



HAL
open science

Macrocanonical Models for Texture Synthesis

Valentin de Bortoli, Agnès Desolneux, Bruno Galerne, Arthur Leclaire

► **To cite this version:**

Valentin de Bortoli, Agnès Desolneux, Bruno Galerne, Arthur Leclaire. Macrocanonical Models for Texture Synthesis. Scale Space and Variational Methods in Computer Vision. SSVM 2019, Jun 2019, Hofgeismar, Germany. 10.1007/978-3-030-22368-7_2 . hal-02093364

HAL Id: hal-02093364

<https://hal.science/hal-02093364>

Submitted on 8 Apr 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.

Macrocanonical Models for Texture Synthesis

Valentin De Bortoli¹, Agnès Desolneux¹, Bruno Galerne², and Arthur Leclaire³

¹ Centre de mathématiques et de leurs applications, CNRS, ENS Paris-Saclay, Université Paris-Saclay, 94235, Cachan cedex, France.

² Institut Denis Poisson, Université d'Orléans, Université de Tours, CNRS

³ Univ. Bordeaux, IMB, Bordeaux INP, CNRS, UMR 5251, F-33400 Talence, France

Abstract. In this article we consider macrocanonical models for texture synthesis. In these models samples are generated given an input texture image and a set of features which should be matched in expectation. It is known that if the images are quantized, macrocanonical models are given by Gibbs measures, using the maximum entropy principle. We study conditions under which this result extends to real-valued images. If these conditions hold, finding a macrocanonical model amounts to minimizing a convex function and sampling from an associated Gibbs measure. We analyze an algorithm which alternates between sampling and minimizing. We present experiments with neural network features and study the drawbacks and advantages of using this sampling scheme.

Keywords: Texture synthesis, Gibbs measure, Monte Carlo methods, Langevin algorithms, Neural networks

1 Introduction

In image processing a texture can be defined as an image which contains repetitive patterns but also randomness in the pattern placement or in the pattern itself. This vague and unformal definition covers a large class of images such as the ones of terrain, plants, minerals, fur and skin. Exemplar-based texture synthesis aims at synthesizing new images of arbitrary size which have the same perceptual characteristics as a given input texture. It is a challenging task to give a mathematical framework which is not too restrictive, thus describing many texture images, and not too broad, so that computations are numerically feasible. In the literature two classes of exemplar-based texture synthesis algorithms have been considered: the parametric and the non-parametric texture algorithms. Non-parametric texture methods do not rely on an explicit image model in order to produce outputs. For instance copy-paste algorithms such as [6] fill the output image with sub-images from the input. Another example is given by [8] in which the authors apply optimal transport tools in a multiscale patch space.

In this work we focus on parametric exemplar-based texture synthesis algorithms. In contrast to the non-parametric approach they provide an explicit

image model. Output textures are produced by sampling from this image model. In order to derive such a model perceptual features have to be carefully selected along with a corresponding sampling algorithm. There have been huge progress in both directions during the last twenty years.

First, it should be noted that textures which do not exhibit long-range correlations and are well described by their first and second-order statistics can be modeled with Gaussian random fields [25], [9]. These models can be understood as maximum entropy distributions given a mean and a covariance matrix. Their simplicity allows for fast sampling as well as good mathematical understanding of the model. However, this simplicity also restricts the class of textures which can be described. Indeed, given more structured inputs, these algorithms do not yield satisfactory visual results. It was already noted by Gagalowicz [7] that first and second-order statistics are not sufficient to synthesize real-world textures images. In [3] the authors remark that multiscale features capture perceptual characteristics. Following this idea, algorithms based on steerable pyramids [13], wavelet coefficients [20] or wavelet coefficients combined with geometrical properties [19] provide good visual results for a large class of textures. Using Convolutional Neural Networks (CNN), and especially the VGG model [22], Gatys et al. in [10] obtain striking visual results using Gram matrices computed on the layers of the neural network. All these models are called microcanonical textures according to Bruna and Mallat [2], in the sense that they approximately match statistical constraints almost surely (a.s.). Indeed, the previously introduced algorithms start from a noisy input containing all the randomness of the process, then use a (deterministic) gradient descent (or any other optimization algorithm) in order to fit constraints.

On the other hand, models relying on constraints in expectation have been considered in [26]. They correspond to macrocanonical textures according to [2]. They have the advantage to be described by exponential distributions and thus, since their distribution can be made explicit up to some parameters, standard statistical tools can be used for mathematical analysis. However, as noted in [2] they often rely on Monte Carlo algorithms which can be slow to converge. Zhu et al. [26] consider a bank of linear and non-linear filters in order to build an exponential model. Texture images are supposed to be quantized and a Gibbs sampler on each pixel is used in order to update the image. In [17] the authors propose to use first-order statistics computed on CNN outputs. They also suggest to use a Langevin algorithm in order to update the whole image at each iteration. It has also been remarked in [23] that specific Generative Adversarial Networks (GAN) [15] which produce satisfying outputs from a perceptual point of view but lack mathematical understanding can be embedded in an expectation constraint model using the Maximum Mean Discrepancy principle [12].

Our contribution is both theoretical and experimental. After recalling the definition of microcanonical models in Section 2.1 we give precise conditions under which macrocanonical models, *i.e.* maximum entropy models, can be written as exponential distributions in Section 2.2. In Section 2.3, we examine how these conditions translate into a neural network model. Assuming that the maximum

entropy principle is satisfied we then turn to the search of the parameters in such a model. The algorithm we consider, which was already introduced without theoretical proof of convergence in [17], relies on the combination of a gradient descent dynamic, see Section 3.1, and a discretized Langevin dynamic, see Section 3.2. Using new results on these Stochastic Optimization with Unadjusted Kernel (SOUK) algorithms [4] convergence results hold for the algorithm introduced in [17], see Section 3.3. We then provide experiments and after assessing the empirical convergence of our algorithm in Section 4.1 we investigate our choice of models in Section 4.2. We draw the conclusions and limitations of our work in Section 5.

2 Maximum entropy models

2.1 Microcanonical models

Let x_0 be a given input texture. For ease of exposition we consider that $x_0 \in \mathbb{R}^d$, with $d \in \mathbb{N}$, but our results extend to images and color images. We aim at sampling x from a probability distribution satisfying $f(x) \approx f(x_0)$, where $f : \mathbb{R}^d \rightarrow \mathbb{R}^p$ are some statistics computed over the images. However if such a probability distribution exists it is not necessarily unique. In order for the problem to be well-posed we introduce a reference function $J : \mathbb{R}^d \rightarrow (0, +\infty)$ such that $\int_{\mathbb{R}^d} J(x) d\lambda(x) < +\infty$ and we associate to J a probability distribution Π_J such that $\Pi_J(A) = Z_J^{-1} \int_A J(x) d\lambda(x)$ with $Z_J = \int_{\mathbb{R}^d} J(y) d\lambda(y)$ for any $A \in \mathcal{B}(\mathbb{R}^d)$, the Borel sets of \mathbb{R}^d . Let \mathcal{P} be the set of probability distributions over $\mathcal{B}(\mathbb{R}^d)$. If $\Pi \in \mathcal{P}$ is absolutely continuous with respect to the Lebesgue measure λ we denote by $\frac{d\Pi}{d\lambda}$ the probability density function of Π . We introduce the J -entropy, see [14], $H_J : \mathcal{P} \rightarrow [-\infty, +\infty)$ such that for any $\Pi \in \mathcal{P}$

$$H_J(\Pi) = \begin{cases} -\int_{\mathbb{R}^d} \log \left[\frac{d\Pi}{d\lambda}(x) J(x)^{-1} \right] \frac{d\Pi}{d\lambda}(x) d\lambda(x) & \text{if } \frac{d\Pi}{d\lambda} \text{ exists ;} \\ -\infty & \text{otherwise.} \end{cases}$$

The quantity H_J is closely related to the Kullback-Leibler divergence between Π and Π_J . We recall that, if Π is absolutely continuous with respect to λ , we have $\text{KL}(\Pi|\Pi_J) = \int_{\mathbb{R}^d} \log \left[\frac{d\Pi}{d\lambda}(x) \frac{d\Pi_J}{d\lambda}(x)^{-1} \right] \frac{d\Pi}{d\lambda}(x) d\lambda(x)$, and $+\infty$ otherwise. Since $\frac{d\Pi_J}{d\lambda}(x) = Z_J^{-1} J(x)$ we obtain that for any $\Pi \in \mathcal{P}$, $H_J(\Pi) = -\text{KL}(\Pi|\Pi_J) + \log(Z_J)$. The following definition gives a texture model for which statistical constraints are met a.s.

Definition 1. *The probability distribution function $\tilde{\Pi} \in \mathcal{P}$ is a microcanonical model associated with the exemplar texture $x_0 \in \mathbb{R}^d$, statistics $f : \mathbb{R}^d \rightarrow \mathbb{R}^p$ and reference J if*

$$H_J(\tilde{\Pi}) = \max \{ H_J(\Pi), \Pi \in \mathcal{P}, f(X) = f(x_0) \text{ a.s. if } X \sim \Pi \} . \quad (1)$$

Most algorithms which aim at finding a microcanonical model apply a gradient descent algorithm on the function $x \mapsto \|f(x) - f(x_0)\|^2$ starting from an initial

white noise. The intuition behind this optimization procedure is that the entropy information is contained in the initialization and the constraints are met asymptotically. There exists few theoretical work on the subject with the remarkable exception of [2] in which the authors prove that under technical assumptions the limit distribution has its support on the set of constrained images, *i.e.* the constraints are met asymptotically, and provide a lower bound on its entropy.

2.2 Macrocanonical models

Instead of considering a.s. constraints as in (1) we can consider statistical constraints in expectation. This model was introduced by Jaynes in [14] and formalized in the context of image processing by Zhu et al. in [26].

Definition 2. *The probability distribution function $\tilde{\Pi} \in \mathcal{P}$ is a macrocanonical model associated with the exemplar texture $x_0 \in \mathbb{R}^d$, statistics $f : \mathbb{R}^d \rightarrow \mathbb{R}^p$ and reference J if*

$$H_J(\tilde{\Pi}) = \max \{H_J(\Pi), \Pi \in \mathcal{P}, \Pi(f) = f(x_0)\}, \quad (2)$$

where $\Pi(f) = \mathbb{E}_\Pi(f)$.

Macrocanonical models can be seen as a relaxation of the microcanonical ones. A link between macrocanonical models and microcanonical models is highlighted by Bruna and Mallat in [2]. They show that for some statistics, macrocanonical and microcanonical models have the same limit when the size of the image goes to infinity. This transition of paradigm has important consequences from a statistical point of view. First, the constraints in (2) require only the knowledge of the expectation of f under a probability distribution Π . Secondly, in Theorem 1 we will show that the macrocanonical model can be written as a Gibbs measure, *i.e.* $\frac{d\tilde{\Pi}}{d\lambda}(x) \propto \exp(-\langle \tilde{\theta}, f(x) - f(x_0) \rangle) J(x)$ for some $\tilde{\theta} \in \mathbb{R}^p$. Given $\theta \in \mathbb{R}^p$, when it is defined we denote by Π_θ the probability distribution defined by

$$Z(\theta) = \int_{\mathbb{R}^d} e^{-\langle \theta, f(x) - f(x_0) \rangle} J(x) d\lambda(x) \quad \text{and} \quad \frac{d\Pi_\theta}{d\lambda}(x) = \frac{e^{-\langle \theta, f(x) - f(x_0) \rangle}}{Z(\theta)} J(x).$$

Theorem 1 (Maximum entropy principle). *Assume that for any $\theta \in \mathbb{R}^p$ we have*

$$\int_{\mathbb{R}^d} e^{\|\theta\| \|f(x)\|} J(x) d\lambda(x) < +\infty \quad \text{and} \quad \lambda(\{x \in \mathbb{R}^d, \langle \theta, f(x) \rangle < \langle \theta, f(x_0) \rangle\}) > 0. \quad (3)$$

Then there exists $\tilde{\theta} \in \mathbb{R}^p$ such that $\Pi_{\tilde{\theta}}$ is a macrocanonical model associated with the exemplar texture $x_0 \in \mathbb{R}^d$, statistics f and reference J . In addition, we have

$$\tilde{\theta} \in \arg \min \left\{ \log \left[\int_{\mathbb{R}^d} \exp(-\langle \theta, f(x) - f(x_0) \rangle) J(x) d\lambda(x) \right], \theta \in \mathbb{R}^p \right\}. \quad (4)$$

Proof. Without loss of generality we assume that $f(x_0) = 0$. First we show that there exists $\tilde{\theta} \in \mathbb{R}^p$ such that $\Pi_{\tilde{\theta}}$ is well-defined and $\Pi_{\tilde{\theta}}(f) = f(x_0)$. The first condition in (3) implies that $Z(\theta) = \int_{\mathbb{R}^d} \exp(-\langle \theta, f(x) \rangle) J(x) d\lambda(x)$ is defined for all $\theta \in \mathbb{R}^p$. Let $\theta_0 \in \mathbb{R}^p$ we have for any $\theta \in B(\theta_0, 1)$, the unit ball centered on θ_0 , and $i \in \{1, \dots, p\}$, using that for any $t \in \mathbb{R}$, $t \leq e^t$ and the Cauchy-Schwarz inequality,

$$\begin{aligned} \int_{\mathbb{R}^d} \left| \frac{\partial}{\partial \theta_i} [\exp(-\langle \theta, f(x) \rangle) J(x)] \right| d\lambda(x) &\leq \int_{\mathbb{R}^d} \|f(x)\| \exp(-\langle \theta, f(x) \rangle) J(x) d\lambda(x) \\ &\leq \int_{\mathbb{R}^d} \exp((\|\theta_0\| + 2)\|f(x)\|) J(x) d\lambda(x) < +\infty. \end{aligned}$$

Therefore $\theta \mapsto \log(Z)(\theta)$ is differentiable and we obtain that for any $\theta \in \mathbb{R}^p$, $\nabla \log(Z)(\theta) = -\mathbb{E}_{\Pi_\theta}(f) = -\Pi_\theta(f)$. In a similar fashion we obtain that $\log(Z) \in C^2(\mathbb{R}^p, \mathbb{R})$ and we have $\frac{\partial^2 \log(Z)}{\partial \theta_i \partial \theta_j}(\theta) = \Pi_\theta(f_i f_j) - \Pi_\theta(f_i) \Pi_\theta(f_j)$. The Hessian of $\log(Z)$ evaluated at θ is the covariance matrix of $f(X)$ where $X \sim \Pi_\theta$ and thus is non-negative which implies that $\log(Z)$ is convex. We also have for any $\theta \in \mathbb{R}^p$ and $t > 0$

$$\begin{aligned} \log(Z)(t\theta) &= \log \left[\int_{\mathbb{R}^d} \exp(-t\langle \theta, f(x) \rangle) J(x) d\lambda(x) \right] \\ &\geq \log \left[\int_{\langle \theta, f(x) \rangle < 0} \exp(-t\langle \theta, f(x) \rangle) J(x) d\lambda(x) \right] \xrightarrow{t \rightarrow +\infty} +\infty, \end{aligned} \quad (5)$$

where we use the first condition in (3) and the monotone convergence theorem. Therefore $\log(Z)$ is coercive along each direction of \mathbb{R}^p . Let us show that $\log(Z)$ is coercive, *i.e.* for any $M > 0$, there exists $R > 0$ such that for all $\|\theta\| \geq R$, $\log(Z)(\theta) \geq M$. Suppose that $\log(Z)$ is not coercive then there exists a sequence $(\theta_n)_{n \in \mathbb{N}}$ such that $\lim_{n \rightarrow +\infty} \|\theta_n\| = +\infty$ and $\log(Z)(\theta_n)_{n \in \mathbb{N}}$ is upper-bounded by some constant $M \geq 0$. We can suppose that $\theta_n \neq 0$. Upon extracting a subsequence we assume that $(\theta_n / \|\theta_n\|)_{n \in \mathbb{N}}$ admits some limit $\theta^* \in \mathbb{R}^p$ with $\|\theta^*\| = 1$. Let $M_0 = \max[M, \log(Z)(0)]$, we have the following inequality for all $t > 0$

$$\log(Z)(t\theta^*) \leq \inf_{n \in \mathbb{N}} [|\log(Z)(t\theta^*) - \log(Z)(t\theta_n / \|\theta_n\|)| + \log(Z)(t\theta_n / \|\theta_n\|)] \leq M_0,$$

where we used the continuity of $\log(Z)$ and the fact that for n large enough $t < \|\theta_n\|$ and therefore by convexity $\log(Z)(t\theta_n / \|\theta_n\|) \leq t / \|\theta_n\| \log(Z)(0) + (1 - t / \|\theta_n\|) \log(Z)(\theta_n) \leq M_0$. Hence for all $t > 0$, $\log(Z)(t\theta^*)$ is bounded which is in contradiction with (5). We obtain that $\log(Z)$ is continuous, convex, coercive and defined over \mathbb{R}^p . This ensures us that there exists $\tilde{\theta}$ such that $\log(Z)(\tilde{\theta})$ is minimal and therefore $\nabla_\theta \log(Z)(\tilde{\theta}) = -\Pi_{\tilde{\theta}}(f) = 0$. Note that we have

$$H_J(\Pi_{\tilde{\theta}}) = \int_{\mathbb{R}^d} \langle \tilde{\theta}, f(x) \rangle \frac{d\Pi_{\tilde{\theta}}}{d\lambda}(x) d\lambda(x) + \log(Z)(\tilde{\theta}) = \log(Z)(\tilde{\theta}).$$

Now let $\Pi \in \mathcal{P}$ such that $\Pi(f) = 0$. If Π is not absolutely continuous with respect to the Lebesgue measure, then $H_J(\Pi) = -\infty < H_J(\Pi_{\tilde{\theta}})$. Otherwise if Π is absolutely continuous with respect to the Lebesgue measure we have the following inequality

$$\begin{aligned} H_J(\Pi) &= - \int_{\mathbb{R}^d} \log \left[\frac{d\Pi}{d\lambda}(x) J(x)^{-1} \right] \frac{d\Pi}{d\lambda}(x) d\lambda(x) \\ &= - \int_{\mathbb{R}^d} \log \left[\frac{d\Pi}{d\lambda}(x) \left(\frac{d\Pi_{\tilde{\theta}}}{d\lambda}(x) \right)^{-1} \right] \frac{d\Pi}{d\lambda}(x) - \log \left[\frac{d\Pi_{\tilde{\theta}}}{d\lambda}(x) J(x)^{-1} \right] \frac{d\Pi}{d\lambda}(x) d\lambda(x) \\ &= -\text{KL}(\Pi | \Pi_{\tilde{\theta}}) + \log(Z)(\tilde{\theta}) \leq \log(Z)(\tilde{\theta}) = H_J(\Pi_{\tilde{\theta}}), \end{aligned}$$

which concludes the proof.

Theorem 1 gives a method for finding the optimal parameters $\theta \in \mathbb{R}^p$ by solving the convex problem (4). We address this issue in Section 3.

2.3 Some feature examples

In the framework of exemplar-based texture synthesis, f is defined as the spatial statistics of some image feature. For instance, let $\mathcal{F} : \mathbb{R}^d \rightarrow \prod_{i=1}^p \mathbb{R}^{d_i}$ be a measurable mapping lifting the image $x \in \mathbb{R}^d$ in a higher-dimensional space $\prod_{i=1}^p \mathbb{R}^{d_i}$. Classical examples include wavelet transforms, power transforms or neural network features. Let $(\mathcal{F}_i)_{i=1,\dots,p}$ such that for any $x \in \mathbb{R}^d$, $\mathcal{F}(x) = (\mathcal{F}_1(x), \dots, \mathcal{F}_p(x))$ and $\mathcal{F}_i : \mathbb{R}^d \rightarrow \mathbb{R}^{d_i}$. Then the statistics f can be defined for any $x \in \mathbb{R}^d$ as follows

$$f(x) = \left(d_1^{-1} \sum_{k=1}^{d_1} \mathcal{F}_1(x)(k), \dots, d_p^{-1} \sum_{k=1}^{d_p} \mathcal{F}_p(x)(k) \right). \quad (6)$$

Note that this formulation includes histograms of bank of filters [20], wavelet coefficients [19] and scattering coefficients [2]. The model defined by such statistics is stationary, *i.e.* translation invariant, since we perform a spatial summation. In the following we focus on first-order features, which will be used in Section 4.1 to assess the convergence of our sampling algorithm, and neural network features, extending the work of [17].

Neural Network features. We denote by $\mathcal{A}_{n_2, n_1}(\mathbb{R})$ the vector space of the affine operators from \mathbb{R}^{n_1} to \mathbb{R}^{n_2} . Let $(A_j)_{j \in \{1, \dots, M\}} \in \prod_{j=1}^M \mathcal{A}_{n_{j+1}, n_j}(\mathbb{R})$, where we let $(n_j)_{j \in \{1, \dots, M+1\}} \in \mathbb{N}^{M+1}$, with $M \in \mathbb{N}$ and $n_1 = d$. Let $\varphi : \mathbb{R} \rightarrow \mathbb{R}$. We define for any $j \in \{1, \dots, M\}$, the j -th layer feature $\mathcal{G}_j : \mathbb{R}^d \rightarrow \mathbb{R}^{n_j}$ for any $x \in \mathbb{R}^d$ by

$$\mathcal{G}_j(x) = (\varphi \circ A_j \circ \varphi \circ A_{j-1} \circ \dots \circ \varphi \circ A_1)(x),$$

where φ is applied on each component of the vectors. Let $p \in \{1, \dots, M\}$ and $(j_i)_{i \in \{1, \dots, p\}} \in \{1, \dots, M\}^p$ then we can define \mathcal{F} as in (6) by

$$f(x) = \left(n_{j_1}^{-1} \sum_{k=1}^{n_{j_1}} \mathcal{G}_{j_1}(x)(k), \dots, n_{j_p}^{-1} \sum_{k=1}^{n_{j_p}} \mathcal{G}_{j_p}(x)(k) \right).$$

Assuming that φ , the non-linear unit, is $\mathcal{C}^1(\mathbb{R})$ we obtain that f is $\mathcal{C}^1(\mathbb{R}^d, \mathbb{R}^p)$. The next proposition gives conditions under which Theorem 1 is satisfied. We denote by df the Jacobian of f .

Proposition 1 (Differentiable neural network maximum entropy). *Let $x_0 \in \mathbb{R}^d$ and assume that $df(x_0)$ has rank $\min(d, p) = p$. In addition, assume that there exists $C \geq 0$ such that for any $x \in \mathbb{R}$, $|\varphi(x)| \leq C(1 + |x|)$. Then the conclusions of Theorem 1 hold for any $J(x) = \exp(-\varepsilon\|x\|^2)$ with $\varepsilon > 0$.*

Proof. The integrability condition is trivially checked since f is sub-linear using that $|\varphi(x)| \leq C(1 + |x|)$. Turning to the proof of the second condition, since $f \in \mathcal{C}^1(\mathbb{R}^d, \mathbb{R}^p)$ and $df(x_0)$ is surjective we can assert the existence of an open set U as well as $\Phi \in \mathcal{C}^1(U, \mathbb{R}^d)$ with $f(x_0) \in U$ such that for any $y \in U$, $f(\Phi(y)) = y$. Now consider $\theta \in \mathbb{R}^p$. If $\theta \in f(x_0)^\perp$ then for $\varepsilon > 0$ small enough $f(x_0) - \varepsilon\theta \in U$ and we obtain that $\langle \theta, f(\Phi(f(x_0) - \varepsilon\theta)) \rangle = -\varepsilon\|\theta\|^2 < 0$. If $\theta \notin f(x_0)^\perp$ then there exists $\varepsilon > 0$ small enough such that $[f(x_0)(1 - \varepsilon), f(x_0)(1 + \varepsilon)] \subset U$. Then for any $\alpha \in (-\varepsilon, \varepsilon)$ we get that $\langle \theta, f(\Phi((1 + \alpha)f(x_0))) \rangle - \langle \theta, f(x_0) \rangle = \alpha \langle \theta, f(x_0) \rangle$. By choosing $\alpha > 0$, respectively $\alpha < 0$, if $\langle \theta, f(x_0) \rangle > 0$, respectively $\langle \theta, f(x_0) \rangle < 0$, we obtain that for any $\theta \in \mathbb{R}^p$, there exists $x \in \mathbb{R}^d$ such that $\langle \theta, f(x) \rangle < \langle \theta, f(x_0) \rangle$. We conclude using the continuity of f .

3 Minimization and sampling algorithm

3.1 Maximizing the entropy

In order to find $\tilde{\theta}$ such that $\Pi_{\tilde{\theta}}$ is the macrocanonical model associated with the exemplar texture x_0 , statistics f and reference J we perform a gradient descent on $\log(Z)$. Let $\theta_0 \in \mathbb{R}^p$ be some initial parameters. We define the sequence $(\theta_n)_{n \in \mathbb{N}}$ for any $n \in \mathbb{N}$ by

$$\theta_{n+1} = P_\Theta [\theta_n - \delta_{n+1} \nabla \log(Z)(\theta_n)] = P_\Theta [\theta_n + \delta_{n+1} (\Pi_{\theta_n}(f) - f(x_0))] , \quad (7)$$

where $(\delta_n)_{n \in \mathbb{N}}$ is a sequence of step sizes with $\delta_n \geq 0$ for any $n \in \mathbb{N}$ and P_Θ is the projection over Θ . The introduction of the projection operator P_Θ is a technical condition in order to guarantee the convergence of the algorithm in Section 3.3. Implementing the algorithm associated to (7) requires the knowledge of the moments of the statistics f for any Gibbs measure Π_θ with $\theta \in \mathbb{R}^p$. The more complex the texture model is the more difficult it is to compute the expectation of the statistics. This expectation can be written as an integral and techniques such as the ones presented in [18] could be used. We choose to approximate this expectation using a Monte Carlo strategy. Assuming that $(X_n)_{n \in \mathbb{N}}$ are samples from Π_θ , we have that $n^{-1} \sum_{k=1}^n f(X_k)$ is an unbiased estimator of $\Pi_\theta(f)$.

3.2 Sampling from Gibbs measures

We now turn to the problem of sampling from Π_θ . Unfortunately, most of the time there is no easy way to produce samples from Π_θ . Nonetheless, using the

ergodicity properties of specific Markov Chains we can still come up with estimators of $\Pi_\theta(f)$. Indeed, if $(X_n)_{n \in \mathbb{N}}$ is a homogeneous Markov chain with kernel K and invariant probability measure Π_θ , *i.e.* $\Pi_\theta K = \Pi_\theta$ we obtain under suitable conditions over f and K that $\lim_{n \rightarrow +\infty} \mathbb{E} \left[n^{-1} \sum_{k=1}^n f(X_k) \right] = \Pi_\theta(f)$. This leads us to consider the following Langevin dynamic for all $n \in \mathbb{N}$

$$X_{n+1} = X_n - \gamma_{n+1} \sum_{i=1}^p \theta_i \nabla f_i(X_n) + \sqrt{2\gamma_{n+1}} Z_{n+1} \quad \text{and} \quad X_0 \in \mathbb{R}^d, \quad (8)$$

where $(Z_n)_{n \in \mathbb{N}^*}$ is a collection of independent d -dimensional zero mean Gaussian random variables with covariance matrix identity and $(\gamma_n)_{n \in \mathbb{N}}$ is a sequence of step sizes with $\gamma_n \geq 0$. The, possibly inhomogeneous, Markov Chain $(X_n)_{n \in \mathbb{N}}$ is associated with the sequence of kernels $(R_{\gamma_n})_{n \in \mathbb{N}}$ with $R_{\gamma_n}(x, \cdot) = \mathcal{N}(x - \gamma_n \sum_{i=1}^p \theta_i \nabla f_i(x), 2\gamma_n \text{Id})$. Note that (8) is the Euler-Maruyama discretization of the continuous dynamic $dX_t = - \sum_{i=1}^p \theta_i \nabla f_i(X_t) dt + \sqrt{2} dB_t$ where $(B_t)_{t \geq 0}$ is a d -dimensional Brownian motion.

3.3 Combining dynamics

We now combine the gradient dynamic and the Langevin dynamic. This algorithm is referred as Stochastic Optimization with Unadjusted Langevin (SOUL) algorithm in [4] and is defined for all $n \in \mathbb{N}$ and $k \in \{0, m_n - 1\}$ by the following recursion

$$X_{k+1}^n = X_k^n - \gamma_{n+1} \sum_{i=1}^p \theta_i \nabla f_i(X_k^n) + \sqrt{2\gamma_{n+1}} Z_{k+1}^n, X_0^n = X_{m_n-1}^{n-1}, n \geq 1, \\ \theta_{n+1} = P_\Theta \left[\theta_n + \delta_{n+1} m_n^{-1} \sum_{k=1}^{m_n} (f(X_k^n) - f(x_0)) \right],$$

with $X_0^0 \in \mathbb{R}^d$, $\theta_0 \in \mathbb{R}^p$, $(\delta_n)_{n \in \mathbb{N}}$, $(\gamma_n)_{n \in \mathbb{N}}$ real positive sequences of step sizes and $(m_n)_{n \in \mathbb{N}} \in \mathbb{N}^{\mathbb{N}}$, the number of Langevin iterations. In [4] the authors study the convergence of these combined dynamics.

4 Experiments

In this section, we present experiments conducted with neural network features. Texture synthesis with these features has been first done in [10] in which the authors compute Gram matrices, *i.e.* second-order information, on different layers of a network. The underlying model is microcanonical. In [17] the authors consider a macrocanonical model with convolutional neural network features corresponding to the mean of filters at a given layer. In our model we consider the features described in Section 2.3 where the linear units and rectifier units are given by the VGG-19 model [22] which contains 16 convolutional layers. We consider the following settings and notations:

- Trained (**T**) or Gaussian (**G**): if option T is selected the weights used in the VGG convolutional units are given by a classical pretraining for the classification task on the ImageNet dataset [5]. If option G is selected we replace the pretrained weights with Gaussian random variables such that the weights of each channel and each layer have same mean and same standard deviation.
- Shallow (**3**), Mid (**6**), Deep (**8**): in our experiments we consider different settings regarding the number of layers and, more importantly, the influence of their depth. In the Shallow (3) setting we consider the linear layers number 3, 4 and 5. In the Mid (6) setting we consider the linear layers number 3, 4, 5, 6, 7 and 11. In the Deep (8) setting we consider the linear layers number 3, 4, 5, 6, 7, 11, 12 and 14.

4.1 Empirical convergence of the sampling algorithm

We assert the experimental convergence of the SOUL algorithm in Figure 1. We choose $\delta_n = \mathcal{O}(n^{-1})$, $\gamma_n = \mathcal{O}(n^{-1})$, $m_n = 1$, $\Theta = \mathbb{R}^p$ and $\varepsilon = 0$, *i.e.* $J = 1$. The algorithm is robust for these fixed parameters for a large number of images. Note that even if this case is not covered by the theoretical results of [4], the convergence is improved using these rates. The drawbacks of not using parameter projection ($\Theta = \mathbb{R}^p$) or image regularization ($\varepsilon = 0$) is that the algorithm may diverge for some images, see Figure 2.

Interestingly, while neural network features capture perceptual details of the texture input they fail to restore low frequency components such as the original color histogram. In order to alleviate this problem, we perform a histogram matching of the output image, as in [11]. In the next section we investigate the advantages of using CNN channel outputs as features.

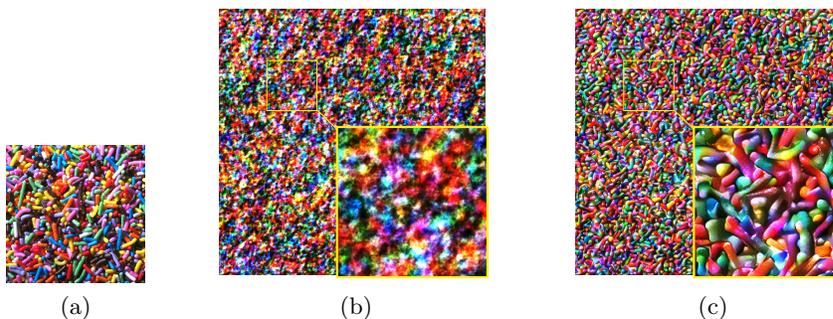


Fig. 1. Empirical convergence. In (a) we present a 512×512 objective texture. In (b) we show the initialization of our algorithm, a 1024×1024 Gaussian random fields with same mean and covariance as a zero-padded version of (a). In (c) we present the result of our algorithm after 5000 iterations with (T-8). In the bottom-right corner of each image (b) and (c) we present a $\times 3$ magnification of some details of the images.

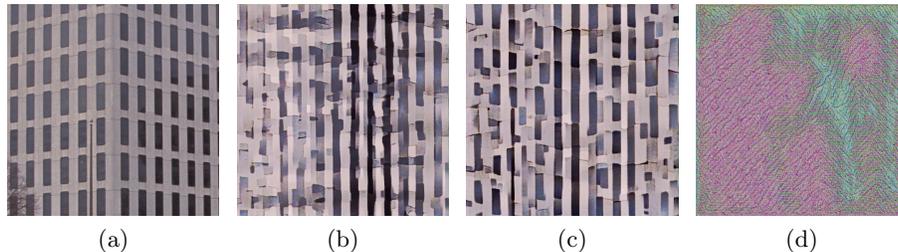


Fig. 2. Depth influence. The image in (a) is the original texture. In (b), (c) and (d) we consider the outputs of the sampling algorithms (T-3), (T-6) and (T-8) algorithms. Note that more geometrical structure is retrieved in (c) than in (b) and that the model has diverged in (d).

4.2 Neural network features

Number of layers. We start by investigating the influence of the number of layers in the model by running the algorithm for different layer configurations. If too many layers are considered the algorithm diverges. However, if the number of layers considered in the model is reduced we observe different behaviors. This is illustrated in Figure 2 where the objective image exhibits strong mid-range structure information.

Model choice. In all previous experiments the weights considered in the CNN architecture are pretrained on a classification task as in [10] and [17]. It is natural to ask if such a pretraining is necessary. In accordance with the results obtained by Gatys et al. [10] we find that a model with no pretraining does not produce perceptually satisfying texture samples, see Figure 3. Note that in [24] the authors obtain good results with random convolutional layers and a microcanonical approach.

4.3 Comparison with state-of-the art methods

To conclude this experimental study we provide a comparison of our results with state-of-art texture synthesis methods in Figure 4. Regarding regular textures our model misses certain geometrical constraints, which are encoded by the Gram matrices in [10] for instance. However, our model relies only on 2k features, using (T-8), whereas Gatys et al. use at least 10k parameters. One way to impose the lost geometrical constraints could be to project the spectrum of the outputs at each step of the algorithm as it was done by Liu et al. [16] in a microcanonical model.

5 Perspectives

There still exists a gap between the theoretical analysis of those algorithms which relies on control theory tools [2], stochastic optimization techniques [1] or general

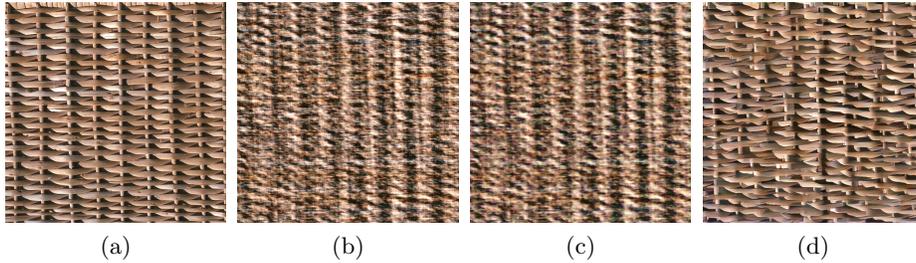


Fig. 3. Noisy weights. In (a) we present the input image, in (b) the initialization of the algorithm and in (c), respectively (d), the output of the algorithm (G-8), respectively (T-8) after 5000 iterations. Note that no spatial structure is retrieved in (c) which is close to its Gaussian initialization.

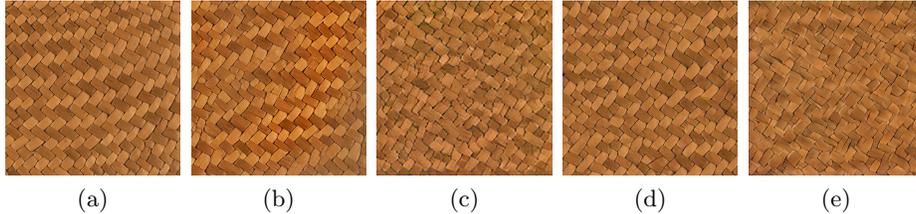


Fig. 4. Comparison. The input image is shown in (a). In (b) we show the output of the algorithm in [10], in (c) the output of the DeepFRAME method [17] and in (d) we present a result obtained with GAN texture synthesis algorithm [15]. Our result (e) after 5000 iterations of (T-8) is comparable but lacks spatial organization. Images (a)–(d) extracted from [21].

state space Markov chain results and the experimental study. Indeed the class of functions which is handled by these theoretical results is constrained (regularity assumptions, drift conditions...) and has yet to be extended to more general CNN features. In addition, they scale badly with the dimension of the data which is high in our image processing context. In a future work we wish to extend our theoretical understanding of SOUL algorithms applied to macrocanonical models and draw parallels with the microcanonical results obtained in [2].

References

1. Atchadé, Y.F., Fort, G., Moulines, E.: On perturbed proximal gradient algorithms. *J. Mach. Learn. Res.* **18**, Paper No. 10, 33 (2017)
2. Bruna, J., Mallat, S.: Multiscale Sparse Microcanonical Models. arXiv e-prints arXiv:1801.02013 (Jan 2018)
3. Cano, D., Minh, T.: Texture synthesis using hierarchical linear transforms. *Signal Processing* **15**(2), 131 – 148 (1988)

4. De Bortoli, V., Durmus, A., Pereyra, M., Fernandez Vidal, A.: Stochastic optimization with unadjusted kernel: the SOUK algorithm. preprint (2019)
5. Deng, J., Dong, W., Socher, R., Li, L., Li, K., Li, F.: Imagenet: A large-scale hierarchical image database. In: CVPR. pp. 248–255 (2009)
6. Efros, A.A., Leung, T.K.: Texture synthesis by non-parametric sampling. In: ICCV. pp. 1033–1038 (1999)
7. Gagalowicz, A., Ma, S.D.: Model driven synthesis of natural textures for 3-d scenes. *Computers & Graphics* **10**(2), 161–170 (1986)
8. Galerne, B., Leclaire, A., Rabin, J.: A texture synthesis model based on semi-discrete optimal transport in patch space. *SIIMS* **11**(4), 2456–2493 (2018)
9. Galerne, B., Gousseau, Y., Morel, J.: Random phase textures: theory and synthesis. *IEEE Trans. Image Processing* **20**(1), 257–267 (2011)
10. Gatys, L.A., Ecker, A.S., Bethge, M.: Texture synthesis using convolutional neural networks. In: NIPS. pp. 262–270 (2015)
11. Gatys, L.A., Ecker, A.S., Bethge, M., Hertzmann, A., Shechtman, E.: Controlling perceptual factors in neural style transfer. In: CVPR. pp. 3730–3738 (2017)
12. Gretton, A., Borgwardt, K.M., Rasch, M.J., Schölkopf, B., Smola, A.J.: A kernel method for the two-sample-problem. In: NIPS. pp. 513–520 (2006)
13. Heeger, D.J., Bergen, J.R.: Pyramid-based texture analysis/synthesis. In: ICIP. pp. 648–651 (1995)
14. Jaynes, E.T.: Information theory and statistical mechanics. *Phys. Rev.* **106**, 620–630 (1957)
15. Jetchev, N., Bergmann, U., Vollgraf, R.: Texture synthesis with spatial generative adversarial networks. *CoRR* (2016)
16. Liu, G., Gousseau, Y., Xia, G.: Texture synthesis through convolutional neural networks and spectrum constraints. In: ICPR. pp. 3234–3239 (2016)
17. Lu, Y., Zhu, S., Wu, Y.N.: Learning FRAME models using CNN filters. In: AAAI. pp. 1902–1910 (2016)
18. Ogden, H.E.: A sequential reduction method for inference in generalized linear mixed models. *Electron. J. Stat.* **9**(1), 135–152 (2015)
19. Peyré, G.: Texture synthesis with grouplets. *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(4), 733–746 (2010)
20. Portilla, J., Simoncelli, E.P.: A parametric texture model based on joint statistics of complex wavelet coefficients. *IJCV* **40**(1), 49–70 (2000)
21. Raad, L., Davy, A., Desolneux, A., Morel, J.: A survey of exemplar-based texture synthesis. *Annals of Mathematical Sciences and Applications* **3**, 89 – 148 (2018)
22. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *CoRR* (2014)
23. Ulyanov, D., Lebedev, V., Vedaldi, A., Lempitsky, V.S.: Texture networks: Feed-forward synthesis of textures and stylized images. In: ICML. pp. 1349–1357 (2016)
24. Ustyuzhaninov, I., Brendel, W., Gatys, L.A., Bethge, M.: Texture synthesis using shallow convolutional networks with random filters. *CoRR* (2016)
25. van Wijk, J.J.: Spot noise texture synthesis for data visualization. In: SIGGRAPH. pp. 309–318 (1991)
26. Zhu, S.C., Wu, Y.N., Mumford, D.: Filters, random fields and maximum entropy (FRAME): towards a unified theory for texture modeling. *IJCV* **27**(2), 107–126 (1998)