

# Toward an Unsupervised Colorization Framework for Historical Land Use Classification

Rémi Ratajczak, Carlos F Crispim-Junior, Élodie Faure, Béatrice Fervers, Laure Tougne

▶ **To cite this version:**

Rémi Ratajczak, Carlos F Crispim-Junior, Élodie Faure, Béatrice Fervers, Laure Tougne. Toward an Unsupervised Colorization Framework for Historical Land Use Classification. International Geoscience and Remote Sensing Symposium (IGARSS 2019), IEEE, Jul 2019, Yokohama, Japan. 10.1109/IGARSS.2019.8898438 . hal-02122014v2

**HAL Id: hal-02122014**
**https://hal.science/hal-02122014v2**

Submitted on 21 Oct 2019

**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.

# TOWARD AN UNSUPERVISED COLORIZATION FRAMEWORK FOR HISTORICAL LAND USE CLASSIFICATION

*R. Ratajczak*[*]*, C.F. Crispim-Junior, L. Tougne*

Univ Lyon, Lyon 2
LIRIS F-69676
Lyon, France

*E. Faure, B. Fervers*

Unité Cancer et Environnement
Centre Léon Bérard
Lyon, France

## ABSTRACT

We present an unsupervised colorization framework to improve both the visualization and the automatic land use classification of historical aerial images. We introduce a novel algorithm built upon a cyclic generative adversarial neural network and a texture replacement method to homogeneously and automatically colorize unpaired VHR images. We apply our framework on historical aerial images acquired in France between 1970 and 1990. We demonstrate that our approach helps to disentangle hard to classify land use classes and hence improves the overall land use classification.

***Index Terms*—** Colorization, Classification, Land Use, Deep Learning, Texture filters

## 1. INTRODUCTION

Retrospective land use land cover (LULC) analysis is an ongoing research subject which has known a growing interest in the epidemiological community to assess the environmental factors leading to long latency diseases like cancer. In recent work, researchers developed per-subject exposition scores to agricultural pesticides based on geographical information system (GIS) metrics. These methods have proved their efficiency to overcome the absence of georeferenced epidemiological data at the scale of a country. A prerequisite to the calculation of these metrics is the availability of crop acreage semantics. However, historical LULC are hardly available with sufficient details (more than three land use classes) before 1990 in Europe, start of the Corine Land Cover program.

To face this limitation, automated computer vision methods have been investigated to estimate land use classes directly from monochromatic historical aerial images. They used low level texture filters, classical machine learning algorithms and deep convolutional neural networks. In these previous studies, monochromatic images have been used due to a lack of publicly available color and multispectral acquisitions before 1990 [1]. But what if we could automatically colorize grayscale images to improve land use classification?

[*]Thanks to the ADEME, Angers, France and to the LABEX IMU (ANR-10-LABX-0088/ ANR-11-IDEX-0007) for funding this work.

In this study, we present an unsupervised framework to automatically colorize historical VHR images. This framework is *unsupervised*, meaning that it can handle the absence of historical color images that would have served as ground truth data otherwise. It is also fully automated: It requires neither a reference image nor user scribbles at inference. Moreover, our framework aims at colorizing VHR images that cannot fit in the memory during the colorization process (up to 12k x 12k pixels in our experiments). As a result, we demonstrate that our framework is suitable for large scale GIS analysis by (1) generating qualitatively relevant color images, that should be easier to analyze for a human, and that it (2) helps to classify land use data.

The remaining of this paper is organized as follow. Section 2 describes related work. Section 3 presents our unsupervised colorization framework. Section 4 details our experiments and discusses the results we obtained.

## 2. RELATED WORK

We recall that the goal of a colorization algorithm is to regress a realistic $N$ dimensional color value for all the pixels of a grayscale image (*e.g.* R, G, B yield $N = 3$). Previous studies proposed to solve this problematic with user guided, hybrid [2] and automatic methods [3, 4]. Among them, authors proposed the use of deep convolutional neural networks (DCNN) following an encoder-decoder architecture in a supervised manner. In particular, [3] proposed to jointly learn a colorization and a classification model to improve the results' quality of the colorization network. In [2], the authors proposed an efficient way to train and use a DCNN with supplementary local hints (*i.e.* user scribbles) and global hints (*i.e.* reference images). On the other hand, generic image to image translation methods have been developed, like Pix2Pix [5] and CycleGan [6] (unsupervised). However, to our knowledge, this study is the first attempt to exploit an unsupervised colorization approach as a mean to improve the performance of an external classification pipeline in the context of LULC analysis.

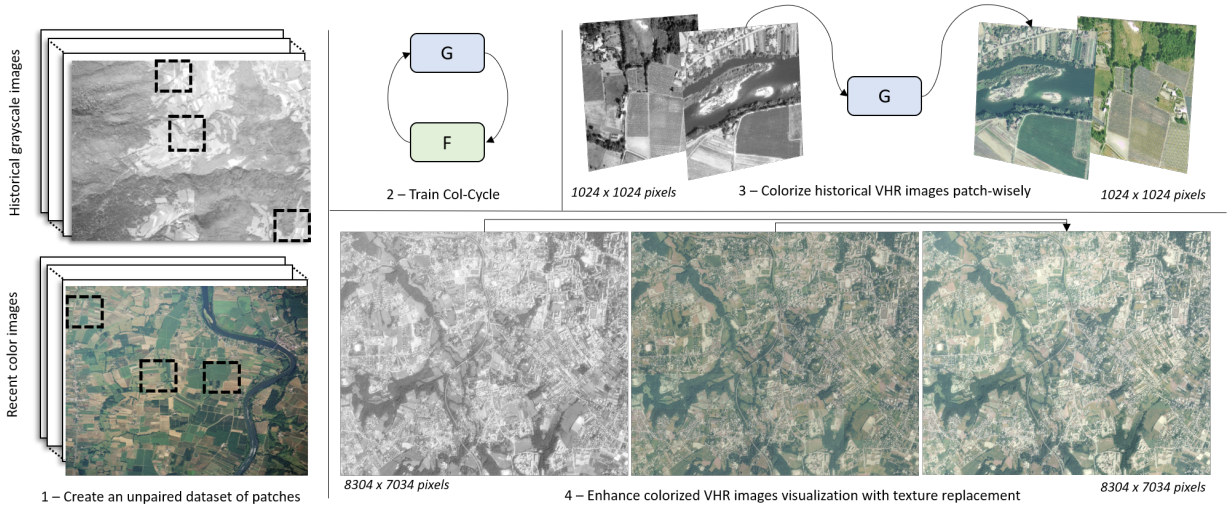



**Fig. 1**. The proposed unsupervised colorization framework.

## 3. FRAMEWORK

There are two essential components in our colorization framework. The first is Col-Cycle, an unsupervised generative adversarial network based on cycle consistent representations. The second is a simple yet efficient texture replacement approach to improve the visualization of VHR images made of a mosaic of patches that have been colorized independently. We decided to work patch-wise for practical computational reasons: Patches allow a sparse representation of VHR images, yielding to memory efficient and significantly faster processing. Our complete framework is schematized on Figure 1.

### 3.1. Col-Cycle Network

Col-Cycle is a relatively small unsupervised network based on a cyclic combination of generative adversarial neural networks (GANs). Similarly to existing unsupervised image-to-image translation approaches, our colorization model is based on two GANs that constrain a shared latent space. Both GANs have the same architecture, but they don't share their weights. The architecture of the generators in Col-Cycle follows an encoder-decoder architecture with only 1 input layer, 2 downsizing layers, 3 residual layers, 2 upsampling layers and 1 output layer. The input layer generates 64 feature maps using filters of 7x7 pixels. The output layer generates 3 feature maps, also using 7x7 pixels filters. All other layers are made of filters of 3x3 pixels. Downsizing layers divide the size of the input by two using the stride value. They also double the number of channels from the previous layer. The upsampling layers operate the opposite operation of the downsizing layers. We decided to use bilinear interpolation and classical convolutions instead of deconvolution filters in the upsampling layers to avoid checkerboard artifact effects. We used mean square error loss ($\mathcal{MSE}$) for the GAN losses defined by equation (1) and (2), and $L_1$ losses for the other constraints (see equations (3) and (4)) [6].

Let $X$ stands for the domain of the grayscale images, and let $Y$ stands for the domain of the color images, which is the RGB color space in our case. We define $GAN_{X\rightarrow Y} = \{G, D_Y\}$ and $GAN_{Y\rightarrow X} = \{F, D_X\}$ as the GANs that are respectively responsible for the grayscale to color and color to grayscale translations. In the above definitions, $G$ and $F$ represent the generative networks (*i.e.* generators) and $D_Y$ and $D_X$ represent the discriminative networks (*i.e.* discriminators). The generators aim to convert an image from a domain to another (*i.e.* grayscale $\leftrightarrow$ color). The discriminators try to classify the images between real and fake ones (*e.g.* colorized) in the domain specified by their indices. In a GAN manifold, the goal of the generator is to fool the discriminator. The generator and the discriminator are trained through GAN losses, respectively represented with equation (1) and (2) for $G$ and $D_Y$ (equivalent equations for domain $X$).

$$\mathcal{L}_{G_{GAN}} = \mathcal{MSE}(1, D_Y(G(I_X))) \tag{1}$$

$$\begin{aligned}\mathcal{L}_{D_{Y_{GAN}}} =& \mathcal{MSE}(1, D_Y(I_Y)) \\ &+\mathcal{MSE}(0, D_Y(G(I_X)))\end{aligned} \tag{2}$$

Additionally, a cycle is defined by the consecutive translation of an image from domain $X$ to $Y$ to $X$ (and vice versa), such as $\hat{I}_Y = G(I_X)$ and $\hat{I}_{X_{cycle}} = F(\hat{I}_Y)$, with $I_X \in X$ and $I_Y \in Y$. Cycle consistency is therefore expressed by targeting $I_X = \hat{I}_{X_{cycle}}$ (*resp.* $I_Y = \hat{I}_{Y_{cycle}}$). It is represented with a loss function applied to the generators and defined in equation (3).

$$\mathcal{L}_{cycle} = L_1(I_{Y_{cycle}}, I_Y) + L_1(I_{X_{cycle}}, I_X) \tag{3}$$

Moreover, to constrain the shared latent space and avoid initialization problems, we follow [6] by introducing an identity

loss as defined by equation (4). This loss helps to constrain the generator to produce realistic images according to the targeted domain.

$$L_{id} = L_1(G(I_Y), I_Y) + L_1(F(I_X), I_X) \quad (4)$$

### 3.2. VHR Image Reconstruction and Texture Replacement

The second step of our framework aims at generating VHR color images using color patches generated with Col-Cycle. To achieve this goal, we follow a simple yet efficient approach. We extract all possible non-overlaping patches of 1024 x 1024 pixels from a grayscale VHR image and we store their region of interest coordinates. We use Col-Cycle to colorize all these patches independently. Finally, we create a colorized VHR image by concatenating the colorized patches based on their original coordinates. This process is easy to implement and computationally efficient (no-overlap means no redundancy). However, we can observe on Figure 2 that non-overlaping patches seem to produce local non-linearity resulting in mosaic-like effects. These effects may be induced by differences in both the generated texture (*i.e.* lightness) and color components.

This observation lead us to the conclusion that 1) Col-Cycle did not learn to preserve the grayscale intensities (textures) very well and 2) may sometime produce VHR images that have spatially inconsistent colors. In this study, we only addressed the observation 1) using an intuitive approach. Starting from a VHR image colorized to the RGB color space, we translate it to the LAB color space. Then, we discard the L channel, and we replace it with the original grayscale image $I_X$. We call this approach texture replacement.

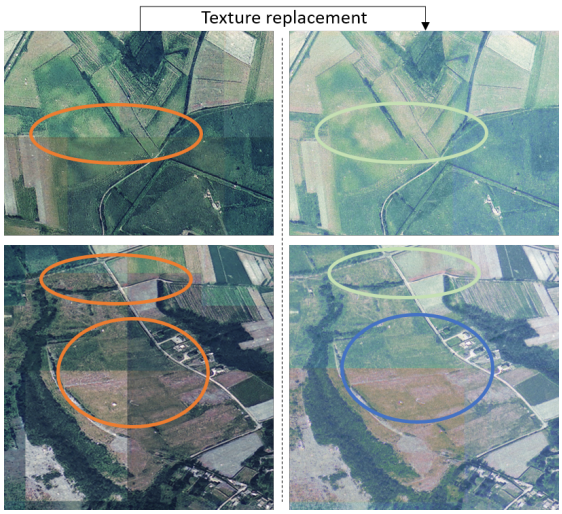



**Fig. 2**. Visual results with mosaic like effect (orange circles). Before (left) and after (right) texture replacement. Green circles indicate corrections, blue circles indicate remaining effects from color components.

As observed on Figure 2, texture replacement allows us to improve the visualization quality of the generated VHR images by significantly reducing the mosaic-like effect (green circles on Figure 2). However, it does not allow to erase the visible inconsistency in the color space, for which further work may be required (blue circles on Figure 2).

## 4. RESULTS AND DISCUSSION

We trained Col-Cycle for 200 epochs using Pytorch 0.4 and two Nvidia 1080 TI Graphical Processing Units. The training step was performed using 1702 patches of 1024 x 1024 pixels extracted from recent color images and 572 patches of the same size extracted from historical aerial images of France. Texture replacement was applied only at inference, directly on the colorized VHR images. We set up two experiments to evaluate our approach.

### 4.1. Mean Opinion Score study

We performed a Mean Opinion Score (MOS) study to evaluate the colorization quality on a patch-wise basis after 60 epochs of training with Col-Cycle.

We asked anonymous annotators to subjectively evaluate the color quality of 50 randomly selected images of 1024 x 1024 pixels (15 real, 35 colorized). We told the annotators that some of these images were real color images while other images were colorized using an algorithm. Based on this sole knowledge, annotators had to provide a quality score between 1 (worst) and 5 (better) for each image without knowing if the image was a real color image or not. Annotators had to indicate if they had prior experience working with remotely sensed data or not. In the following, annotators with prior experience will be call Experts (resp. Non-Experts).

A total of 28 annotators answered to our study, comprising 53.57% percent of Experts and 46.43% percent of Non-Experts. We summarized the mean opinion scores of these annotators on Figure 3. From an overall perspective, the colorized images obtained a MOS almost equal with the real color images, with a slight difference of only 0.1 point (3.08 versus 3.18) in favor of the real ones: The color quality of the generated images seems correct.

### 4.2. Application to classification

To evaluate our framework on a classification task, we applied Col-Cycle with texture replacement to colorize all the 81 VHR images used to create the HistAerial dataset[1]. Using these colorized images, we followed [1] and we extracted images of 100 x 100 pixels for 7 LULC classes, resulting in

[1]http://eidolon.univ-lyon2.fr/ remi1/HistAerialDataset/

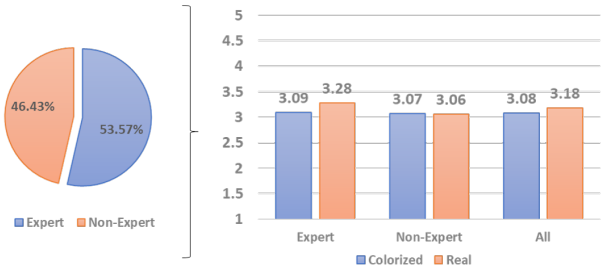



**Fig. 3**. Results of the Mean Opinion Score study to assess the color quality. Obtained with 28 annotators on 50 images of 1024x1024 pixels. Highest is best.

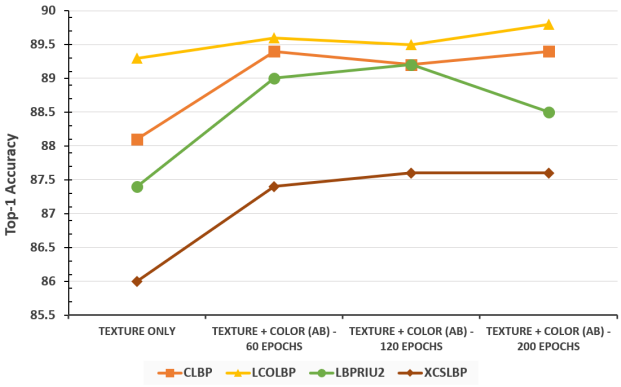



**Fig. 4**. Top-1 accuracy on the HistAerial dataset (7 classes) using low level texture descriptors and color features. Results obtained with Col-Cycle and texture replacement at different epochs.

a colorized version of the size-balanced subset of HistAerial with 6 000 images per class for this image size. These images were separated into train (63%), validation (27%) and test sets (10%) according to the original HistAerial dataset. For each colorized image represented in the LAB color space, we retrieved texture features on the L channel and color statistics on the AB channels. For the texture features, we compared the best performing handcrafted texture filter from [1] (LCOLBP) and three of its opponents (CLBP, XCSLBP, $\text{LBP}_{riu^2}$) using a 3-radii neighborhood. For the color features, we computed the mean and the standard deviation for the A and B channels independently. As supplementary features, we extracted 4 statistics (kurtosis, skewness, variance, maximum) from the histograms of the A and B channels, for a total of 12 color features per patch. We concatenated the texture and color features, over which we trained a random forest classifier with 100 decision trees, following the best classifier evaluated in [1]. Other parameters of the random forest were fixed using gridsearch from Scikit-learn.

The results are displayed on Figure 4 for different epochs of the Col-Cycle training. We observe that colorizing the historical aerial images tends to improve the top-1 classification on the dataset by an average of 1.3%. We also observe that training the network for much longer does not necessarily provide higher results. This phenomenon may be explained by the lack of control on the training of GAN-like networks: As opposed to classification network, it is difficult to know when to stop. Another explanation could be the absence of explicit constraints on the network to produce colors that could explicitly help the classification process.

## 5. CONCLUSION

In this article, we presented a framework to colorize VHR historical aerial images with qualitatively relevant colors for both GIS Experts and Non-Experts. We demonstrated that the generated colors help to classify LULC classes on historical aerial images, improving the state of the art by 1.3% in average. Thanks to these encouraging results, future work will focus on addressing the mosaic-like effects in the color space of the colorized VHR images by improving the colorization network and adding spatial consistency constraints. Multi-spectral image to image translation will also be investigated to further disentangle LULC classes. Our framework will be applied to define LULC classes required to assess environmental exposures in epidemiological studies.

## 6. REFERENCES

[1] R. Ratajczak, C. Crispim-Junior, E. Faure, B. Fervers, and L. Tougne, "Automatic Land Cover Reconstruction From Historical Aerial Images: An Evaluation of Features Extraction and Classification Algorithms," *IEEE TIP*, 2019.

[2] R. Zhang, J.-Y. Zhu, P. Isola, X. Geng, A. S. Lin, T. Yu, and A. A. Efros, "Real-time user-guided image colorization with learned deep priors," *ACM TOG*, 2017.

[3] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Let there be Color!: Joint End-to-end Learning of Global and Local Image Priors for Automatic Image Colorization with Simultaneous Classification," *ACM TOG*, 2016.

[4] G. Larsson, M. Maire, and G. Shakhnarovich, "Learning representations for automatic colorization," in *IEEE ECCV*, 2016.

[5] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *IEEE CVPR*, 2017.

[6] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *IEEE ICCV*, 2017.