



**HAL**  
open science

# Multispectral Texture Synthesis using RGB Convolutional Neural Networks

Sélim Ollivier, Yann Gousseau, Sidonie Lefebvre

► **To cite this version:**

Sélim Ollivier, Yann Gousseau, Sidonie Lefebvre. Multispectral Texture Synthesis using RGB Convolutional Neural Networks. 2024. hal-04719766v2

**HAL Id: hal-04719766**

**<https://hal.science/hal-04719766v2>**

Preprint submitted on 21 Oct 2024

**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.

# Multispectral Texture Synthesis using RGB Convolutional Neural Networks

Sélim Ollivier, Yann Gousseau, and Sidonie Lefebvre

**Abstract**—State-of-the-art RGB texture synthesis algorithms rely on *style distances* that are computed through statistics of deep features. These deep features are extracted by classification neural networks that have been trained on large datasets of RGB images. Extending such synthesis methods to multispectral images is not straightforward, since the pre-trained networks are designed for and have been trained on RGB images. In this work, we propose two solutions to extend these methods to multispectral imaging. Neither of them require additional training of the neural network from which the second order neural statistics are extracted. The first one consists in optimizing over batches of random triplets of spectral bands throughout training. The second one projects multispectral pixels onto a 3 dimensional space. We further explore the benefit of a color transfer operation upstream of the projection to avoid the potentially abnormal color distributions induced by the projection. Our experiments compare the performances of the various methods through different metrics. We demonstrate that they can be used to perform exemplar-based texture synthesis, achieve good visual quality and comes close to state-of-the art methods on RGB bands.

**Index Terms**—Texture Synthesis, Multispectral Imaging.

## I. INTRODUCTION

IN recent years, a great deal of effort has been devoted to the development and design of multispectral imagers, capable of capturing images of a scene in multiple (usually between 2 and 20) spectral bands in the infrared or visible range. These devices serve various purposes, such as detecting aircraft and UAVs, and analyzing satellite data for Earth observation, including vegetation, atmosphere, ocean, and land assessments. Moreover, numerous deep learning methods have been developed in response to growing interest for multispectral imaging and its applications.

Yet, the training of efficient classifiers and detection systems requires access to a large amount of data. Similarly, the augmentation of existing databases is highly beneficial for the performance assessment of image processing algorithms and optical sensors. In both cases, it is necessary to have reference scenes and to be able to model the variability of objects of interest and background environments. Given the significant costs associated with data acquisition and annotation, the ability to rapidly synthesize images that accurately replicate the radiometric properties and textures of real-world backgrounds is highly advantageous.

S. Ollivier is with the LTCI, Télécom Paris, Institut Polytechnique de Paris, 91120 Palaiseau, France and also with DOTA/AILab, ONERA, Université Paris-Saclay, 91120, Palaiseau, France (e-mail: selim.ollivier@onera.fr).

Y. Gousseau is with the LTCI, Télécom Paris, Institut Polytechnique de Paris, 91120 Palaiseau, France.

S. Lefebvre is with the DOTA/AILab, ONERA, Université Paris-Saclay, 91120, Palaiseau, France.

Texture modeling and analysis was initially limited to mono-band images for which tools such as local statistics extraction, co-occurrence matrices, wavelet transforms or fractal based techniques have been developed to account for the spatial behaviour and patterns. Extensions of these techniques to color and multi/hyperspectral images followed afterwards. In this context, the color or spectral statistics need to be taken into account. For this purpose, dimensionality reduction [19] or 3D power spectral density and histogram matching [29] have been introduced for texture characterization or synthesis.

Since the foundational work of Gatys et al [8], pre-trained convolutional neural networks (CNN) have been the state-of-the-art in the field of RGB exemplar-based texture synthesis. Most subsequent works used statistics extracted from the feature maps of these networks to model texture through *style distances*. This approach is particularly effective because it capitalizes on the network’s powerful description and recognition abilities, which stem from training on large datasets. Recently, generative networks for the synthesis of satellite data and multi/hyperspectral images have emerged. These models are often based on architectures that have been proven to be effective for RGB images and manage to generate realistic multispectral images by training new neural networks on multi/hyperspectral images [15] [20] [18] [1]. However, to the best of our knowledge, no method has been proposed to perform multi/hyperspectral texture synthesis that leverage style distances relying on statistics extracted from pre-trained networks as those used for RGB images. Indeed, This is not straightforward, as the widely used VGG19 network, trained on ImageNet, is designed for 3-channel images. Although interesting works have tried to bypass pre-trained networks as VGG-19 to define style distances, results remain of lower quality [34]. Moreover, training or fine-tuning a new network would need large multi/hyperspectral texture databases.

To avoid this, the present work introduces two different strategies to build multispectral style distances based on pre-trained CNN. These distances are then successfully used to perform multispectral texture synthesis. Our approach can be applied to a wide range of natural textures. Our numerical analyses target cloud backgrounds, that are useful for various applications : cloud cover forecasting to inform acquisition decisions, and thus optimize the planning of agile satellite constellations for Earth observation, but also to predict the availability of optical communication links between ground stations and satellites. Additional use cases involve the detection of small objects, such as UAVs, against a clouded sky background. Another important application area lies in the context of cloud removal in optical satellite images, which requires pairs of images with and without clouds for training

algorithms. Czerkawski et al. [7] designed an open-source python toolbox based on Perlin noise enabling the synthesis of diverse simulated pair data, with controllable parameters to adjust cloud appearance. This tool has been evaluated as a source of training images supply for cloud detection and cloud removal algorithms, and has led to good performances compared to the same algorithms trained only on real data.

Our contributions can be summarized as the following:

- 1) we introduce two style distances: a stochastic style distance and a projected style distance that can be computed from a pre-trained RGB CNN. This in turn allows to synthesize multispectral images using an optimization process similar to the one of [8]. Our approach is not specific to texture modeling and could also be used to extend other distances based on pre-trained RGB networks to multispectral imaging without having to train new networks for that purpose.
- 2) to correct the color domain shift induced by the projection, we propose to rely on color transfer, meant to align the projected image with a more typical color distribution seen during training by the pre-trained CNN involved in the style distance. We then empirically investigate its usefulness.

Section II reviews related work on multispectral texture modeling, and section III presents our proposed methodology in detail. Texture synthesis experiments on 130 multispectral cloud field images from the Sentinel-2 mission are conducted in section IV. We conclude this paper in section V.

## II. RELATED WORK

*a) Multi and hyperspectral texture modeling:* Texture synthesis and modeling first emerged in applications linked to monochromatic images and extensions to colour images were proposed later on. For color texture synthesis, considering each color channel independently is often not satisfying due to inter-channel correlations. Early methods for color texture synthesis apply a Principal Component Analysis (PCA) before working channel-by-channel [10], possibly modeling inter-channel correlations as in [23]. For hyperspectral images, Mercier et al. [19] uses an Independent Component Analysis before applying wavelet decomposition for texture characterization. Sarkar et al. [29] proposes to constrain the multi-dimensional histograms of images together with a 3D Fourier transform accounting for intra-band spatial dependency to perform texture synthesis.

From a modeling perspective, textures were originally often analysed through dependencies between nearby pixels, which led to the use of Markov models. Adding the assumption of a Gaussian response for each vertex of the graph representing the image to be analysed led to a model well-known as a random Gauss-Markov field. These models were first leveraged for anomaly detection in multispectral and hyperspectral images by Schweizer and Moura [30] [31]: the inter-pixel Markov component allows textures to be taken into account while the underlying Gaussian assumption allows a simple modeling of the spectral response. Some Gauss-Markov

models with different way of accounting for second order spectral information were proposed by Rellier et al [27] [26], for hyperspectral textures classification. All these models are based on sets of simplifying assumptions taking into account the properties of the hyperspectral images analysed. These assumptions relate in particular to the constancy across the spectrum of the Markov parameters governing the textures and on the spectral covariance modeling of the textures across the spectrum. Spatial parameters are typically assumed to be constant throughout the spectrum, which means that the multispectral texture is considered to be identical at each wavelength, up to a factor of variance. The Schweizer and Moura formulation implies that each scalar pixel interacts with its neighbours in all three dimensions (spatial and spectral) and the inclusion of an inter-band Markov link in the model implicitly imposes identical inter-band behaviour between all pairs of neighbouring bands. Rellier et al proposed to restrict the pixel neighbourhood to spatial neighbours. The aim is then to model the spatial interactions between hyperpixels, and thus the expression of the spatial part of the precision matrix loses its interband component. Although more flexible variations have been proposed, these underlying simplifying assumptions limit the expressiveness of the textures generated, and texture synthesis methods based on generative networks offer a promising alternative.

To assess the quality of the synthesis, conventional texture metrics, such as local binary pattern and co-occurrence matrices, have been exploited, with within and between channel interactions, in a recent comparison of color imaging and hyperspectral imaging for texture classification [22]. Recently, Chu and co-authors focused on accounting simultaneously for spectral and spatial distribution. They proposed a four dimensional texture metrics, GHOST [3], encompassing joint metrological assessment of spectral and spatial properties. They also conceived an alternative feature, Relative Spectral Difference Occurrence Matrix (RSDOM) [4] [5], which considers the joint probability of spectral distribution and of pixels spatial arrangement and consists of pixel differences with spectral reference and spatial neighbors, and emphasized its contribution in the hyperspectral image texture classification scheme on HyTexiLa database [13].

*b) Deep generative models for multispectral and hyperspectral images:* The most recent generative neural networks methods such as GANs (Generative adversarial Networks) and diffusion models have only very recently started to be applied on multispectral satellite data. The recent work of [1] compares, both with a spectral signature analysis and a modified Fréchet inception distance, the performances of StyleGAN3 and proGAN architectures to reproduce Sentinel-2 images. Khanna et al. [14] propose DiffusionSat, a generative foundation model for remote sensing data based on the latent-diffusion model architecture of StableDiffusion [28], dedicated to inverse problems such as multispectral input super-resolution, temporal prediction, and in-painting. In order to alleviate the computational complexity due to the high spectral dimensionality, UnmmixDiff [37] has recently been proposed, and consists in coupling an unmixing

autoencoder and a diffusion model. The synthesis is thus performed in the abundance space, which enables to obtain realistic objects’ distribution in abundance maps. However, this method focuses on the spectral profiles consistency rather than on the texture representativity of the different natural background components, such as forest, sky, grassland... Another publication [18] emphasizes the use of a diffusion model for the synthesis of hyperspectral images, but with a conditional RGB image as input. They introduce a conditional vector quantized generative adversarial network (VQGAN) to map hyperspectral data into a low-dimensional latent space, in which the diffusion model is easier to train. They pay particular attention to the diversity of the generated hyperspectral images, by proposing two new metrics focusing on spectral content diversity. The syntheses are evaluated using both pixel-wise metrics such as RMSE, spectral angle mapper (SAM), and PSNR, and the structure similarity index metric SSIM for evaluating the quality of spatial information. Additionally, foundation models—trained on vast unlabelled datasets and fine-tuned for specific tasks using limited annotated data—have seen increased use in satellite imagery over the past two years. FG-MAE [35] and SatMAE [6] have been developed for multispectral Sentinel-2 images, and DOFA [36] has been very recently proposed to adaptively integrate various data modalities into a single framework, leveraging the concept of neural plasticity in brain science. These foundation models will undoubtedly contribute to strengthen multispectral image analysis capabilities. These models have only been applied on satellite data, and mostly on data taken over cities, not on textured natural backgrounds.

*c) Neural texture synthesis:* The work from Gatys et al. [8] was a major breakthrough for exemplar-based texture synthesis. The main idea of this work, that we will detail in Section III-A, is to constrain some statistics on deep image features extracted by classical classification networks that have been trained on large image datasets (such as the VGG19 network trained on ImageNet). More precisely, second order statistics between deep features are computed on an exemplar image through Gram matrices and images are then synthesized by optimizing the pixel responses to match these matrices. This results in very realistic synthesized texture images, although at the price of a relatively heavy optimization procedure for each synthesis. Many works have followed, all exploiting the central idea of measuring distance between texture images through the squared difference between their Gram matrices, distances that in this paper we will call *style distances*. A particular interest was devoted to develop feed-forward architecture to speed-up the synthesis [33], use the style distances in the GAN framework [12] or to develop auto-encoder texture synthesizers [2]. To the best of our knowledge, none of these state-of-the-art texture synthesis approaches have been adapted to the multispectral or hyperspectral framework, which is the purpose of the present paper.

### III. METHODS

#### A. Neural statistics for texture synthesis

As recalled in the previous section, the method from Gatys et al. [8] rely on the use of statistics of deep features. These features are classically obtained from the VGG19 network. It consists of a succession of convolutional layers. Each layer  $l$  is made of  $N_l$  convolutional filters and provides a feature map  $F^l \in \mathbb{R}^{N_l \times M_l}$  (where  $M_l$  is the number of pixels of the feature map). The Gram matrices were chosen by the authors to form a summary statistic that is blind to spatial information:

$$\forall l \in \{1, \dots, L\}, \quad G^l = F^l (F^l)^T. \quad (1)$$

The set  $G = \{G^1, \dots, G^L\}$  of Gram matrices computed from the network in response to an image constitutes a stationary multi-scale description of the texture. It can further be used to define a style distance between two images  $I$  and  $\hat{I}$  with Gram matrices  $G$  and  $\hat{G}$ :

$$\mathcal{L}_{style}(I, \hat{I}) = \sum_l w_l (G^l - \hat{G}^l)^2, \quad (2)$$

where the  $w_l$  are chosen weighting factors.

This style distance was used in [8] to achieve a clear breakthrough for texture synthesis. It was also later used to train other neural networks as a loss function [33] [16] [2]. In this work, we follow the exemplar synthesis process presented by [8]: given an exemplar image  $I$ , a synthetic image  $\hat{I}$  is initialized with a Gaussian white noise and then gradient descent is used to minimize the style distance between the two images  $I$  and  $\hat{I}$ . The result of this optimization is expected to match the Gram matrices of the exemplar. In the following, we chose to initialize the synthetic image following a Gaussian noise with the mean and covariance of the original image as it showed better performances.

This style distance makes use of the Gram matrices to describe the feature maps distributions, but other statistics as covariance matrices (i.e. Gram matrices of the centered feature maps) have also shown good performances for texture synthesis.

We chose to use the latter in our work.

The main obstacle to the extension of CNN based texture synthesis processes to multispectral imaging is the design and the training of a multispectral CNN from which to extract feature maps statistics. To address this, we introduce a stochastic style distance in section III-B and a style distance that makes use of projection in section III-C. In section III-D, we add a color operation to the projection to align the color distributions with those of the images the network has been trained on. These distances enables one to use the VGG19 network, pre-trained on ImageNet, that has become standard for 3-channels images.

#### B. Stochastic style distance

The difficulty to extend the approach recalled in the previous paragraph to multispectral images is to constrain the statistics of an image of dimension  $N$  with an optimization procedure



that relies on a RGB CNN. To do so, we decide to constrain the statistics of all possible 3-channels images obtained from the channels of the multispectral image.

We consider an exemplar multispectral image  $I$  with  $N$  spectral bands and  $\hat{I}$  its synthetic counterpart. We propose to randomly draw a triplet  $J$  of integers in  $\{1, \dots, N\}$  at each iteration of the optimization. Then, to evaluate the style distance  $\mathcal{L}_{style}(I_J, \hat{I}_J)$  between the images  $I_J$  and  $\hat{I}_J$  made of the spectral bands indexed by the integers of  $J$ . Finally, the gradient of this quantity is used to take an optimization step on  $\hat{I}$ .

During the optimization, one can average the style distance over batches of uniformly drawn triplets in a way that is reminiscent with stochastic gradient descent algorithms. It ensures to take into account more spectral bands and correlations at each iterations. Moreover, it regularizes the objective function.

The global minimization objective of our algorithm is the expectation of the style distance between 3-channels images corresponding to the spectral bands indexed by a triplet  $J$  that follows the uniform law. This amounts to averaging the style distance over all possible triplets :

$$\mathcal{L}_{style}^{MS}(I, \hat{I}) = \mathbb{E}_{J \sim \mathcal{U}(\mathcal{J}_N)} \left[ \mathcal{L}_{style}(I_J, \hat{I}_J) \right] \quad (3)$$

$$= \frac{1}{N_J} \sum_{J \in \mathcal{J}_N} \mathcal{L}_{style}(I_J, \hat{I}_J) \quad (4)$$

Where :

- $\mathcal{J}_N$  is the set of triplets of integers in  $\{1, \dots, N\}$  of cardinal  $N_J$ ,
- $\mathcal{U}(\mathcal{J}_N)$  is the uniform distribution on  $\mathcal{J}_N$ ,
- $\mathbb{E}_{J \sim \mathcal{U}(\mathcal{J}_N)}$  is the mathematical expectation under the condition that  $J$  follows a uniform distribution,
- if  $J$  is a triplet,  $I_J$  is the 3-channels image made of the 3 spectral bands of  $I$  indexed by  $J$ ,
- $\mathcal{L}_{style}$  is the classical style loss for RGB images using a VGG19 pre-trained network described by [8], recalled in Equation (2).

The proposed synthesis algorithm is detailed in Algorithm 1 for  $n$  iterations of the gradient descent algorithm and  $B$  triplets drawn at every iteration. The mean and covariance of  $I$  are denoted as  $\mu_I$  and  $\Sigma_I$ .

---

**Algorithm 1** Texture synthesis with stochastic style loss

---

```

 $\hat{I} \sim \mathcal{N}(\mu_I, \Sigma_I)$  ▷ Gaussian noise initialization
for  $i \in \{1, \dots, n\}$  do
   $\mathcal{L} \leftarrow 0$ 
  for  $j \in \{1, \dots, B\}$  do
     $J \sim \mathcal{U}(\mathcal{J}_N)$  ▷ Uniformly drawn triplet
     $\mathcal{L} \leftarrow \mathcal{L} + \mathcal{L}_{style}(I_J, \hat{I}_J)$ 
  end for
  Take a gradient step on  $\hat{I}$  with gradient  $\nabla \mathcal{L}$ 
end for
return  $\hat{I}$ 

```

---

Our method ensures to browse every spectral bands and consider the correlations between each of them in the course

of the optimization. However, a drawback of this method is that it may need a relatively large number of triplets to reach a good overall control over multi-spectral statistics, resulting in a high computational cost. In the next section, we introduce a more simple and faster alternative approach.

### C. Projected style distance

The second style distance that we propose aims to take advantage of the redundancy of the spectral information. Indeed, the spectral information that is contained in  $N$  dimensions often can be reduced in a lower number of dimensions. Therefore, we consider a projector  $P : \mathbb{R}^N \rightarrow \mathbb{R}^3$  that maps multispectral images with  $N$  spectral bands into RGB-like images with 3-channels. The proposed style distance consists then in the classical style distance computed on projected images :

$$\mathcal{L}_{style}^P(I, \hat{I}) = \mathcal{L}_{style}(P(I), P(\hat{I})) \quad (5)$$

The use of the projected style distance for exemplar-based texture synthesis is detailed in algorithm 2.

---

**Algorithm 2** Texture synthesis with projected style loss

---

```

 $\hat{I} \sim \mathcal{N}(\mu_I, \Sigma_I)$  ▷ Gaussian noise initialization
for  $i \in \{1, \dots, n\}$  do
   $I_P \leftarrow P(I)$ 
   $\hat{I}_P \leftarrow P(\hat{I})$ 
   $\mathcal{L} \leftarrow \mathcal{L}_{style}(I_P, \hat{I}_P)$ 
  Take a gradient step on  $\hat{I}$  with gradient  $\nabla \mathcal{L}$ 
end for
return  $\hat{I}$ 

```

---

### D. Color transfer for neural statistics extraction

We observe that the use of a projection can lead to unusual color distributions that can have an influence on the quality of the description provided by the feature maps as the network may not have been trained on such distributions. This domain shift can affect the quality of the description provided by the CNN. Examples of the color that are induced by a PCA projection for cloud field images taken from Sentinel-2 mission are available in Figure 1.

To circumvent this, we propose to perform a color transfer operation after the projection and before the use of the VGG-19 network. The target distribution of this transfer operation is simply the color distribution of a natural image, that therefore acts as a color palette. The idea here is to send back the considered 3-channels image in regions of the color space that are likely to be closer to the regions that have been explored during the training of the network.

To define such a color transfer operation, we chose to describe the color distributions of the images by the first and second order statistics, assuming that it gives a first satisfactory and sufficient approximation for our application. Hence, we consider an affine transformation that maps the projected

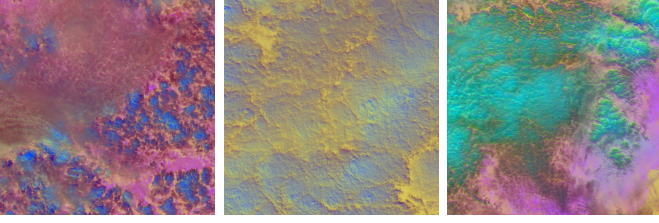


Fig. 1: Examples of unusual color distributions induced by the PCA projection of Sentinel-2 cloud field images.

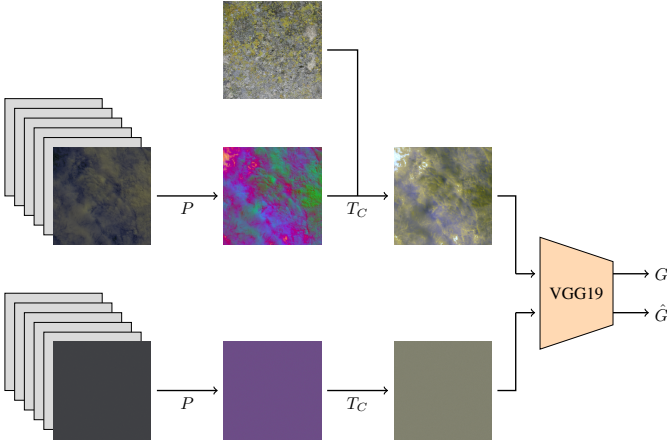


Fig. 2: Scheme of the proposed use of color transfer for multispectral texture synthesis. A projection  $P$  is used to obtain 3-channels images from the exemplar and synthetic multispectral images. Then, a color transfer operator  $T_C$  is applied to both images to match their color statistics to the one of a reference palette image. Finally, statistics of deep features are extracted from the so-obtained images to perform texture synthesis.

image  $I_P$  with color statistics  $(\mu_P, \Sigma_P)$  to an image matching the color statistics  $(\mu_C, \Sigma_C)$  of a palette image  $I_C$ . We denote  $T_{P \rightarrow C}$  such a transformation and, for an image  $I$ , we have:

$$T_{P \rightarrow C}(I) = \mu_C + L_C L_P^{-1} (I - \mu_P) \quad (6)$$

where the  $L_P$  and  $L_C$  matrices are obtained using Cholesky decomposition of covariances matrices:

$$\Sigma_P = L_P L_P^T \quad (7)$$

$$\Sigma_C = L_C L_C^T \quad (8)$$

As it is possible to define the converse transformation  $T_{C \rightarrow P}$ , such a transformation happens to be reversible so that it conserves the spatial information of the images and only act on the color of the pixels to match the first and second order statistics.

The corresponding style distance is therefore (see Figure 2 for illustration):

$$\mathcal{L}_{style}^{P+C}(I, \hat{I}) = \mathcal{L}_{style}(T_{P \rightarrow C}(P(I)), T_{P \rightarrow C}(P(\hat{I}))) \quad (9)$$

$$= \mathcal{L}_{style}(P_C(I), P_C(\hat{I})) \quad (10)$$

with  $T_{P \rightarrow C}$  being the color transformation introduced earlier, that maps the color statistics of the projection of the exemplar texture  $P(I) = I_P$  into the color statistics of the palette color image  $I_C$ , and  $P_C = T_C \circ P$  being the new projection operator.

#### E. Multispectral texture synthesis with deep features statistics

To perform multispectral texture synthesis, we initialize a synthetic multispectral image of  $N$  channels with a Gaussian noise of mean and covariance of the exemplar image as it showed better results than a white noise initialization. Then, we use L-BFGS [17] to update the synthetic multispectral image in order to minimize a style distance. At each iteration of the algorithm, the gradients of one of the aforementioned style distances are used to update the pixels of the  $N$  channels synthetic image. The style distance uses a CNN trained on RGB images to extract deep features statistics, covariance matrices here.

## IV. EXPERIMENTS

In this section, we give a brief description of the dataset we consider, the experimental settings and the metrics used to compare the texture syntheses. Comprehensive comparisons of syntheses using the different style distances that we formerly introduced as objective functions are then presented and discussed.

### A. Dataset

We consider a dataset composed of 130 512×512 Sentinel-2 images of cloud fields. The Sentinel-2 mission is an Earth observation mission from the Copernicus Program that acquires optical imagery at a spatial resolution ranging from 10 m to 60 m over land and coastal waters. It captures images in 12 spectral channels in the visible/near infrared (VNIR) and short wave infrared spectral range (SWIR), including 9 proper observation bands and 3 atmospheric correction bands (near 442, 945 and 1375 nm), for which the atmospheric absorption is high. The last band at 1375 nm present particularly high absorption percentage and is not exploitable. As a consequence, only 11 spectral bands are available.

### B. Experimental settings

We use the different style distances that we introduced in this paper as objective functions to perform texture synthesis on the 11-band images. Additional results over the 9 observation bands are also available in Appendix A.

Experimental choices and hyperparameters for each objective are as follows:

- **Stochastic** : we use a batch size of 10 triplets at each iterations. We show further experiments varying the batch size in Appendix B.
- **PCA** : we use Principal Component Analysis to encode the spectral information contained in 11 bands into 3 dimensions. In our case, these 3 first components are sufficient to explain 99% of the variance (see Figure 3 for eigenvalues and eigenvectors).

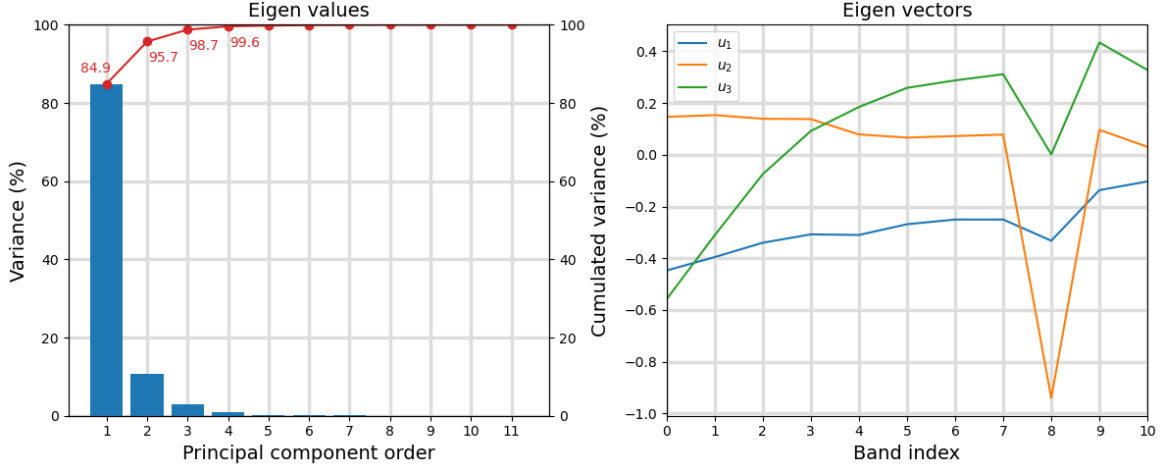


Fig. 3: (Left) Eigenvalues of the used PCA over 11-band Sentinel-2 images. The first component encompasses 95% of the variance, the 3 first 99%. (Right) 3 first eigenvectors of the used PCA over 11-band Sentinel-2 images (denoted as  $u_i$ ).



Fig. 4: Examples of palette images used for the proposed color transfer.

- **PCA + Color** : We use 3 different color exemplar images to perform color transfer upstream of the PCA projection : *Wall*, *Pebbles* and *Fabrics*. The images are available in Figure 4.

For style distances, we use a VGG19 network initialized with the pre-trained weights from [8]. As the authors of the latter, we extract the feature maps of the first convolutional filters at each spatial resolution (i.e. Conv layers with indices 1,3,5,8,11) and use quadratic decaying loss weights  $w_l = \frac{1}{N_l^2}$  where  $N_l$  is the number of feature maps of the layer  $l$ . As mentioned in section III-A, we use the covariance matrices as deep features statistics instead of the Gram matrices and we initialize the synthetic image with a Gaussian noise with the mean and covariance of the exemplar image. We use L-BFGS for 500 iterations to perform optimization.

### C. Metrics

Unlike traditional pixel-by-pixel reconstruction, we want to generate a new texture image that conserves the statistics, the properties and the general visual aspect of the original image while allowing another spatial distribution. Therefore, traditional pixel-wise metrics such as the PSNR are not adapted. In the following, we introduce different metrics used to evaluate the performance of our methods.

1) *Style*: We measure different style distances between original and synthetic images :

- the stochastic style distance denoted as  $\mathcal{L}_{style}^{MS}$ . It is computed by averaging over every possible triplets. This is precisely the metric corresponding to the stochastic style distance.
- the projected style distance using a PCA, denoted as  $\mathcal{L}_{style}^{PCA}$ , corresponding to the second style distance introduced in this paper.
- the style distance for the bands 2,3 and 4 which are the closest to RGB wavelengths. It is denoted as  $\mathcal{L}_{style}^{RGB}$

2) *Multispectral distribution*: In the context of RGB images, color distribution is a central feature in texture description. For multispectral images, we similarly want to synthesize images having a distribution of pixel values as close as possible to the exemplar image. An efficient way to compare multidimensional distribution is the Wasserstein distance. Because it is very costly to compute such distances as soon as we are in dimension greater than two, we either rely on the sliced Wasserstein distance [24] or to band-wise distances:

- sliced Wasserstein distance  $\mathcal{L}_{hist}$ :

$$\begin{aligned} \mathcal{L}_{hist}(I, \hat{I})^2 &= SW_2(I, \hat{I})^2 \\ &= \mathbb{E}_{v \sim \mathcal{U}(\mathbb{S}^{N-1})} \left[ W_2(I \cdot v, \hat{I} \cdot v)^2 \right] \\ &= \mathbb{E}_{v \sim \mathcal{U}(\mathbb{S}^{N-1})} \left[ \left\| \text{sort}(I \cdot v) - \text{sort}(\hat{I} \cdot v) \right\|_2^2 \right], \end{aligned}$$

where  $\mathbb{S}^{N-1}$  is the unit sphere in dimension  $N$ ,  $SW_2$  the 2-sliced Wasserstein distance,  $W_2$  the 2-Wasserstein distance,  $I \cdot v$  the scalar product between  $I$  and  $v$ , and  $\text{sort}$  the sorting operator.

- band-wise Wasserstein distance  $\mathcal{L}_{hist}^\lambda$  for a band  $\lambda$ :

$$\begin{aligned} \mathcal{L}_{hist}^\lambda(I, \hat{I}) &= W_2(I_\lambda, \hat{I}_\lambda) \\ &= \left\| \text{sort}(I_\lambda) - \text{sort}(\hat{I}_\lambda) \right\|_2 \end{aligned}$$

3) *Gaussian hypothesis*: Anomaly detection is one of the most common analyses performed on multispectral images. The historical multispectral anomaly detection criterion, the *RX* [25], makes a Gaussian assumption for the whole image, or at least for limited windows, in order to define a statistical test. Hence, we also measure the difference in the first and second order spectral statistics, using :

- the  $L_2$  norm between the means and covariances, respectively  $\mathcal{L}_\mu$  and  $\mathcal{L}_\Sigma$  :

$$\begin{aligned}\mathcal{L}_\mu(I, \hat{I}) &= \|\mu - \hat{\mu}\|_2 \\ \mathcal{L}_\Sigma(I, \hat{I}) &= \left\| \Sigma - \hat{\Sigma} \right\|_2\end{aligned}$$

- the Wasserstein distance between the two gaussian distributions given by the means and covariances which, for the sake of simplicity, we label *RX* even though it is not equivalent to the *RX* criterion (we do not compute a Mahalanobis distance):

$$\mathcal{L}_{RX}(I, \hat{I}) = \sqrt{(\mu - \hat{\mu})^2 + \text{Tr} \left( \Sigma + \hat{\Sigma} - 2\Sigma^{\frac{1}{2}} \hat{\Sigma} \Sigma^{\frac{1}{2}} \right)} \quad (11)$$

where  $I$  and  $\hat{I}$  have respectively  $(\mu, \Sigma)$  and  $(\hat{\mu}, \hat{\Sigma})$  as statistics.

4) *Spectrum*: Cloud fields images exhibit fractal behaviors, which can be analyzed through the decay of the azimuthal modulus of the Fourier's transform of the image. which is defined as follows :

$$\mathcal{F}_{rad}(I)(r) = \frac{1}{N_r} \int_0^{2\pi} \|\mathcal{F}(I)(r \cos \theta, r \sin \theta)\| d\theta \quad (12)$$

where  $I$  is a single channel image and  $\mathcal{F}$  the Fourier's transform operator. As one expects a power-law decay  $\mathcal{F}_{rad}(r) \sim \frac{1}{r^p}$ , the differences in fractal behaviours is quantified by the  $l_2$  distance between the two logarithm of the azimuthal spectra :

$$\mathcal{L}_{sp}(I, \hat{I}) = \left\| \log \left( \mathcal{F}_{rad}(I) \right) - \log \left( \mathcal{F}_{rad}(\hat{I}) \right) \right\|_2 \quad (13)$$

We compute this distance on the band-wise image  $I_\lambda$  for each spectral band  $\lambda$  and on the grey-scale image  $I_{mean}$  :

$$\mathcal{L}_{sp}^\lambda(I, \hat{I}) = \mathcal{L}_{sp}(I_\lambda, \hat{I}_\lambda) \quad (14)$$

$$\mathcal{L}_{sp}^{mean}(I, \hat{I}) = \mathcal{L}_{sp}(I_{mean}, \hat{I}_{mean}) \quad (15)$$

5) *Gradients*: The final features we consider are image gradients, which are key to several cloud image applications [32] [21] [9]. Gradient calculation in the two spatial directions (horizontal and vertical) on an image  $I$ , leads to two multispectral images  $\nabla_x I$  and  $\nabla_y I$  that can be used together to obtain the magnitude resulting in another multispectral image  $|\nabla I|$ . A proper reconstruction should exhibit a distribution of gradients as close as possible to the original one. Hence, we use the sliced Wasserstein distance to evaluate the deviation between the two distributions :

$$\mathcal{L}_{grad}(I, \hat{I}) = SW_2 \left( |\nabla I|, |\nabla \hat{I}| \right) \quad (16)$$

Method	$\mathcal{L}_{style}^{RGB} \downarrow$	$\mathcal{L}_{sp}^{mean} \downarrow$	$\mathcal{L}_{grad} (\cdot 10^{-3}) \downarrow$	$\mathcal{L}_{hist} (\cdot 10^{-3}) \downarrow$
RGB	<b>0.019</b>	2.03	28.53	14.09
Stochastic	0.099	<b>1.68</b>	<b>9.33</b>	12.41
PCA	85.15	2.09	20.55	13.02
PCA + Wall	55.65	1.90	15.27	11.75
PCA + Pebbles	44.71	1.87	13.46	<b>9.32</b>
PCA + Fabrics	95.68	2.29	27.73	12.35

TABLE I: Evaluation of our methods on RGB bands. The line RGB refers to a baseline where images are obtained by optimizing only on the 3 RGB bands to match deep features statistics. All metrics are computed on the RGB bands.

We further compute the Wasserstein distance over the spectral bands in a similar way :

$$\mathcal{L}_{grad}^\lambda(I, \hat{I}) = W_2(\nabla I_\lambda, \nabla \hat{I}_\lambda) \quad (17)$$

A total of 1000 random directions is used to compute the sliced Wasserstein distances,  $\mathcal{L}_{grad}$  and  $\mathcal{L}_{hist}$ .

#### D. Results

We show qualitative results with reconstruction examples in Figure 5 and quantitative results in Table II. In addition, we show band-wise comparison for distribution, gradients and spectral metrics in Figure 6.

1) *Visual quality* : Our methods enable faithful texture reconstruction for cloud field multispectral images. However, results differ depending on the method used. Indeed, we observe more frequently grid-like artifacts on the images synthesized using a projected style distance (either PCA or PCA + Color), while the ones synthesized using the stochastic style distance do not exhibit this kind of artifacts. The PCA method is the one showing the coarsest artifacts and the use of color transfer significantly enhances the visual aspect of the synthetic images.

2) *Comparison with RGB texture synthesis*: As a sanity check, we compare our method with the result of applying the original method from [8] to the three multispectral bands that are closer to the RGB spectral content. For this, we consider bands 2, 3 and 4. We perform texture synthesis only using these 3 bands. Then, we compare the obtained 3-channels images with the images obtained by keeping only the same 3 bands after optimizing the multispectral images using the aforementioned multispectral style distances. The quantitative comparison on the 3 RGB bands is given in Table I. The considered metrics are computed only on the RGB bands.

Without surprise, the  $\mathcal{L}_{style}^{RGB}$  loss is smaller when optimizing the RGB image. Also, the stochastic style distance performs much better than the projected ones for this metric. Next, one observes that all proposed style distances yield a better or similar preservation of the gradient, the histogram and the spectrum. This may be due to a better account of inter-band dependency when optimizing over all bands.





Fig. 5: Examples of multispectral texture synthesis using different style distances as optimization objective. The images displayed are obtained by pooling the 11 spectral bands to obtain 3-channels images allowing visual representation of multispectral images (channel 1: bands 1,2,3,4; channel 2: bands 5,6,7,8; channel 3: bands 9,11,12). Exemplar textures are shown on the left, and each following column correspond to a previously presented method of synthesis, label are shown on the bottom. Images that optimize the stochastic loss allow a visual appearance more faithful to the original textures than the ones minimizing a projected style distance, with no grid-like artifacts observed in the first ones. The introduction of color transfer upstream of a projection enables far less of these defects in the generated images.

Method	$\mathcal{L}_{style}^{MS} \downarrow$	$\mathcal{L}_{style}^{PCA} \downarrow$	$\mathcal{L}_{sp}^{mean} \downarrow$	$\mathcal{L}_{grad} (\cdot 10^{-3}) \downarrow$	$\mathcal{L}_{hist} (\cdot 10^{-3}) \downarrow$	$\mathcal{L}_{\mu} (\cdot 10^{-3}) \downarrow$	$\mathcal{L}_{\Sigma} (\cdot 10^{-3}) \downarrow$	$\mathcal{L}_{RX} (\cdot 10^{-3}) \downarrow$
Stochastic	<b>0.17</b>	676.84	<b>1.96</b>	<b>13.52</b>	12.63	17.36	<b>14.02</b>	18.62
PCA	115.38	<b>0.59</b>	2.33	26.52	15.02	12.20	21.45	15.68
PCA + Wall	98.38	2158.46	2.35	21.92	14.39	9.24	21.63	12.65
PCA + Pebbles	68.00	4872.01	2.21	18.94	<b>11.70</b>	<b>4.58</b>	21.01	<b>7.87</b>
PCA + Fabrics	102.09	11008.26	2.33	29.78	15.32	12.80	22.16	16.83

TABLE II: Evaluation of multispectral texture synthesis through CNN trained on RGB images. Metrics were computed and averaged over 130 multispectral images.

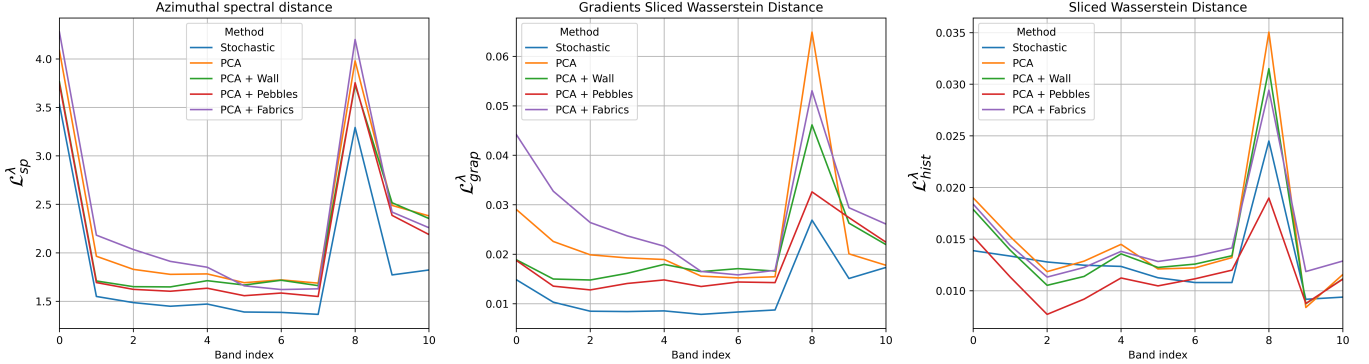


Fig. 6: Band wise comparison of the proposed methods.

3) *Comparison between methods* : We now use all metrics introduced above to compare the three methods that we have introduced to perform neural multispectral texture synthesis: by using the stochastic style distance, by using the projected distance, and finally by using the projected style distance together with color transfer. Of course, each of these methods perform best when using the metric that has been optimized by the method, and we therefore focus on the remaining metrics to reach some conclusions. Results are gathered in Table II.

First, we observe that the stochastic style distance enforces greater respect of the Fourier spectra and of the gradients in the generated textures, two features that are related to the spatial structure of the images. We can see in Figure 6 that this is the case across all spectral bands.

For the multispectral distribution, two methods, namely *Stochastic* and *PCA + Pebbles*, stand out, showing competitive SWD  $\mathcal{L}_{hist}$ . However, these results find different explanations if we look at the different statistics. The projected style distance ensures a better respect of the mean  $\mathcal{L}_{\mu}$  while its stochastic counterpart seems to allow a better respect of the higher order statistics, as seen for the covariance  $\mathcal{L}_{\Sigma}$ . Then, we also observe different dynamics across the spectral bands with the *Stochastic* method outperforming *PCA + Pebbles* for 4 out of 11 spectral bands.

Finally, the band wise analysis clearly reveals that all methods under-perform on the 2 correction bands, at 442nm and 945nm (see Figure 6). This was expected as they present more variability than other bands. However, the two most effective methods are less sensitive to this high variability, showing greater ability to take multispectral information into account.

## V. CONCLUSION

In this paper, we have introduced several strategies to generalize state-of-the-art RGB texture synthesis methods to multispectral images. The main difficulty to do so is to generalize style distances obtained from CNN pre-trained on 3-channels images to a larger number of bands.

The new introduced style distances were used to perform exemplar-based texture synthesis on single multispectral images of cloud fields from the Sentinel-2 mission. They show satisfactory results, both visually and numerically.

The stochastic style distance ensures the best visual reconstruction and globally shows the best numerical results among the tested methods, but comes at a higher computational cost. The use of a projection is faster but less efficient. However, color transfer upstream of the projection happens to largely enhance the quality of the reconstruction.

Finally, all of the proposed methods do not require any additional training, since they leverage an RGB CNN that has proven itself to perform well for texture synthesis and style transfer. The solutions we propose are not specific to texture modeling and can be applied to a wide range of style-like distances that take advantage of the capabilities of pre-trained neural networks. The most direct perspective we can highlight is style transfer. A second perspective is the use of the proposed style distances for feed-forward approaches, to texture synthesis or to more general restoration or editing tasks. Eventually, common distances based on neural networks such as the widely used FID [11] or LPIPS [38] can also be extended to multi/hyperspectral imaging with the strategies proposed in this paper, opening interesting avenues for multispectral image quality evaluation. Another perspective lies in the use of the texture models presented here to train new networks to perform multi-texture synthesis.

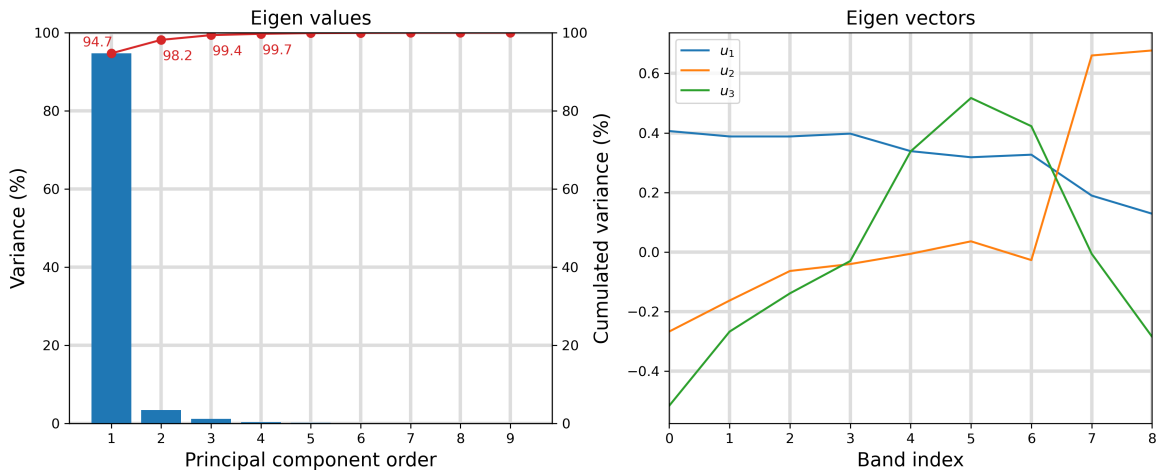


Fig. 7: (Left) Eigenvalues of the used PCA over 9-band Sentinel-2 images. The first component encompasses 95% of the variance, the 3 first 99%. (Right) 3 first eigenvectors of the used PCA over 9-band Sentinel-2 images (denoted as  $u_i$ ).

Method	$\mathcal{L}_{style}^{MS} \downarrow$	$\mathcal{L}_{style}^{PCA} \downarrow$	$\mathcal{L}_{sp}^{mean} \downarrow$	$\mathcal{L}_{grad} (\cdot 10^{-3})$	$\mathcal{L}_{hist} (\cdot 10^{-3}) \downarrow$	$\mathcal{L}_{\mu} (\cdot 10^{-3}) \downarrow$	$\mathcal{L}_{\Sigma} (\cdot 10^{-3}) \downarrow$	$\mathcal{L}_{RX} (\cdot 10^{-3}) \downarrow$
Stochastic	<b>0.09</b>	<b>161.51</b>	<b>1.72</b>	<b>12.46</b>	9.19	14.69	<b>4.93</b>	15.27
PCA + Pebbles	27.92	745.75	1.88	16.30	<b>7.66</b>	<b>3.10</b>	7.08	4.04

TABLE III: Mean principal metrics results for multispectral texture synthesis on 9-band Sentinel-2 images. Sliced distances  $\mathcal{L}_{sp}^{mean}$  and  $\mathcal{L}_{hist}$  were computed using 1000 random directions.

Number of bands	Method	$\mathcal{L}_{style}^{RGB} \downarrow$	$\mathcal{L}_{sp}^{mean} \downarrow$	$\mathcal{L}_{grad} (\cdot 10^{-3}) \downarrow$	$\mathcal{L}_{hist} (\cdot 10^{-3}) \downarrow$
3	RGB	<b>0.019</b>	2.03	3.09	1.56
9	Stochastic	0.059	<b>1.58</b>	<b>9.32</b>	9.75
	PCA + Pebbles	31.15	1.68	11.63	<b>8.25</b>
11	Stochastic	0.099	1.68	9.33	12.41
	PCA + Pebbles	44.71	1.87	13.46	9.32

TABLE IV: Mean principal metrics results computed on RGB bands. The line RGB refers to the benchmark where the 3 bands images were optimized to match deep features statistics

#### APPENDIX A 9 BANDS DATASET

In addition to the results presented for the 11-band Sentinel-2 images, we present in this section the results of texture synthesis for 9-band images. These images correspond to the same images keeping only the proper observation bands and excluding the 2 spectral bands that are usually used for atmospheric correction (see IV-A). We chose to test this new dataset only on the 2 best performing methods, namely *Stochastic* and *PCA + Pebbles*. Figure 7 presents the eigenvectors and eigenvalues of PCA used for the 9-band dataset.

Numerical results are shown in Table III. These results are in line with the previous results on the 11-band images with regard to the hierarchy of the two methods in the various metrics. Moreover, as the two bands that presented the more variability are removed, the synthesis task becomes easier. Indeed, those bands shown themselves to be harder to synthesize for the 11-band images and the results on the 9-band images show a quantitative improvement for both of

Batch size	$\mathcal{L}_{style}^{MS} \downarrow$	$\mathcal{L}_{style}^{PCA} \downarrow$	$\mathcal{L}_{sp}^{mean} \downarrow$	$\mathcal{L}_{grad} (\cdot 10^{-3}) \downarrow$	$\mathcal{L}_{hist} (\cdot 10^{-3}) \downarrow$
1	18.32	4254.55	2.09	22.98	68.11
3	3.00	2435.40	2.05	13.74	60.51
5	0.68	1228.71	2.01	13.65	27.59
7	0.27	726.56	1.98	13.62	15.60
10	<b>0.17</b>	<b>676.84</b>	<b>1.96</b>	<b>13.52</b>	<b>12.63</b>

TABLE V: Mean principal metrics results for multispectral texture synthesis on 11-band Sentinel-2 images when varying the the number of triplets drawn at each iteration

the tested methods. The same goes when we compare the synthesis on the 3 RGB bands with the benchmark (consisting of optimizing the images only on these 3 bands) and the 11-band dataset: a fewer number of bands to consider enhance the quality of the reconstruction for both tested methods.

#### APPENDIX B STOCHASTIC STYLE DISTANCE BATCH SIZE

In this section we study the influence of the batch size (i.e. the number of triplets drawn at each iteration) on the synthesis results when using the stochastic style distance as a minimization objective. We performed exemplar synthesis on the same 130 images from the Sentinel-2 mission and vary the batch size.

The numerical results of our experiment are given in Table V and illustrated in Figure 8. We also provide visual examples in Figure 9. They show that the higher the batch size is, the better are the results for all the tested metric. We observe a great interest in increasing the batch size up to at least 5



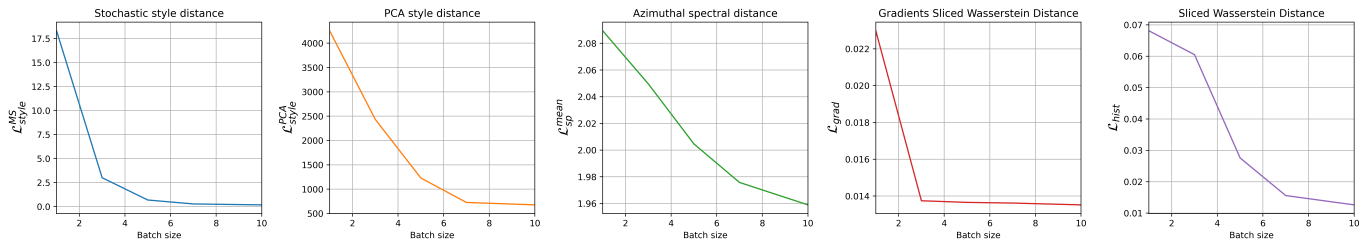


Fig. 8: Influence of the number of triplets drawn at each iteration on texture synthesis quality. Metrics are averaged on 130 images.

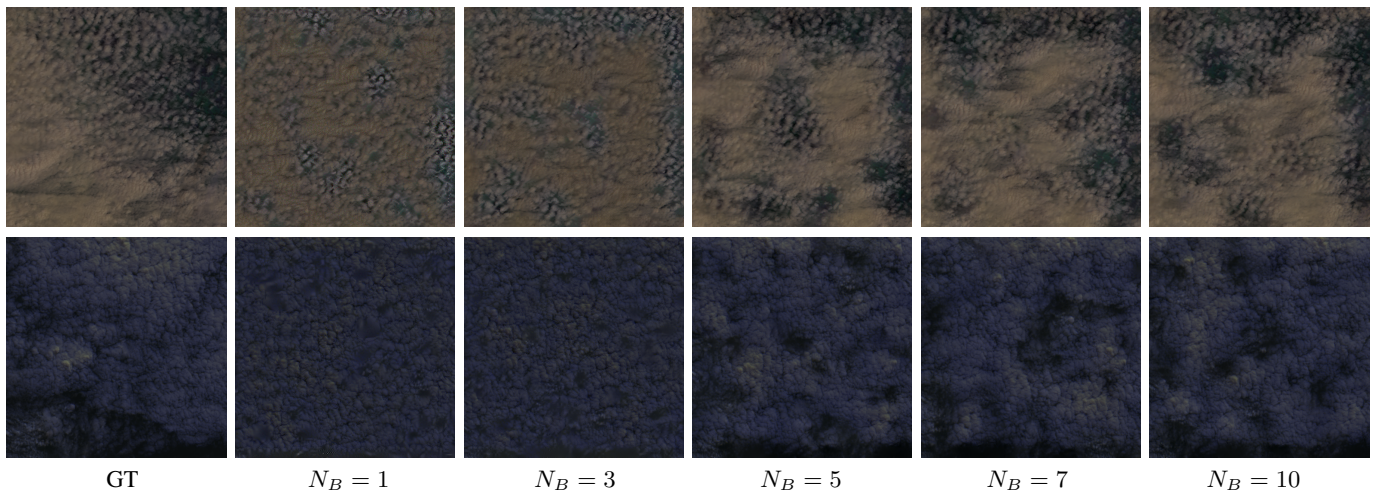


Fig. 9: Examples of multispectral texture reconstruction using stochastic style distance as minimization objective and varying the number of triplets  $N_B$  drawn at each iterations. The images displayed are obtained by pooling the 11 spectral bands to obtain 3-channels (see 5 for details). Exemplar textures are shown on the left, and each following column correspond to a batch size, label are shown on the bottom.

triplets. Indeed, we also observe this threshold in visual quality starting at 5 images for the majority of the tested images.

#### ACKNOWLEDGMENT

This work was supported by the Defence Innovation Agency.

#### REFERENCES

- [1] Michael Alibani, Nicola Acito, and Giovanni Corsini. Multispectral Satellite Image Generation Using StyleGAN3. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17:4379–4391, 2024.
- [2] Pierrick Chatillon, Yann Gousseau, and Sidonie Lefebvre. A Geometrically Aware Auto-Encoder for Multi-texture Synthesis. In Luca Calatroni, Marco Donatelli, Serena Morigi, Marco Prato, and Matteo Santacesaria, editors, *Scale Space and Variational Methods in Computer Vision*, pages 263–275, Cham, 2023. Springer International Publishing.
- [3] Rui Jian Chu, Noel Richard, Hermine Chatoux, Christine Fernandez-Maloigne, and Jon Yngve Hardeberg. Ghost: Gradient Histogram of Spectral Texture. In *2021 11th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pages 1–5, March 2021.
- [4] Rui Jian Chu, Noel Richard, Faouzi Ghorbel, Christine Fernandez-Maloigne, and Jon Yngve Hardeberg. A Metrological Framework For Hyperspectral Texture Analysis Using Relative Spectral Difference Occurrence Matrix. In *2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pages 1–5, September 2019.
- [5] Rui Jian Chu, Noël Richard, Hermine Chatoux, Christine Fernandez-Maloigne, and Jon Yngve Hardeberg. Hyperspectral Texture Metrology Based on Joint Probability of Spectral and Spatial Distribution. *IEEE Transactions on Image Processing*, 30:4341–4356, 2021.
- [6] Yezhen Cong, Samar Khanna, Chenlin Meng, Patrick Liu, Erik Rozi, Yutong He, Marshall Burke, David B. Lobell, and Stefano Ermon. SatMAE: Pre-training Transformers for Temporal and Multi-Spectral Satellite Imagery. October 2022. arXiv 2207.08051.
- [7] Mikolaj Czerkawski, Robert Atkinson, Craig Michie, and Christos Tachtatzis. SatelliteCloudGenerator: Controllable Cloud and Shadow Synthesis for Multi-Spectral Optical Satellite Images. *Remote Sensing*, 15(17):4138, January 2023. Number: 17 Publisher: Multidisciplinary Digital Publishing Institute.
- [8] Leon Gatys, Alexander S Ecker, and Matthias Bethge. Texture Synthesis Using Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, volume 28, 2015.
- [9] Kaiqiang Ge, Jiayin Liu, Feng Wang, Bo Chen, and Yuxin Hu. A Cloud Detection Method Based on Spectral and Gradient Features for SDGSAT-1 Multispectral Images. *Remote Sensing*, 15(1):24, January 2023. Number: 1 Publisher: Multidisciplinary Digital Publishing Institute.
- [10] D.J. Heeger and J.R. Bergen. Pyramid-based texture analysis/synthesis. In *Proceedings., International Conference on Image Processing*, volume 3, pages 648–651 vol.3, October 1995.
- [11] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.
- [12] Nikolay Jetchev, Urs Bergmann, and Roland Vollgraf. Texture Synthesis with Spatial Generative Adversarial Networks, September 2017. arXiv:1611.08207 [cs, stat].
- [13] Haris Ahmad Khan, Sofiane Mihoubi, Benjamin Mathon, Jean-Baptiste



- Thomas, and Jon Yngve Hardeberg. HyTexiLa: High Resolution Visible and Near Infrared Hyperspectral Texture Images. *Sensors*, 18(7):2045, July 2018. Number: 7 Publisher: Multidisciplinary Digital Publishing Institute.
- [14] Samar Khanna, Patrick Liu, Linqi Zhou, Chenlin Meng, Robin Rombach, Marshall Burke, David Lobell, and Stefano Ermon. Diffusion-Sat: A Generative Foundation Model for Satellite Imagery. In *The Twelfth International Conference on Learning Representations*, 2024. arXiv:2312.03606.
- [15] Yin Li and Da Huang. Generating Hyperspectral Data Based on 3D CNN and Improved Wasserstein Generative Adversarial Network Using Homemade High-resolution Datasets. August 2020.
- [16] Jue Lin, Gaurav Sharma, and Thrasylvoulos N. Pappas. Towards Universal Texture Synthesis by Combining Texton Broadcasting with Noise Injection in StyleGAN-2, March 2022. arXiv:2203.04221 [cs].
- [17] Dong C. Liu and Jorge Nocedal. On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, 45(1):503–528, August 1989.
- [18] Liqin Liu, Bowen Chen, Hao Chen, Zhengxia Zou, and Zhenwei Shi. Diverse Hyperspectral Remote Sensing Image Synthesis With Diffusion Models. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–16, 2023.
- [19] G. Mercier and M. Lennon. On the characterization of hyperspectral texture. volume 5, pages 2584–2586 vol.5, February 2002.
- [20] Tharun Mohandoss, Aditya Kulkarni, Daniel Northrup, Ernest Mwebaze, and Hamed Alemohammad. Generating Synthetic Multispectral Satellite Imagery from Sentinel-2, December 2020. arXiv:2012.03108 [cs, eess].
- [21] Saed Moradi, Payman Moallem, and Mohamad Farzan Sabahi. A false-alarm aware methodology to develop robust and efficient multi-scale infrared small target detection algorithm. *Infrared Physics & Technology*, 89:387–397, March 2018.
- [22] Alice Porebski, Mohamed Alimoussa, and Nicolas Vandembroucke. Comparison of color imaging vs. hyperspectral imaging for texture classification. *Pattern Recognition Letters*, 161:115–121, September 2022.
- [23] Javier Portilla and Eero P. Simoncelli. A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients. *International Journal of Computer Vision*, 40(1):49–70, October 2000.
- [24] Julien Rabin, Gabriel Peyré, Julie Delon, and Marc Berton. Wasserstein Barycenter and Its Application to Texture Mixing. In Alfred M. Bruckstein, Bart M. ter Haar Romeny, Alexander M. Bronstein, and Michael M. Bronstein, editors, *Scale Space and Variational Methods in Computer Vision*, pages 435–446, Berlin, Heidelberg, 2012. Springer.
- [25] I.S. Reed and X. Yu. Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(10):1760–1770, October 1990.
- [26] G. Rellier, X. Descombes, F. Falzon, and J. Zerubia. Texture feature analysis using a gauss-Markov model in hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 42(7):1543–1551, July 2004.
- [27] Guillaume Rellier, Xavier Descombes, Josiane Zerubia, and Frédéric Falzon. A Gauss-Markov Model for Hyperspectral Texture Analysis of Urban Areas. In *Proceedings of the 16th International Conference on Pattern Recognition (ICPR'02) Volume 1*, ICPR '02, page 10692, USA, 2002.
- [28] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-Resolution Image Synthesis with Latent Diffusion Models. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10674–10685. IEEE Computer Society.
- [29] Subhadip Sarkar and Glenn Healey. Hyperspectral Texture Synthesis Using Histogram and Power Spectral Density Matching. *IEEE Transactions on Geoscience and Remote Sensing*, 48(5):2261–2270, May 2010. Conference Name: IEEE Transactions on Geoscience and Remote Sensing.
- [30] S.M. Schweizer and J.M.F. Moura. Hyperspectral imagery: clutter adaptation in anomaly detection. *IEEE Transactions on Information Theory*, 46(5):1855–1871, August 2000.
- [31] S.M. Schweizer and J.M.F. Moura. Efficient detection in hyperspectral imagery. *IEEE Transactions on Image Processing*, 10(5):584–597, 2001.
- [32] F. Szczap, Y. Gour, T. Fauchez, C. Cornet, T. Faure, O. Jourdan, G. Penide, and P. Dubuisson. A flexible three-dimensional stratocumulus, cumulus and cirrus cloud generator (3DCLOUD) based on drastically simplified atmospheric equations and the Fourier transform framework. *Geoscientific Model Development*, 7(4):1779–1801, August 2014. Publisher: Copernicus GmbH.
- [33] Dmitry Ulyanov, Vadim Lebedev, Andrea Vedaldi, and Victor Lempitsky. Texture networks: feed-forward synthesis of textures and stylized images. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*, ICML'16, pages 1349–1357, New York, NY, USA, 2016. JMLR.org.
- [34] Ivan Ustyuzhaninov, Wieland Brendel, Leon A. Gatys, and Matthias Bethge. Texture Synthesis Using Shallow Convolutional Networks with Random Filters, May 2016. arXiv:1606.00021 [cs].
- [35] Yi Wang, Hugo Hernández Hernández, Conrad M. Albrecht, and Xiao Xiang Zhu. Feature Guided Masked Autoencoder for Self-supervised Learning in Remote Sensing, October 2023. arXiv:2310.18653.
- [36] Zhitong Xiong, Yi Wang, Fahong Zhang, Adam J. Stewart, Joëlle Hanna, Damian Borth, Ioannis Papoutsis, Bertrand Le Saux, Gustau Camps-Valls, and Xiao Xiang Zhu. Neural Plasticity-Inspired Foundation Model for Observing the Earth Crossing Modalities, March 2024. arXiv:2403.15356.
- [37] Yang Yu, Erting Pan, Yong Ma, Xiaoguang Mei, Qihai Chen, and Jiayi Ma. UnmixDiff: Unmixing-Based Diffusion Model for Hyperspectral Image Synthesis. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–18, 2024. Conference Name: IEEE Transactions on Geoscience and Remote Sensing.
- [38] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. pages 586–595, 2018.



**Sélim Olivier** received the M.Eng in mathematics and data science and the M.Sc in fundamental and applied mathematics from CentraleSupélec, Gif-sur-Yvette, France in 2023. He is currently pursuing the Ph.D degree with the LTCI, Télécom Paris, Palaiseau, France and with the DOTA/AI Lab, ONERA, Palaiseau, France, under the supervision of Yann Gousseau and Sidonie Lefebvre. He is working on the development of deep learning models for texture synthesis applied to multi/hyperspectral and 3D imaging.



**Yann Gousseau** Yann Gousseau received the engineering degree from the École Centrale de Paris, France, the Part III of the Mathematical Tripos from the University of Cambridge, UK, and the Ph.D. degree in Applied Mathematics from the University of Paris-Dauphine. He is currently a professor at Telecom Paris, IP Paris. His research interests include the mathematical modeling of natural images and textures, generative models, computer vision, image and video processing.



**Sidonie Lefebvre** received the Ph.D. degree from Ecole Centrale Paris in 2006. She is currently working as a Research Scientist in statistics at ONERA Palaiseau (France), in the Optics and Associated Techniques Department. Her research interests are centered on the design and modeling of computer experiments, uncertainty quantification and sensitivity analysis, and target detection.