



**HAL**  
open science

## A CNN cloud detector for panchromatic satellite images

Mariano Rodríguez, Jérémy Anger, Carlo de Franchis, Charles Hessel,  
Gabriele Facciolo, Rafael Grompone von Gioi, Jean-Michel Morel

### ► To cite this version:

Mariano Rodríguez, Jérémy Anger, Carlo de Franchis, Charles Hessel, Gabriele Facciolo, et al.. A CNN cloud detector for panchromatic satellite images. IGARSS 2021 - 2021 IEEE International Geoscience and Remote Sensing Symposium, Jul 2021, Bruxelles, Belgium. 10.1109/IGARSS47720.2021.9554059 . hal-03126940

**HAL Id: hal-03126940**

**<https://hal.science/hal-03126940>**

Submitted on 1 Feb 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.

# A CNN CLOUD DETECTOR FOR PANCHROMATIC SATELLITE IMAGES

M. Rodríguez,<sup>†</sup> J. Anger,<sup>§,†</sup> C. de Franchis,<sup>§,†</sup> C. Hessel,<sup>§,†</sup> G. Facciolo,<sup>†</sup> R. Grompone<sup>†</sup> and J.-M. Morel<sup>†</sup>

<sup>†</sup> Université Paris-Saclay, CNRS, ENS Paris-Saclay, Centre Borelli, France  
<sup>§</sup> Kayrros SAS

## ABSTRACT

Cloud detection is a crucial step for automatic satellite image analysis. Some cloud detection methods exploit specially designed spectral bands, other base the detection on time series, or on the inter-band delay in push-broom satellites. Nevertheless many use cases occur where these methods do not apply. This paper describes a convolutional neural network for cloud detection in panchromatic and single-frame images. Only a per-image annotation is required, indicating which images contain clouds and which are cloud-free. Our experiments show that, in spite of using less information, the proposed method produces competitive results.

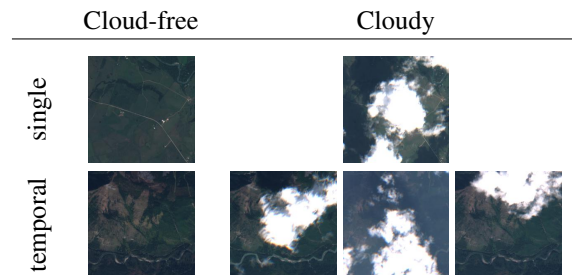
**Index Terms**— Cloud detector, CNN, single-band, panchromatic, satellite images.

## 1. INTRODUCTION

Unless observed for meteorological purposes, clouds are an overwhelming nuisance for optical satellite imagery. Not only they hide the ground, but their detection is also a major concern to avoid detection and interpretation errors in automatic image analysis. Being so numerous and large, satellite images require an automatic and accurate analysis [1, 2].

A number of companies in the Earth observation industry are launching constellations of dozens to hundreds of satellites to get a short revisit time over any region. Most of these satellites only acquire a few spectral bands in the visible light range of the electromagnetic spectrum. Thus, many lack specifically designed spectral bands for cloud detection (such as, for example, the cirrus band B10 of Sentinel-2).

Most well known approaches to detect clouds on satellite images are based on exploiting spatio-temporal information [1, 2]. Satellite cloud detection often exploits spectral bands specifically designed for cloud detection [3, 4, 5, 6, 1, 7, 8]. Alternatively, the inter-band delay in push-broom satellites allows cloud detection by parallax analysis of the color bands [9, 10, 11, 12, 13, 14]. In some cases, local descriptors are used to detect changes, hopefully due to clouds,



**Fig. 1:** Examples of labels and images provided by the single and temporal image datasets [18].

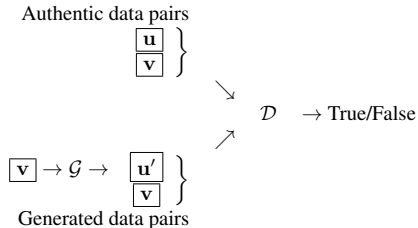
and, in other cases, the clouds' relative altitude is measured with respect to the ground through the interband parallax information [15, 16, 17]. Certainly, the case of cloud detection in panchromatic optical images has none of aforementioned characteristics and presents a challenge as it contains less information to work with.

Cloud-related tasks are currently being learned from data, with the promise of a better performance. Ideally, this learning process uses a well annotated per-pixel dataset (e.g. each pixel has a label : cloud or not-cloud). However, per-pixel annotations of clouds are time consuming and tedious, while per-image annotated datasets only require one click for the whole image. In this paper, we present a cloud detector that has learned from a per-image annotated dataset of panchromatic satellite images, derived from the single-image dataset [18] appearing in Figure 1. The approach is generic and can be extended to any sort of single-band image.

## 2. RELATED WORK

**Feature based techniques.** A classical approach to detect clouds is through features or local descriptors [19, 20, 21]. With the significant progress in the design of features for several image processing applications, the scientific community has achieved good classification performance. The fact that we humans can identify clouds even locally (by texture, shape, color, etc) suggests that the right descriptors might be learned, either implicitly or explicitly. In this paper, our purpose is to let a network learn from scratch the proper descriptors for detec-

Work partly financed by IDEX Paris-Saclay IDI 2016, ANR-11-IDEX-0003-02, Office of Naval research grant N00014-17-1-2552, DGA Astrid project « filmer la Terre » n°ANR-17-ASTR-0013-01, MENRT. The Titan V used for this research was donated by the NVIDIA Corporation.



**Fig. 2:** Proposed GAN training in [18] to obtain the cloud-removal network. Legend :  $\mathbf{v}$  - cloudy image ;  $\mathbf{u}$  - cloud-free image ;  $\mathcal{G}$  - Generator ;  $\mathbf{u}'$  - Fake cloud-free image satisfying  $\mathcal{G}(\mathbf{v}) = \mathbf{u}'$  ;  $\mathcal{D}$  - Discriminator.

ting clouds, by submitting to it a minimally annotated dataset, in which images are simply tagged as cloudy or non-cloudy.

**Generative models for cloud removal.** The task of cloud removal in satellite images seems more challenging than the one of detecting clouds. In [18], a generator CNN, that was trained as in Figure 2, is capable of implicitly detecting clouds in satellite images and inpaints, in a credible way, those zones occluded by clouds. However, it is not clear how to access the implicit per-pixel classification applied by the network before inpainting. This generator was learned on per-image annotated datasets (also published in [18]) for single-image inputs and spatiotemporal inputs. These datasets consist of Sentinel-2 RGB+IR images annotated by thresholding the cloud coverage provided by [22], a well known per-pixel cloud/clear classification algorithm. In this paper, we will use both the generator and the datasets (see Figure 1) presented in [18].

### 3. CLOUD DETECTION

The neural network described in [18] successfully generates realistic cloud-free images  $\mathbf{u}'$  from cloudy images  $\mathbf{v}$ . Some hallucinations are observed, but the main differences between  $\mathbf{v}$  and  $\mathbf{u}'$  are indeed in regions where clouds are present; the network is implicitly detecting clouds. In this section we propose to use [18] to train a network to directly generate a cloud mask.

Let  $\mathcal{M}$  be a network that receives as input a cloudy image and from it computes a cloud mask  $\mathbf{m} := \mathcal{M}(\mathbf{v})$ . The architecture of  $\mathcal{M}$  is based on a ResNet with 9 residual blocks. Each residual block consists of two convolutional layers and a skip connection. While training,  $\mathcal{M}$  will be coupled with a degraded version of the  $\mathcal{G}$  network (only 2 ResNet blocks instead of 6 used in [18]) in order to help us reconstruct  $\mathbf{u}$  from  $\mathbf{v}$ . The idea, shown in Figure 3, is that the mask  $\mathbf{m}$  must identify the parts of  $\mathbf{u}'$  obscured by clouds, which are to be completed by information in  $\mathbf{v}$  :

$$\mathbf{m} * \mathbf{v} + (1 - \mathbf{m}) * \mathbf{u}' \approx \mathbf{u}. \quad (1)$$

Mathematically, we encourage such a mask with a loss func-

tion like

$$f_{\mathbf{u}}(\mathbf{x}) = \|\mathcal{M}(\mathbf{x}) * \mathbf{x} + (1 - \mathcal{M}(\mathbf{x})) * \mathcal{G}(\mathbf{x}) - \mathbf{u}\|_{L_1}, \quad (2)$$

where  $\mathcal{G}(\mathbf{x})$  generates images that look similar to  $\mathbf{x}$ , but not exactly equal, even for clear areas. This last behavior makes the training easier for the mask to be close to 1 in visible areas. Indeed, if the generator  $\mathcal{G}$  was such that  $\mathbf{x} = \mathcal{G}(\mathbf{x})$  for cloud-free areas, then the mask could be any value between  $[0, 1]$  without affecting the minimization in Equation 2.

Of course, if  $\mathcal{M}$  is trained to detect clouds only among cloudy images, there is a risk of failure in generalizing due to biased training data. Indeed, a network trained like this often detects clouds even in cloud-free images. This fact was confirmed in our experiments. At the end of the day, we propose to train the network  $\mathcal{M}$  with the following loss function :

$$\text{Loss}(\mathcal{M}) := f_{\mathbf{u}}(\mathbf{v}) + f_{\mathbf{u}}(\mathbf{u}). \quad (3)$$

This forces the network to correctly handle cloud-free images. Notice that the loss function of Equation 3 is not used to train  $\mathcal{G}$ , and only affects the weights of the network  $\mathcal{M}$ .

The single-image dataset was split randomly in three : training, validation and test. All RGB images were used to produce a panchromatic dataset. We trained the networks at the same time, but both  $\mathcal{G}$  and  $\mathcal{M}$  were detached from each other, i.e., not trained against the other. After training for 10 epochs over the whole training dataset and with a learning rate of 0.0002, results were already satisfactory, see Figure 4.

The output of  $\mathcal{M}$  is a float matrix. However, binary masks are the final result we want to provide. We cannot train directly for binary masks because most weights would remain unchanged due to zero-gradients. We define a simple procedure in order to binarize these float masks. This adds a new hyperparameter to our model, that will be set in validation. For each threshold  $t$ , we define the binarized mask at pixel  $p$  as

$$\mathbf{b}_t(p) := \begin{cases} 1, & \mathbf{m}(p) > t \\ 0, & \mathbf{m}(p) \leq t \end{cases}. \quad (4)$$

The final binary mask,  $\mathbf{b}$ , is selected as the  $\mathbf{b}_t$  minimizing the loss function of Equation 3. In practice, we do a grid search between 0 and 1 with a step of 0.01.  $t = 0.14$  is selected as the optimal threshold. Two visual examples of this passage from float masks  $\mathbf{m}$  to binary masks  $\mathbf{b}$  are shown in Figure 4.

### 4. RESULTS

	Clouds (Ours)	No-clouds (Ours)
Clouds (SIFT)	0.883 (15981177)	0.116 (2109476)
No-clouds (SIFT)	0.408 (641462)	0.591 (928685)

**Table 1:** Normalized confusion matrix over the test dataset (300 random pairs from the single-image dataset [18]).

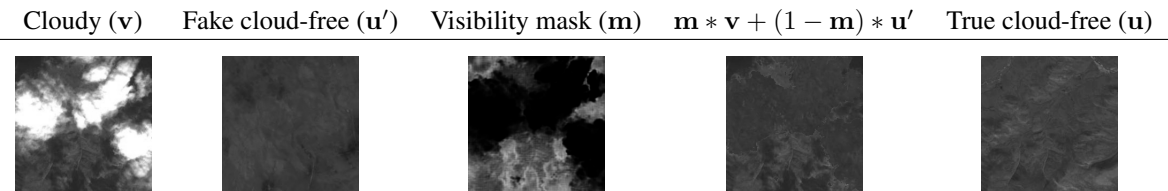


Fig. 3: Interpretation of Equation 1 : Clouds are detected in  $v$  as zones that are not as similar to  $u$  as those zones in  $u'$ .

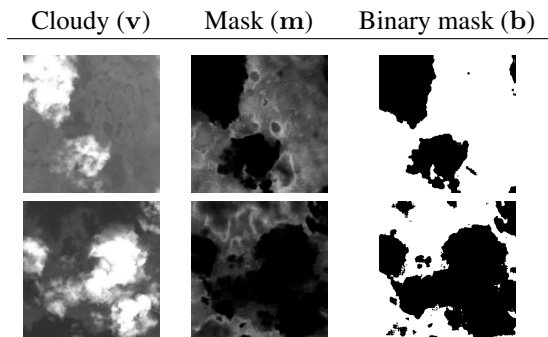


Fig. 4: Float masks  $m$  provided by the network  $\mathcal{M}$ .  $t = 0.14$  was used as threshold to pass from  $m$  to binary masks  $b$ .

	Clouds (Ours)	No-clouds (Ours)
Clouds (SIFT)	0.760 (12922400)	0.239 (4071420)
No-clouds (SIFT)	0.236 (631970)	0.763 (2035010)

Table 2: Normalized confusion matrix over 300 random tuples of images from the temporal dataset [18].

In this section we test the resulting mask from  $\mathcal{M}$  against the SIFT cloud detector [15, 16]. The two datasets (single and temporal) from [18] are used. The SIFT cloud detector receives as input all the information available : the two images  $v$  and  $u$  in the case of the single-image dataset; and the four images  $v, v_2, v_3$  and  $u$ , from the temporal-image dataset. Whereas our network ( $\mathcal{M}$ ) only receives the cloudy image  $v$ . Both cloud detectors are to output a mask for each cloudy image  $v$ .

The precision of  $\mathcal{M}$  in predicting the clouds of the SIFT cloud detector is of 0.96 for the single-image dataset; similarly, 95 percent of what  $\mathcal{M}$  predicts as clouds is indeed predicted as clouds by the SIFT cloud detector in the temporal dataset. The recalls are of 0.883 and 0.76 respectively. Table 1 and Table 2 provide the normalized confusion matrices of our cloud detector  $\mathcal{M}$  against the SIFT cloud-detector for the two datasets : single and temporal. Rows represent SIFT labels (and sum to one), while columns represent  $\mathcal{M}$  predicted labels. The reader will notice the recall of  $\mathcal{M}$  appearing in the top left corner of Tables 1-2.

The main difference (in number) in classification between  $\mathcal{M}$  and the SIFT cloud detector are for clouds. Indeed, 2109476 and 4071420 cloud labels (from the single and temporal datasets) of the SIFT detector are classified as visible

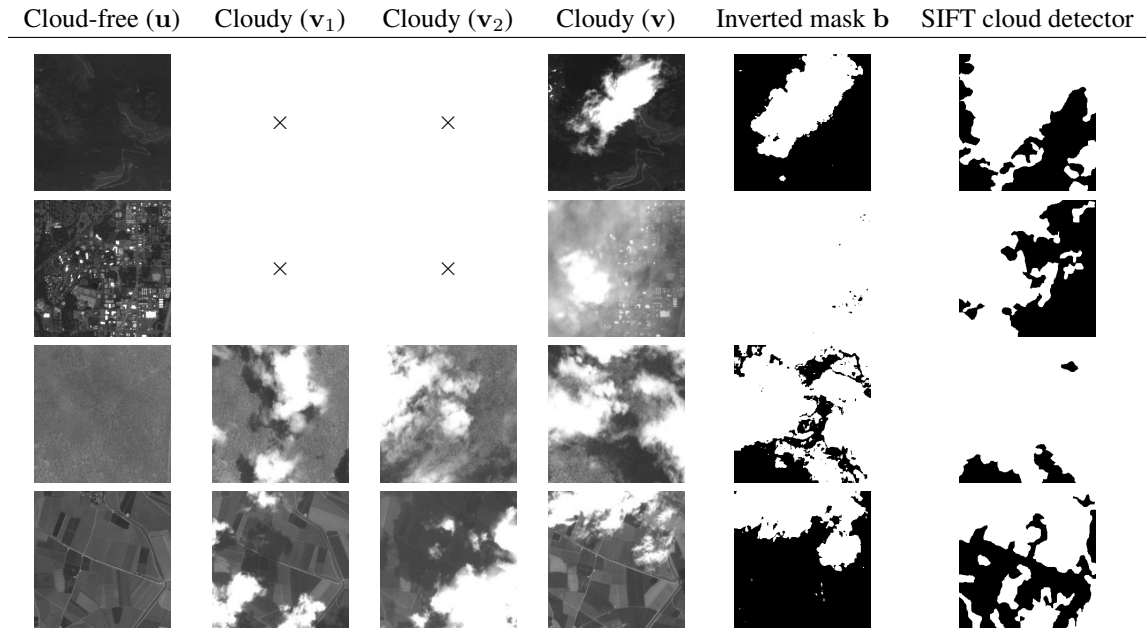
by the proposed cloud detector. This already points out that the SIFT cloud detector is being conservative, i.e. it prefers to classify as cloud rather than to be mistaken. On the other hand, there is also a difference (the biggest in percentage for the single dataset) in classification of visible zones. This behavior is reflected in Figure 5.

## 5. CONCLUSION

A convolutional neural network was proposed for panchromatic and single-frame cloud detection. The method produces a per-pixel classification into cloud or not cloud, even if only a per-image annotation was required for training, indicating which images contain clouds and which are cloud-free. Our experiments compare the proposed method to a state-of-the-art method, obtaining robust results, comparable or better, even if less rich information was used. Future work will focus on extending the method to any kind of input channels.

## 6. REFERENCES

- [1] G. Chandran and C. Jojy, "A survey of cloud detection techniques for satellite images," 2015.
- [2] S. Mahajan and B. Fataniya, "Cloud detection methodologies : variants and development—a review," *Complex & Intelligent Systems*, 2019.
- [3] R. R. Irish, "Landsat 7 automatic cloud cover assessment," 2000.
- [4] W. D. Zhang, M. X. He, and M. W. Mak, "Cloud detection using probabilistic neural networks," in *IGARSS*, 2001, vol. 5, pp. 2373–2375.
- [5] R. R. Irish, J. L. Barker, S. N. Goward, and T. Arvidson, "Characterization of the landsat-7 etm+ automated cloud-cover assessment (acca) algorithm," *PE&RS*, 2006.
- [6] P. L. Scaramuzza, M. A. Bouchard, and J. L. Dwyer, "Development of the landsat data continuity mission cloud-cover assessment algorithms," *TGRS*, April 2012.
- [7] A. Taravat, S. Proud, S. Peronaci, F. Del Frate, and N. Oppelt, "Multilayer perceptron neural networks model for meteosat second generation seviri daytime cloud masking," *Remote Sensing*, 2015.
- [8] A. Hollstein, K. Segl, L. Guanter, M. Brell, and M. Enesco, "Ready-to-use methods for the detection of



**Fig. 5:** Examples of masks generated by  $\mathcal{M}$  and the SIFT cloud detector [15, 16]. First two and last two rows correspond respectively to the single-image and temporal datasets [18]. The SIFT cloud detector receives all first four columns (when available) whereas  $\mathcal{M}$  only the fourth column. The typical convention for cloud-masks is adopted, i.e., white is cloud, black otherwise.

clouds, cirrus, snow, shadow, water and clear sky pixels in sentinel-2 msi images,” *Remote Sensing*, 2016.

- [9] D. Shin and J. K. Pollard, “Cloud height determination from satellite stereo images,” in *IEE Colloquium on Image Processing for Remote Sensing*, 1996.
- [10] C. Panem, S. Baillarin, C. Latry, H. Vadon, and P. Dejean, “Automatic cloud detection on high resolution images,” in *IGARSS*, 2005.
- [11] K. F. Manizade, J. D. Spinhirne, and R. S. Lancaster, “Stereo cloud heights from multispectral ir imagery via region-of-interest segmentation,” *TGRS*, 2006.
- [12] T. Wu, X. Hu, Y. Zhang, L. Zhang, P. Tao, and I. Lu, “Automatic cloud detection for high resolution satellite stereo images and its application in terrain extraction,” *P&RS*, 2016.
- [13] K. Sinclair, B. van Dienenhoven, B. Cairns, J. Yorks, A. Wasilewski, and M. McGill, “Remote sensing of multiple cloud layer heights using multi-angular measurements,” *AMT*, 2017.
- [14] D. Frantz, E. Haß, A. Uhl, J. Stoffels, and J. Hill, “Improvement of the fmask algorithm for sentinel-2 images : Separating clouds from bright surfaces based on parallax effects,” *RSE*, 2018.
- [15] T. Dagobert, J. M. Morel, C. de Franchis, and R. Grompone, “Visibility detection in time series of planetscope images,” in *IGARSS*, 2019.
- [16] T. Dagobert, R. Grompone, J. M. Morel, and C. de Franchis, “Temporal Repetition Detector for Time Series of Spectrally Limited Satellite Imagers,” *IPOL*, 2020.
- [17] R. Grompone, C. Hessel, T. Dagobert, J. M. Morel, and C. de Franchis, “Temporal repetition detection for ground visibility assessment,” in *ISPRS*, 2020.
- [18] V. Sarukkai, A. Jain, B. Uz Kent, and S. Ermon, “Cloud removal from satellite images using spatiotemporal generator networks,” in *WACV*, 2020.
- [19] T. Dagobert, R. Grompone, C. de Franchis, J. M. Morel, and C. Hessel, “Cloud Detection by Luminance and Interband Parallax Analysis for Pushbroom Satellite Imagers,” *IPOL*, 2020.
- [20] B. Tian, M. A. Shaikh, M. R. Azimi-Sadjadi, T. H. V. Haar, and D. L. Reinke, “A study of cloud classification with neural networks using spectral and textural features,” *IEEE transactions on neural networks*, 1999.
- [21] N. Laban, A. Nasr, M. ElSaban, and H. Onsi, “Spatial cloud detection and retrieval system for satellite images,” *IJACSA*, 2012.
- [22] A. Hollstein, K. Segl, L. Guanter, M. Brell, and M. Enesco, “Ready-to-use methods for the detection of clouds, cirrus, snow, shadow, water and clear sky pixels in sentinel-2 msi images,” *Remote Sensing*, 2016.