CSSNET: A LEARNING ALGORITHM FOR THE SEGMENTATION OF COMPRESSED HYPERSPECTRAL IMAGES

M. Biquard, A. Rouxel, S. Lacroix, H. Carfantan†, A. Monmayrant, H. Camon*

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

ABSTRACT

The paper presents a semantic segmentation method which is directly applicable to compressed hyperspectral images acquired with a dual-disperser CASSI instrument. It introduses an algorithm based on a shallow neural network that exploits the spectral filtering performed by the optical system and the compressed hyperspectral images measured by the detector. Encouraging results that exploit 50 to 100 less data than the whole hyperspectral datacube on PaviaU and IndianPines datasets are presented.

***Index Terms*—** Compressed hyperspectral imaging, Dual-Disperser CASSI, Compressed images segmentation

1. INTRODUCTION

Classic hyperspectral imagers that produce a whole data cube pose a twofold challenge in environmental observation and monitoring from space, firstly because of the quantity of data to be transmitted to the ground, and secondly because of the resources required for post-processing analyses. Compressed hyperspectral acquisition systems, where each pixel of the acquired images combines several spectral components, can alleviate these issues, firstly by reducing the amount of data to be acquired and transmitted.

The interpretation of compressed measurements requires the use of algorithms that exploit various assumptions about the observed scene, such as regularity and redundancy, and an a priori knowledge of the spectral filtering performed by the optical system. The classical approach is to reconstruct the complete hyperspectral cube from the compressed data. State-of-the-art methods such as TwIST [1], GAP-TV [2], DeSCI [3] and λ-net [4] obtain impressive results that rely on theoretical work on "Coded-Aperture Snapshot Spectral Imager" systems (CASSI, [5, 6]). Nevertheless, in many cases and particularly in remote sensing, the objective of image acquisition is not to obtain the signal in its raw form, but rather to extract given semantic information from the scene. In such cases, the reconstruction is only an intermediate step in data processing, it can be costly by itself, it still requires post-processing of the reconstructed cube, whereas it can by-passed by directly processing the compressed data.

In this paper, we present a semantic segmentation method which is directly applicable to compressed hyperspectral images acquired with a dual-disperser CASSI instrument. By refraining from reconstructing the raw signal, this work is in line with work done by Davenport [7, 8], Arguello [9, 10, 11] and Arce [12]. We propose an algorithm based on a shallow neural network that exploits the spectral filtering performed by the optical system and the compressed hyperspectral images measured by the detector.

The next section briefly describes the Dual-Disperser CASSI imager. Section 3 introduces the CSSNet architecture, and section 4 presents some results.

2. DUAL-DISPERSER CASSI

Our proposal is based on the use of a DD-CASSI hyperspectral imager, originally introduced in [13], on the basis of which we have developped a prototype composed of two symmetrical spectro-imagers on both sides of a micromirror array (DMD) [14]. The first spectro-imager spatially separates and images the different spectral planes of the scene on the DMD. Each spectral plane is therefore filtered by a different piece of the DMD. The second spectro-imager recombines the filtered spectral planes and re-images them on the camera, cancelling any spectral dispersion (figure 1). The essential property of this system and necessary to our approach is the *co-localization*: whatever the spatial filtering introduced by the DMD, the spectral components present at a point (x, y) of the camera come only from the spectrum present in the scene at the corresponding point. A consequence of co-localization is a direct access to the panchromatic image on the camera by completely opening the DMD.

In the more general case of any filtering pattern on the DMD, we measure at each point of the camera a weighted sum of the different components of the spectrum at this point (equation 1). The weights are directly related to the mask programmed on the DMD and the dispersion of spectral planes on it. It is thus possible from the mask and the dispersion of the system to build a *filtering cube*.

The discretized model of the DD-CASSI imager is expressed as follows:

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

with $I_{r,c}$ the intensity measured by the pixel (r, c) of the

*Airbus Defence and Space, Toulouse, France

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

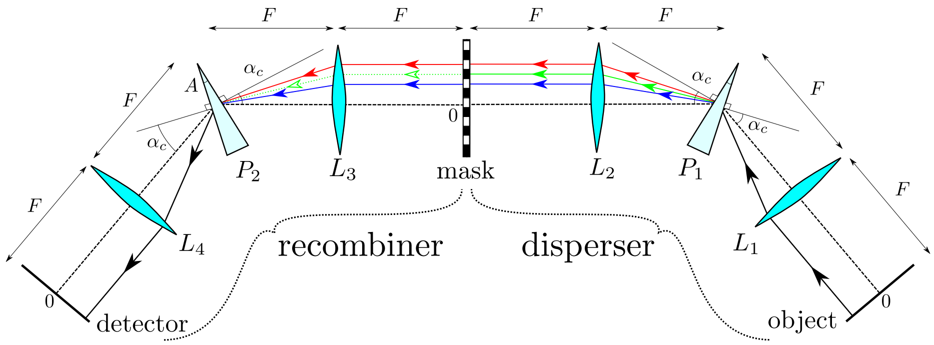

Fig. 1. Dual-disperser CASSI architecture.

detector, $r \in [0, n_r - 1]$ and $c \in [0, n_c - 1]$ with n_r and n_c the numbers of rows and columns on the detector. $n_{r,c,k}$ is the number of spectral components. $h_{r,c,k}$ is the value of the filter cube associated with the k-th spectral component of the pixel (r, c) and $S_{r,c,k}$ is the spectro-spatial intensity of the observed scene, associated with the voxel (r, c, k).

3. CSSNET ARCHITECTURE

3.1. Related work

A number of works aim at extracting semantic information from the full hyperspectral cube via neural networks. Given the lack of labeled cubes, this extraction is done from small patches of the scene. Approaches come to *classify* them using 1D and 2D convolutions [15] [16] [17], or 3D convolutions [18] [19]. Others do semantic segmentation on these patches [20].
Recent works seek, as we do, to segment the scene directly from the compressed images. Approaches inspired by *compressed sensing* segment the scene based on sparcity assumptions [21]. Others rely on Bayesian probabilities to propose iterative segmentation [22]. Hao Zhang et al. use neural networks to co-optimize scene segmentation and mask generation [23].

3.2. Network inputs

The network works with patches of size p. To form the input, we need 3 objects: the filter cube with spatial dimensions of the patch $H \in [0, 1]^{p\times p\times n_\lambda}$, the associated compressed image $I \in \mathbf{R}^{p\times p}$ and the panchromatic image of the scene $P \in \mathbf{R}^{p\times p}$. The output is $O \in \mathbf{R}^{p\times p\times q}$ with q the number of classes. We estimate $S \in \{0, 1, ..., q-1\}^{p\times p}$ with $\hat{S} = (\arg\max O_{i,j})_{0\leq i,j\leq p-1}$.
We want to extract the spectral information of I by removing P. But I is only a percentage of the light flux of P and does not have the same variations. We then define $I_{spec} = \bar{I} - \bar{P}$ with $\bar{I}$ and $\bar{P}$ the normalized versions of I and P.
The input of the network consists of $(\bar{H}, \bar{I}_{spec}, \bar{P})$, $\bar{H}$ and $\bar{I}_{spec}$ being the normalized versions of H and I_{spec}.

3.3. Network architecture

The CSSNet ("Compressed Semantic Segmentation Network") is divided into two parts. The first one processes the spectral information brought by the compressed image and the filtering cube. The second part mutualizes these data with the spatial information extracted from the panchromatic image.
The network is only made of convolution layers on feature maps of size $p \times p$: a sufficient padding is applied to keep the size of these feature maps.
A detailed diagram of the architecture is given in Figure 2.

Spectral processing. This part includes two blocks of convolutions $\mathcal{F}_1$ and $\mathcal{F}_2$. The treatment performed is :

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

where the operation $[., .]$ represents the concatenation along the channels. $\bar{H}$ is of dimension $p \times p \times n_\lambda$: we consider for the input of $\mathcal{F}_1$ that $\bar{H}$ is of dimension $p \times p$ with n_λ channels.

Spatial/spectral processing. The architecture used is inspired by the *DSSNet* [20] used for the segmentation of the complete hyperspectral cube, to which two convolution layers have been added, and the number of channels modified. This part is decomposed into three convolution blocks $\mathcal{F}_3$, $\mathcal{F}_{3,sym}$, and $\mathcal{F}_4$.

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

Optimisation. The loss used is the weighted cross entropy loss to evaluate the difference between
$\hat{S}$ and S:

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

with N the number of patches, $p^s_{i,j,k}$ the softmax probability that pixel (i, j) of patch s belongs to class k, $t^s_{i,j,k}$ the class label k for pixel (i, j), and w_k the inverse of the median frequency, computed over the whole training dataset, of class k.

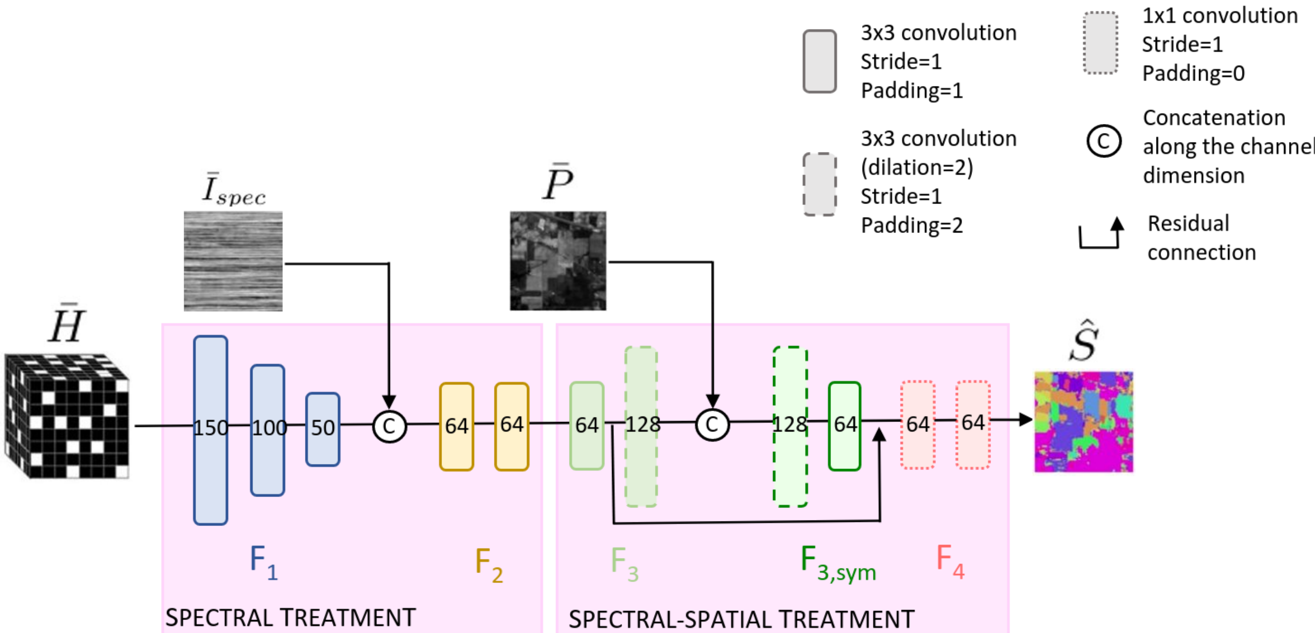

Fig. 2. Architecture of the CSSNet network. Each blue, yellow, green or red rectangle represents a convolution layer. The number inside each convolution layer represents the number of channels. The inputs of the network are: the patch "panchromatic image" normalized $\bar{P}$, the patch "compressed image" normalized $\bar{I}_{spec}$ and the patch "associated filtering cube" normalized $\bar{H}$. The output of the network is $\hat{S}$ which associates to each pixel of the patch a vector of probabilities.

Dataset	Dimensions	Number of bands	Number of classes
PaviaU	610 x 340	103	9
Indian Pines	145 x 145	220	16

Fig. 3. Used datasets

4. RESULTS

4.1. Methodology

Datasets used. The network is tested on two datasets: Indian Pines [24] and Pavia University [25]. To build the database, these two hyperspectral cubes are run through a DD-CASSI system simulator. The characteristics of the two datasets are presented in Figure 3.

Training and inference separation. Training and inference are performed on compressed images from the same hyperspectral cube. The experiments are done using a well defined separation in which no pixel of the hyperspectral cube involved in the inference is seen during the training phase. An area of 10% is reserved for inference, 80% for training (the rest of the patches are located at the border).

Experimental settings. The patches are of size 7×7. For each dataset, 40 compressed images and one panchromatic image are acquired. 80% of the compressed images are used for training and 20% for inference. For inference, the image is scanned with a sliding window of stride 3. The network is trained on 60 epochs with "early stopping". The learning rate is 0.01, the batch size is 100, the weight decay is 5e-4. The optimizer used is SGD (Stochastic Gradient Descent) with a momentum of 0.9.

Metrics. Two metrics are used: the accuracy (OA) and the average f-score (average of the f-scores per class).

Comparison to existing work. As the implementation details leading to the results presented in [23] are not explicit, we only compare ourselves to segmentation networks on full hyperspectral cubes: the 3D CNN of [15] implemented by [26] and DSSNet (our own implementation).

4.2. Segmentation results

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

Fig. 4. Comparison of segmentation algorithm results on PaviaU and Indian Pines. The two shaded lines correspond to the reference algorithms to which we compare ourselves. Unlike our networks, these algorithms take the entire hyperspectral cube as input. The next two rows of the table contain the results of two different versions of our segmentation network. CSSNet (mix) corresponds to a version of CSSNet where the panchromatic image $\bar{P}$ is introduced together with the compressed image $\bar{I}_{spec}$.

Each network has been trained 5 times, and the average of the scores is displayed in Figure 4. The underperformance of 3D-CNN on Pavia University, representing the state of the art, is probably due to a problem - whose source we did not find - of hyperparameter setting.
For each CSSNet training, the metrics are computed for several DMD masks and the average is taken. Note that with the best performing masks, the accuracy on Pavia University averages 86.5.
Compared to the full cube segmentation, there is a loss of performance which is natural because we process 50 and 100 less data volume on Pavia University and Indian Pines. But the results are still acceptable, the details of the errors are shown in Figure 5.

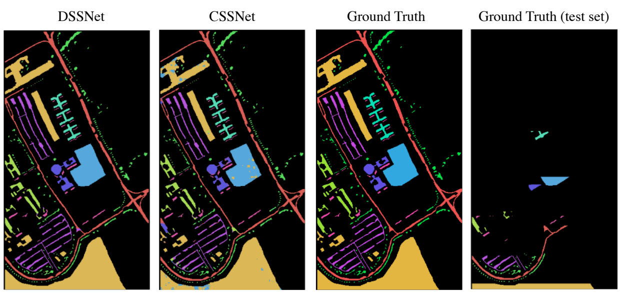

Fig. 5. Segmentation results for DSSNet and CSSNet. Each color represents a class, unlabeled areas are in black. Right: the test areas exploited for the numerical evaluations presented Table 4

On spatial-spectral separation. $\bar{P}$ is introduced only at the end of the network. A consequent part of this one is dedicated only to the extraction of spectral information. We observe, by introducing $\bar{P}$ at the same time as $\bar{I}_{spec}$ (CSSNet (mix) in Figure 1), a less good generalization of the network to inference. The assumption made is that extracting spectral information from $\bar{H}$ and $\bar{I}_{spec}$ is a more complex task than extracting contours from $\bar{P}$. By incorporating $\bar{P}$ earlier, the network gives it too much importance and "over-learns" the spatial part.

5. CONCLUSION

In this paper, we have presented an algorithm for semantic segmentation from compressed hyperspectral data. It is based on a neural network architecture taking into account the specificities of the measurement system. Simulation results indicate lower performances than the state of the art on full hyperspectral cube but the number of acquisitions required to obtain the measurements is drastically reduced by nearly two orders of magnitude. The next steps are to perform iterative segmentation, either with Bayesian fusion or using a RNN type architecture, and perform tests on data sets acquired with our prototype [14].

6. REFERENCES

[1] J.M. Bioucas-Dias and M.A.T. Figueiredo, "A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration," *IEEE Transactions on Image Processing*, vol. 16, no. 12, pp. 2992–3004, dec 2007.

[2] Xin Yuan, "Generalized alternating projection based total variation minimization for compressive sensing," in *2016 IEEE International Conference on Image Processing (ICIP)*. sep 2016, IEEE.

[3] Yang Liu, Xin Yuan, Jinli Suo, David J. Brady, and Qionghai Dai, "Rank minimization for snapshot compressive imaging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 12, pp. 2990–3006, dec 2019.

[4] Xin Miao, Xin Yuan, Yunchen Pu, and Vassilis Athitsos, "lambda-net: Reconstruct hyperspectral images from a snapshot measurement," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 4058–4068.

[5] Shirin Jalali and Xin Yuan, "Snapshot compressed sensing: Performance bounds and algorithms," *IEEE Transactions on Information Theory*, vol. 65, no. 12, pp. 8005–8024, dec 2019.

[6] Claire Boyer, Jérémie Bigot, and Pierre Weiss, "Compressed sensing with structured sparsity and structured acquisition," *Applied and Computational Harmonic Analysis*, vol. 46, no. 2, pp. 312–350, mar 2019.

[7] Mark A. Davenport, Marco F. Duarte, Michael B. Wakin, Jason N. Laska, Dharmpal Takhar, Kevin F. Kelly, and Richard G. Baraniuk, "The smashed filter for compressive classification and target recognition," in *SPIE Proceedings*, Charles A. Bouman, Eric L. Miller, and Ilya Pollak, Eds. feb 2007, SPIE.

[8] M.A. Davenport, P.T. Boufounos, M.B. Wakin, and R.G. Baraniuk, "Signal processing with compressive measurements," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 445–460, apr 2010.

[9] David Alberto Boada, Hector Miguel Vargas, Jaime Octavio Albarracin Ferreira, and Henry Arguello Fuentes, "Algorithm for spectral image target detection from compressive measurements using a joint sparsity model and a CASSI architecture," in *2015 20th Symposium on Signal Processing, Images and Computer Vision (STSIVA)*. sep 2015, IEEE.

[10] Jorge Bacca, Laura Galvis, and Henry Arguello, "Coupled deep learning coded aperture design for compressive image classification," *Optics Express*, vol. 28, no. 6, pp. 8528, mar 2020.

[11] Nelson Diaz, Juan Ramirez, Esteban Vera, and Henry Arguello, "Adaptive multisensor acquisition via spatial contextual information for compressive spectral

image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 9254–9266, 2021.

[12] Hao Zhang, Xu Ma, Xianhong Zhao, and Gonzalo R. Arce, "Compressive hyperspectral image classification using a 3d coded convolutional neural network," *Optics Express*, vol. 29, no. 21, pp. 32875, sep 2021.

[13] M. E. Gehm, R. John, D. J. Brady, R. M. Willett, and T. J. Schulz, "Single-shot compressive spectral imaging with a dual-disperser architecture," *Optics Express*, vol. 15, no. 21, pp. 14013, oct 2007.

[14] Elizabeth Hemsley, Simon Lacroix, Hervé Carfantan, and Antoine Monmayrant, "Calibration of programmable spectral imager with dual disperser architecture," *Optics Communications*, vol. 468, pp. 125767, aug 2020.

[15] Hyungtae Lee and Heesung Kwon, "Going deeper with contextual CNN for hyperspectral image classification," *IEEE Transactions on Image Processing*, vol. 26, no. 10, pp. 4843–4855, oct 2017.

[16] 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)*. jul 2018, IEEE.

[17] 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*, vol. 56, no. 8, pp. 4420–4434, aug 2018.

[18] 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.

[19] Mingyi He, Bo Li, and Huahui Chen, "Multi-scale 3d deep convolutional neural network for hyperspectral image classification," *IEEE*, 2017.

[20] 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*, vol. 17, no. 11, pp. 1968–1972, nov 2020.

[21] 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*, vol. 52, no. 6, pp. 3299–3309, jun 2014.

[22] Matthew Dunlop-Gray, Phillip K. Poon, Dathon Golish, Esteban Vera, and Michael E. Gehm, "Experimental demonstration of an adaptive architecture for direct spectral imaging classification," *Optics Express*, vol. 24, no. 16, pp. 18307, aug 2016.

[23] Hao Zhang, "Joint coded aperture optimization and compressive hyperspectral image classification using 3D coded neural network," *arXiv e-prints arXiv:2009.11948*, 2020.

[24] 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.

[25] Paolo Gamba, "Pavia University hyperspectral dataset from Grupo de Inteligencia Compututacional," .

[26] Nicolas Audebert, Bertrand Le Saux, and Sebastien Lefevre, "Deep learning for classification of hyperspectral data: A comparative review," *IEEE Geoscience and Remote Sensing Magazine*, vol. 7, no. 2, pp. 159–173, jun 2019.

[27] Fahim Irfan Alam, Jun Zhou, Alan Wee-Chung Liew, and Xiuping Jia, "CRF learning with CNN features for hyperspectral image segmentation," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. jul 2016, IEEE.