes sont d'effectuer des tests sur des ensembles de données acquises avec notre prototype [7].

Références

- [1] Nicolas Audebert, Bertrand Le Saux, and Sebastien Lefevre. Deep learning for classification of hyperspectral data : A comparative review. *IEEE Geoscience and Remote Sensing Magazine*, 7(2) :159–173, jun 2019.
- [2] Marion F. Baumgardner, Larry L. Biehl, and David A. Landgrebe. 220 Band AVIRIS Hyperspectral Image Data Set : June 12, 1992 Indian Pine Test Site 3, 2015.
- [3] Matthew Dunlop-Gray, Phillip K. Poon, Dathon Gollish, Esteban Vera, and Michael E. Gehm. Experimental demonstration of an adaptive architecture for direct spectral imaging classification. *Optics Express*, 24(16) :18307, aug 2016.
- [4] Paolo Gamba. Pavia University hyperspectral dataset from Grupo de Inteligencia Compututacional.
- [5] Amina Ben Hamida, Alexandre Benoit, Patrick Lambert, and Chokri Ben Amar. 3-d deep learning approach for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 56(8) :4420–4434, aug 2018.
- [6] Mingyi He, Bo Li, and Huahui Chen. Multi-scale 3d deep convolutional neural network for hyperspectral image classification. *IEEE*, 2017.
- [7] Elizabeth Hemsley, Simon Lacroix, Hervé Carfantan, and Antoine Monmayrant. Calibration of programmable spectral imager with dual disperser architecture. *Optics Communications*, 468 :125767, aug 2020.
- [8] Hyungtae Lee and Heesung Kwon. Going deeper with contextual CNN for hyperspectral image classification. *IEEE Transactions on Image Processing*, 26(10) :4843–4855, oct 2017.
- [9] Ying Li, Haokui Zhang, and Qiang Shen. Spectral-spatial classification of hyperspectral imagery with 3d convolutional neural network. *Remote Sensing*, 9(1) :67, jan 2017.
- [10] Yanan Luo, Jie Zou, Chengfei Yao, Xiaosong Zhao, Tao Li, and Gang Bai. HSI-CNN : A novel convolution neural network for hyperspectral image. In *2018 International Conference on Audio, Language and Image Processing (ICALIP)*. IEEE, jul 2018.
- [11] Bin Pan, Xia Xu, Zhenwei Shi, Ning Zhang, Huanlin Luo, and Xianchao Lan. DSSNet : A simple dilated semantic segmentation network for hyperspectral imagery classification. *IEEE Geoscience and Remote Sensing Letters*, 17(11) :1968–1972, nov 2020.
- [12] Ana Ramirez, Henry Arguello, Gonzalo R. Arce, and Brian M. Sadler. Spectral image classification from optimal coded-aperture compressive measurements. *IEEE Transactions on Geoscience and Remote Sensing*, 52(6) :3299–3309, jun 2014.
- [13] Hao Zhang. Joint coded aperture optimization and compressive hyperspectral image classification using 3D coded neural network. *arXiv e-prints arXiv :2009.11948*, 2020.