CONSISTENT COLORIZATION OF SENTINEL-2 IMAGES FOR GLOBAL PRODUCT GENERATION

Thomas Corpetti, Thomas Cusson, Antoine Lefebvre

CONSISTENT COLORIZATION OF SENTINEL-2 IMAGES FOR GLOBAL PRODUCT GENERATION

Thomas Corpetti[1,2], Thomas Cusson[2], Antoine Lefebvre[2]

[1]: CNRS, UMR 6554 LETG, Université Rennes 2, 35000 Rennes, France
[2]: Kermap, 1137a Av. des Champs Blancs, 35510 Cesson-Sévigné, France

ABSTRACT

This paper is interested with consistent rendering of 16-bits SENTINEL-2 images into 8-bits ones in order to produce visually sound global products (useful for basemap for example). Though the colorization of a single 16-bits data can efficiency be done with histogram stretching techniques for example, applying a unique transformation able to generate consistent data with enough details/contrasts whatever the content of the original image remains challenging. We propose here to train a neural network on a wide variety of images that have been manually enhanced to derive a unique model able to consistently recolor images and generate sound global products without mosaic effects.

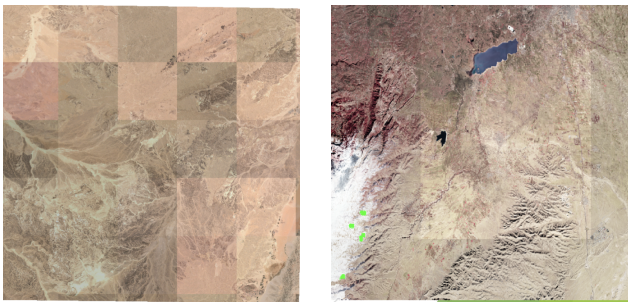
***Index Terms*—** Colorization, Sentinel-2, neural networks, feature losses, global products, basemaps

1. INTRODUCTION

SENTINEL-2 (S2) satellites constitute a breakthrough in the management of remote sensing data thanks to their continuous acquisition process and their high temporal resolution. Since the launching of S2 in 2015, many new use cases have been explored for a wide range of applications related either to agriculture monitoring and gesture [1, 2, 3], imperviousness estimation [4], urban survey (water areas [5], vegetation [6], mapping [7], human settlement [8] or Local Climate Zones for example [9]) or even water quality [10] and land cover classification [11].

Today, S2 images are widely used by the scientific community, and many operating systems have emerged over the 5 previous years. Faced with the massive flow of data to be managed, S2 images require reviewing the processes for managing such data: storage, diffusion, processing, and visualization among others. For this reason, a large number of global products has been developed to facilitate access to these data, including ergonomic platforms for visualizing and obtaining global information on sentinel data. These platforms can either be issued from public institutions (as the French Theia Platform, cf `https://www.theia-land.fr/`) or from private companies, such as the Nimbo platform from Kermap (`https://nimbo.earth/`).

This work has been partly supported by the ANR HIATUS (ANR-18-CE23-0025), financed by the French National Research Agency and by Kermap.



(a) (b)

Fig. 1. Illustration of mosaicing problems when a similar correction is applied to each snippet. Some bloc effects appear as well as a saturation in some areas (bottom left part of right figure)

Although Generative Adversarial Networks have enabled to manipulate in an efficient way the (re)-colorization of images in various contexts (RGB to multispectral, Radar to RGB, ...), few studies have been focused on the consistent mosaicking for remote sensing and in particular SENTINEL-2 images [12]. To generate consistent images from a visual point of view, an old problem has reappeared: the recolorization of 16-bit images into consistent RGB 8-bits data for visualization in platforms. Even if the visualization of a single tile can easily be performed by techniques based on scaling/stretching 16-bits data [13], the generalization over the entire earth lacks in efficiency. Such approaches are indeed focused on the correct reconstruction of a single image: the internal parameters (of stretching, of enhancement of the histograms, etc.) depend on each snippet and the global mosaic generates drastic consistency problems, as illustrated in Figure 1. On the other hand, because of the complexity and the variety of landscapes over the earth (especially mountains, deserts, ...), getting common parameters for the full range of images and colors available is almost impossible.

In this paper, we propose a colorization model efficient for all areas over the earth. To this end, we design a resid-

ual neural network that takes into account 16-bits 512×512 images and outputs 8-bits well-colorized images without any consistency problems. To train the network, a database of (16-bits,8-bits) pairs of data has been generated for a large variety of images all over the earth (forest, desert, mountains, cities, ...). Each 8-bit data has been manually recolored using specific editing software to provide images with visually pleasant properties (brightness, colors, etc).

2. MATERIALS AND METHOD

2.1. Colorization Neural Network

The network we designed is a residual U-net neural network summarized in Fig. 2. It is a quite usual network for this kind of regression problem.

2.2. Loss functions

Though the network in Fig. 2 is quite usual, the loss functions used to optimize it have to be taken carefully in order to render consistent visual properties of images. We used a threefold loss $\mathcal{L}(\hat{y}, y)$ between predicted ($\hat{y}$) and true (y) 8-bits images defined as:

$$\mathcal{L}(\hat{y}, y) = \lambda_m \mathcal{L}_m(\hat{y}, y) + \lambda_l \mathcal{L}_l(\hat{y}, y) + \lambda_f \mathcal{L}_f(\hat{y}, y) \quad (1)$$

with:

- $\mathcal{L}_m$ the mean absolute error (MAE) $|\hat{y} - y|$ inside a batch;
- $\mathcal{L}_l$ the mean local laplacian error $|\frac{\partial \hat{y}}{\partial x_1^2} + \frac{\partial \hat{y}}{\partial x_2^2} - \frac{\partial y}{\partial x_1^2} - \frac{\partial y}{\partial x_2^2}|$ inside a batch;
- $\mathcal{L}_f$ the mean feature loss $\|VGG(\hat{y}) - VGG(y)\|^2$ inside a batch;
- $\lambda_m + \lambda_l + \lambda_f = 1$, $\lambda_o \in [0, 1], \quad \forall o \in \{m, l, f\}$ the associated weights.

The rule of $\mathcal{L}_m$ is to ensure a global consistency of the image colors. However, as will be seen in the experimental section, two images with similar R, G, B values may have different perceptual colors and this loss alone is not optimal. The rule of the Laplacian loss is to ensure more local consistency since the Laplacian captures the information related to main edges in the images. Lastly, the term $\mathcal{L}_f$ extracts features issued from VGG-19 neural network [14] to keep meaningful information in terms of visual consistency [15].

3. EXPERIMENTAL EVALUATION & DISCUSSION

Our network has been trained on 5000 pairs of images taken worldwide and manually corrected using dedicated software. We used an Adam optimizer with a learning rate of 10^{-3}.

We evaluate the benefits of the different loss functions. We propose only a visual inspection since it is quite difficult to have a unique perceptual criteria for all possible areas around the world. One of the main criteria is to prevent from mosaic effects illustrated in figure 1 while ensuring a reliable visual consistency.

We present in the top of Fig. 3 some illustrations with the MAE loss only (i.e. $\lambda_m = 1, \lambda_l = 0, \lambda_f = 0$) for different epochs of the training. As can be seen, along epochs, though the loss function decreases, the representation of colours is not optimal. In addition, minimizing the MAE only tends to saturate the luminance, as seen in white areas that are more and more saturated. In the bottom of Fig. 3, the saturation at convergence is obvious for the two illustrations.

To overcome these difficulties, we have added the Laplace loss function to the MAE ($\lambda_m = 0.5, \lambda_l = 0.5, \lambda_f = 0$). Results can be seen in Figure 4. As can be observed, the saturation effect has disappeared and contrasts appear more clearly (as can be seen in the center of the image). However, as observed in highlighted areas in red, unrealistic colours have been added in some parcels. These surprising results may come from the different gradients but need to be investigated.

Lastly, when adding the feature loss, results are more consistent with the ground truth, as illustrated in Fig. 5 on a simple image. We also depicted in Fig. 6 some comparisons using products generated with the Theia plateform in France (see : `https://www.theia-land.fr/en`) with our results. Some large scale visualisations all over the world can be seen online in the Nimbo plateform developed by Kermap : `https://nimbo.earth/` and a snapshot is visible in Fig. 7

4. REFERENCES

[1] Tianxiang Zhang, Jinya Su, Cunjia Liu, Wen-Hua Chen, Hui Liu, and Guohai Liu, "Band selection in sentinel-2 satellite for agriculture applications," in *2017 23rd International Conference on Automation and Computing (ICAC)*. IEEE, 2017, pp. 1–6.

[2] Amanda Veloso, Stéphane Mermoz, Alexandre Bouvet, Thuy Le Toan, Milena Planells, Jean-François Dejoux, and Eric Ceschia, "Understanding the temporal behavior of crops using sentinel-1 and sentinel-2-like data for agricultural applications," *Remote sensing of environment*, vol. 199, pp. 415–426, 2017.

[3] Francesco Vuolo, Martin Neuwirth, Markus Immitzer, Clement Atzberger, and Wai-Tim Ng, "How much does multi-temporal sentinel-2 data improve crop type classification?," *International journal of applied earth observation and geoinformation*, vol. 72, pp. 122–130, 2018.

[4] Antoine Lefebvre, Christophe Sannier, and Thomas Corpetti, "Monitoring urban areas with sentinel-2a data: Application to the update of the copernicus high resolution layer imperviousness degree," *Remote Sensing*, vol. 8, no. 7, pp. 606, 2016.

[5] Xiucheng Yang, Shanshan Zhao, Xuebin Qin, Na Zhao, and Ligang Liang, "Mapping of urban surface water bodies from sentinel-2 msi imagery at 10 m resolution via ndwi-based image sharpening," *Remote Sensing*, vol. 9, no. 6, pp. 596, 2017.

[6] Michael Mau Fung Wong, Jimmy Chi Hung Fung, and Peter Pak Shing Yeung, "High-resolution calculation of the urban vegetation fraction in

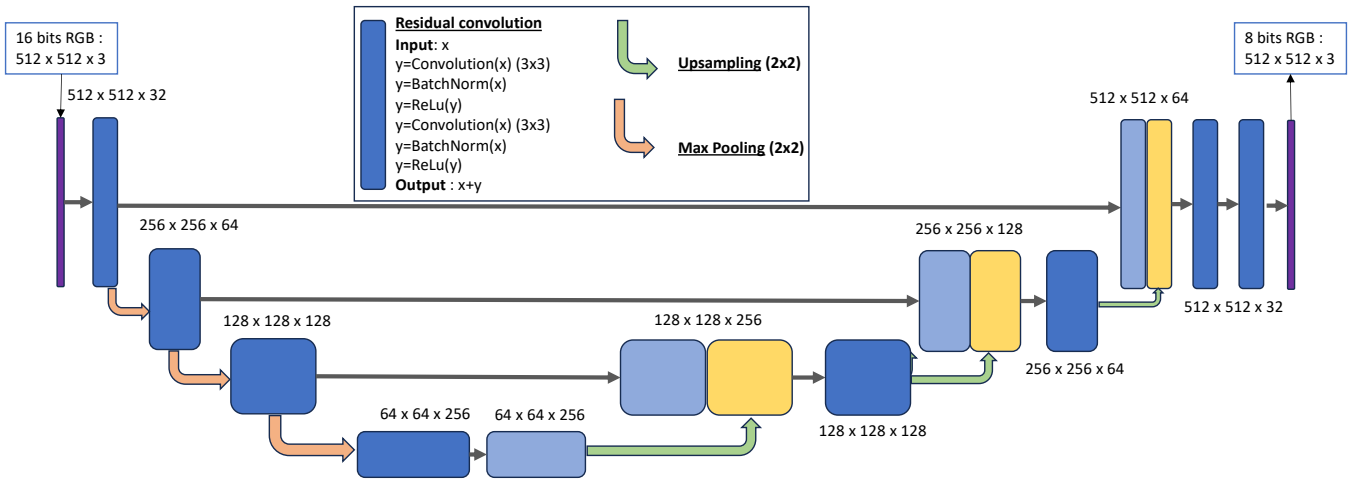
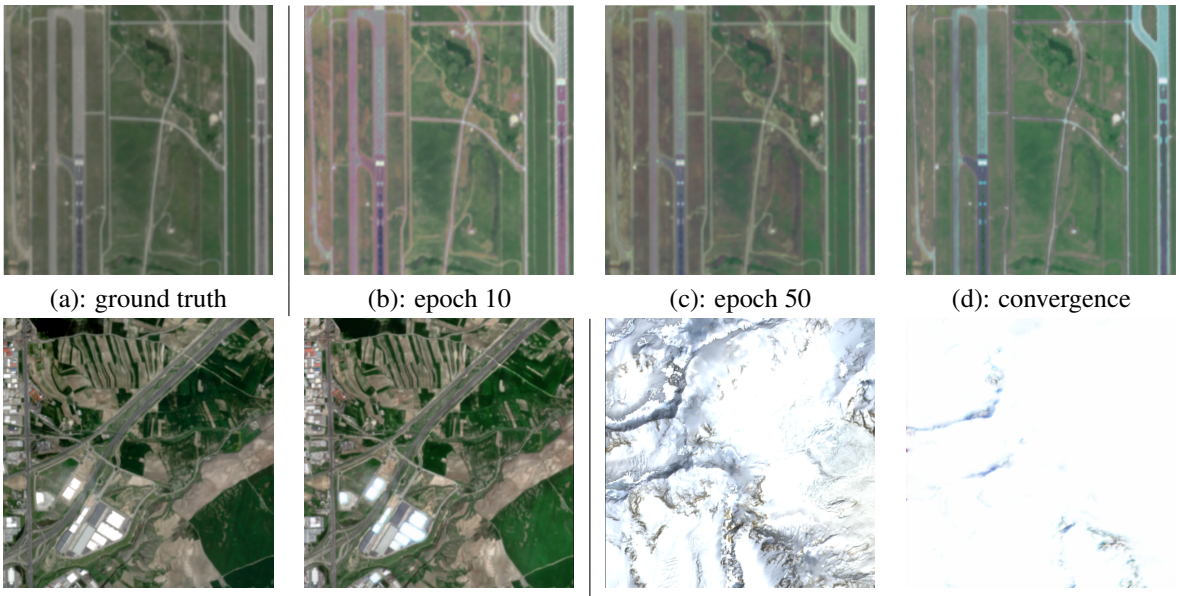


Fig. 2. Residual U-NET neural network for colorization of (512×512) 16-bits RGB images



(a): ground truth (b): epoch 10 (c): epoch 50 (d): convergence

(e): ground truth (f): convergence (g): ground truth (h): convergence

Fig. 3. **Illustration of difficulties with MAE loss only**. TOP. (a) : the ground truth image to retrieve; (b) : the reconstruction after 10 epochs; (c) the reconstruction after 50 epochs ; (d) the reconstruction at convergence. BOTTOM : (e,g) : two ground truths to retrieve from their corresponding 16-bits images; (f,h): estimated 8-bits images

(a): ground truth (b): convergence

Fig. 4. **Illustration of difficulties with MAE and Laplace losses**. (a) : the ground truth image to retrieve; (b) : the reconstruction at convergence.

(a): ground truth (b): convergence

Fig. 5. **Colorization with a combination of the three losses**. (a) : the ground truth image to retrieve; (b) : the reconstruction at convergence.

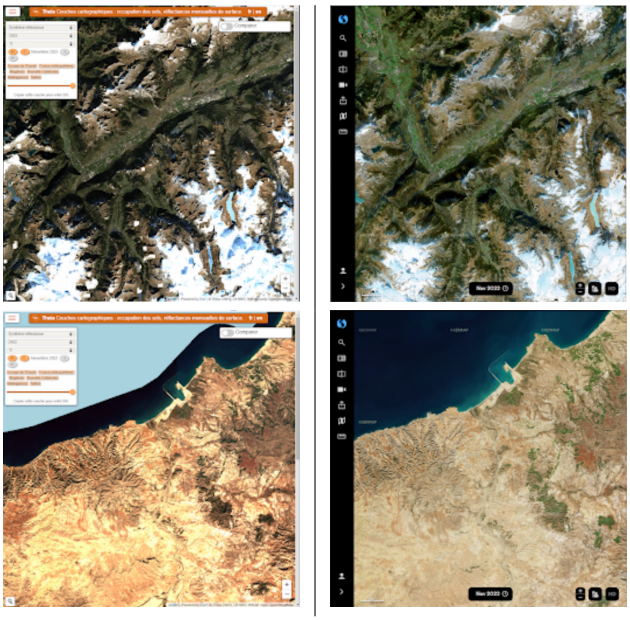


Fig. 6. **Comparisons of colorization with Theia (left) and Nimbo (right) plateforms**. As can be observed, with Theia plateform, images are more saturated and less details can be observed.

Fig. 7. **A global colorization over the world via mosaic of S2 tiles with our technique, visible in the nimbo platform**. No mosaic effects appear and colors are globally consistent.

the pearl river delta from the sentinel-2 ndvi for urban climate model parameterization," *Geoscience Letters*, vol. 6, no. 1, pp. 1–10, 2019.

[7] Zhongchang Sun, Ru Xu, Wenjie Du, Lei Wang, and Dengsheng Lu, "High-resolution urban land mapping in china from sentinel 1a/2 imagery based on google earth engine," *Remote Sensing*, vol. 11, no. 7, pp. 752, 2019.

[8] Chunping Qiu, Michael Schmitt, Christian Geiß, Tzu-Hsin Karen Chen, and Xiao Xiang Zhu, "A framework for large-scale mapping of human settlement extent from sentinel-2 images via fully convolutional neural networks," *Isprs Journal of Photogrammetry and Remote Sensing*, vol. 163, pp. 152–170, 2020.

[9] Chunping Qiu, Lichao Mou, Michael Schmitt, and Xiao Xiang Zhu, "Local climate zone-based urban land cover classification from multi-seasonal sentinel-2 images with a recurrent residual network," *ISPRS J. of Pho. and Rem. Sens.*, vol. 154, pp. 151–162, 2019.

[10] Kaire Toming, Tiit Kutser, Alo Laas, Margot Sepp, Birgot Paavel, and Tiina Nõges, "First experiences in mapping lake water quality parameters with sentinel-2 msi imagery," *Remote Sensing*, vol. 8, no. 8, pp. 640, 2016.

[11] Phan Thanh Noi and Martin Kappas, "Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using sentinel-2 imagery," *Sens.*, vol. 18, no. 1, pp. 18, 2018.

[12] S Liu, H Li, X Wang, L Guo, and R Wang, "Study on mosaic and uniform color method of satellite image fusion in large srea," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, pp. 1099–1102, 2018.

[13] Erik Reinhard, Wolfgang Heidrich, Paul Debevec, Sumanta Pattanaik, Greg Ward, and Karol Myszkowski, *High dynamic range imaging: acquisition, display, and image-based lighting*, M. Kaufmann, 2010.

[14] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[15] Alexey Dosovitskiy and Thomas Brox, "Generating images with perceptual similarity metrics based on deep networks," *Advances in neural information processing systems*, vol. 29, 2016.