LEARNING GENERATIVE MODELS WITH OPTIMAL TRANSPORT

Antoine Houdard, Arthur Leclaire, Nicolas Papadakis, Julien Rabin

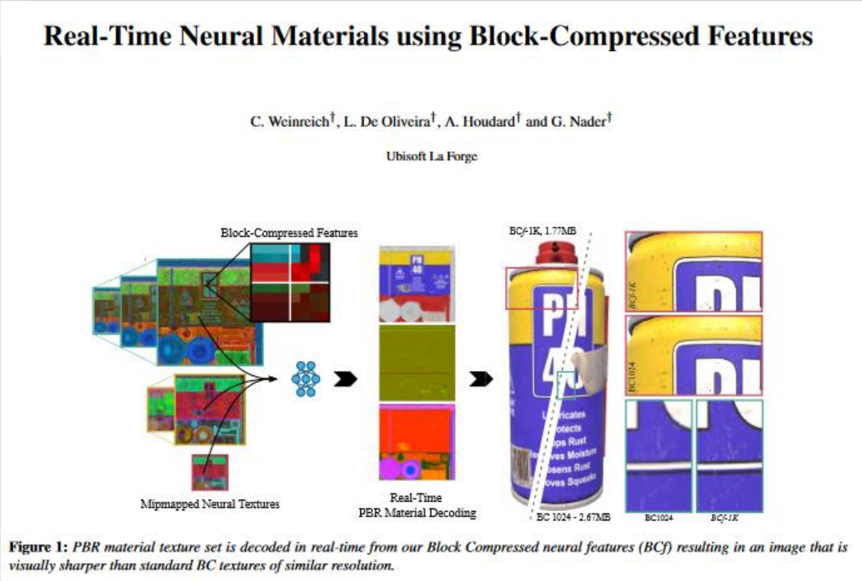
CANUM 2024 Symposium

ABOUT ME

Since 2021 — Research Scientist at Ubisoft

3D rendering, Computer Graphics, Image Processing

New paper about real time compression at EG2024


Real-Time Neural Materials using Block-Compressed Features

C. Weinreich†, L. De Oliveira†, A. Houdard† and G. Nader†

Ubisoft La Forge

Figure 1: *PBR material texture set is decoded in real-time from our Block Compressed neural features (BCf) resulting in an image that is visually sharper than standard BC textures of similar resolution.*

2019-2021 — Post-doc at IMB, Univ. Bordeaux

Optimal Transport, Generative Models, Texture Synthesis

On the Gradient Formula for learning Generative Models with Regularized Optimal Transport Costs

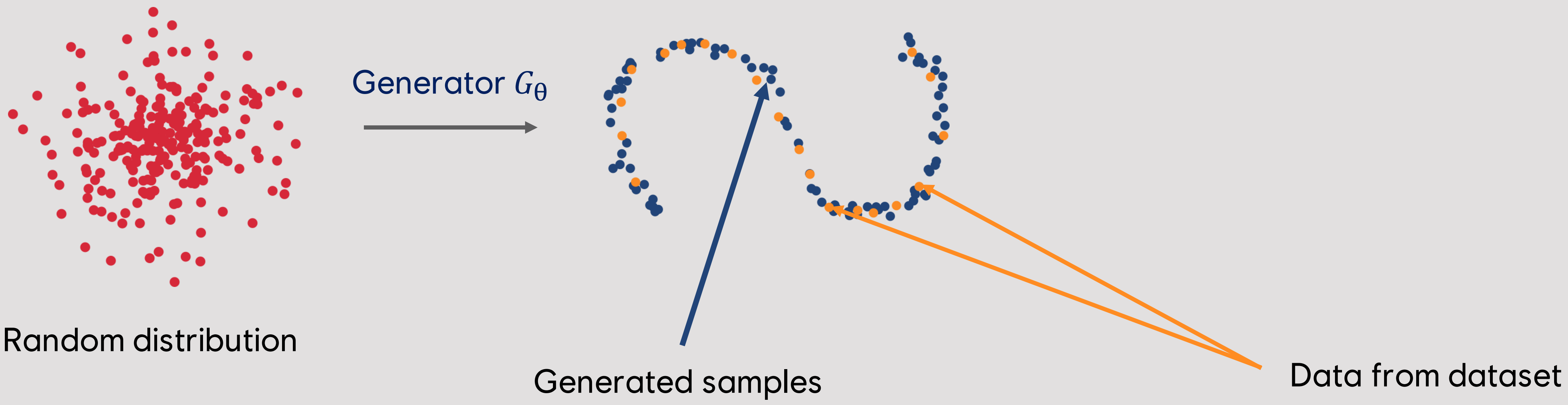
Antoine Houdard, Arthur Leclaire, Nicolas Papadakis, Julien Rabin

Published: 15 Jul 2023, Last Modified: 15 Jul 2023 Accepted by TMLR Everyone Revisions BibTeX

Abstract: Learning a Wasserstein Generative Adversarial Networks (WGAN) requires the differentiation of the optimal transport cost with respect to the parameters of the generative model. In this work, we provide sufficient conditions for the existence of a gradient formula in two different frameworks: the case of semi-discrete optimal transport (i.e. with a discrete target distribution) and the case of regularized optimal transport (i.e. with an entropic penalty). In both cases the gradient formula involves a solution of the semi-dual formulation of the optimal transport cost. Our study makes a connection between the gradient of the WGAN loss function and the Laguerre diagrams associated to semi-discrete transport maps. The learning problem is addressed with an alternating algorithm, which is in general not convergent. However, in most cases, it stabilizes close to a relevant solution for the generative learning problem. We also show that entropic regularization can improve the convergence speed but noticeably changes the shape of the learned generative model.

This talk

GENERATIVE MODELS



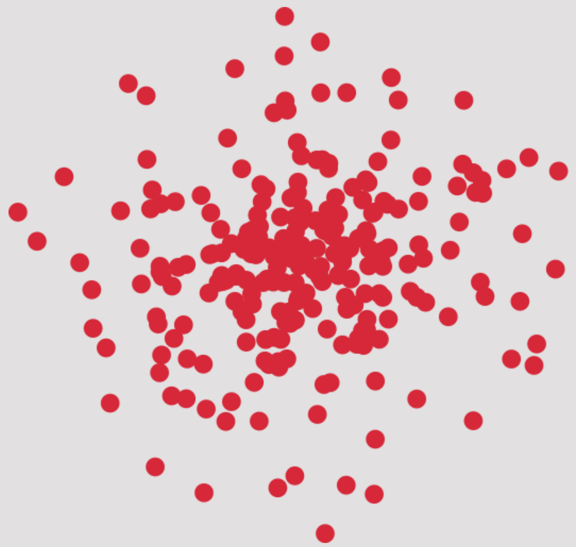
Goal: optimize θ such that generated samples match the dataset distribution

Ex: GANs use a discriminator D_η and try to solve $\min_\theta \max_\eta L(G_\theta, D_\eta)$

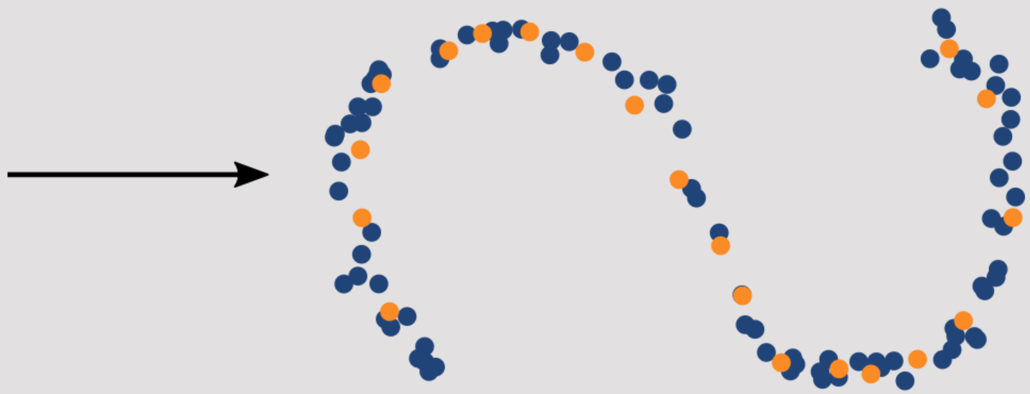
Ex: Diffusion models learn a step-conditioned generator to iteratively generate target distribution

SEEN AS STATISTICAL ESTIMATION

- Data $\{y_1, \ldots, y_n\}$ sampled from $Y \sim \nu$
- Distribution image $\mu_\theta = g_\theta \# \zeta$ defined through a generative model



samples of $Z \sim \zeta$



samples of $g_\theta(Z) \sim g_\theta \# \zeta$

Goal: find an estimate of θ s.t. μ_θ and ν *are close*

?

OPTIMAL TRANSPORT COST

Find an estimate of θ s.t. μ_θ *is close to* **ν** for the **optimal transport** cost

$$\hat{\theta} = \min_{\theta} \mathrm{OT}_c\,(\mu_\theta, \nu)$$

using **semi-duale** formulation

$$\mathrm{OT}_c(\mu_\theta, \nu) = \max_{\psi} E_{X \sim \mu_\theta}[\psi^c(X)] + E_{Y \sim \nu}[\psi(Y)]$$

where $\psi^c(x) = \min_{y}[c(x,y) - \psi(y)]$

with $c(x,y) = |x-y|$, we have $\psi^c = -\psi$ and we get the formulation from **Wassestein GAN**

TITLE	CITED BY	YEAR
Wasserstein gan M Arjovsky, S Chintala, L Bottou	16061 *	2017

How to use **OT cost as a loss**? General case for **any cost**? Get **rid of the neural network** for ψ?

MINIMIZE OPTIMAL TRANSPORT LOSS

- **Goal:** minimize w.r.t. θ

$$W(\theta) = \mathrm{OT}_c(g_\theta \# \zeta, \nu) = \max_{\psi} E_{Z\sim\zeta}[\psi^c(g_\theta(Z))] + E_{Y\sim\nu}[\psi(Y)]$$

- distribution **ν is known (data)**
- one **can sample from $\boldsymbol{\mu_\theta = g_\theta \# \zeta}$ (forward generative model)**

Can we compute a **stochastic gradient of $\boldsymbol{W(\theta)}$?**

Proposition (envelop theorem):

Under some regularity conditions of g_θ and c, if ψ_0^* is an optimal potential for θ_0 then

$$\nabla_\theta W(\theta_0) = \nabla_\theta E_{Z\sim\zeta}\left[\psi_0^{*c}\left(g_{\theta_0}(Z)\right)\right]$$

Whenever both terms are well-defined

WASSERSTEIN GAN THEOREM 3

WGAN paper uses this to derive a **gradient formula**

Theorem 3. *Let $\mathbb{P}_r$ be any distribution. Let $\mathbb{P}_\theta$ be the distribution of $g_\theta(Z)$ with Z a random variable with density p and g_θ a function satisfying assumption 1. Then, there is a solution $f : \mathcal{X} \to \mathbb{R}$ to the problem*

$$\max_{\|f\|_L \leq 1} \mathbb{E}_{x \sim \mathbb{P}_r}[f(x)] - \mathbb{E}_{x \sim \mathbb{P}_\theta}[f(x)]$$

and we have

$$\nabla_\theta W(\mathbb{P}_r, \mathbb{P}_\theta) = -\mathbb{E}_{z \sim p(z)}[\nabla_\theta f(g_\theta(z))]$$

when both terms are well-defined.

Proof. See Appendix C $\square$

This formula may never hold!

"Whenever both terms are well-defined"

WASSERSTEIN GAN THEOREM 3

Let $f \in X^*(\theta)$, which we knows exists since $X^*(\theta)$ is non-empty for all θ. Then, we get

$$\begin{aligned}\nabla_\theta W(\mathbb{P}_r, \mathbb{P}_\theta) &= \nabla_\theta V(f, \theta) \\ &= \nabla_\theta[\mathbb{E}_{x\sim\mathbb{P}_r}[f(x)] - \mathbb{E}_{z\sim p(z)}[f(g_\theta(z))] \\ &= -\nabla_\theta \mathbb{E}_{z\sim p(z)}[f(g_\theta(z))]\end{aligned}$$

under the condition that the first and last terms are well-defined. The rest of the proof will be dedicated to show that

$$-\nabla_\theta \mathbb{E}_{z\sim p(z)}[f(g_\theta(z))] = -\mathbb{E}_{z\sim p(z)}[\nabla_\theta f(g_\theta(z))] \tag{5}$$

when the right hand side is defined. For the reader who is not interested in such technicalities, he or she can skip the rest of the proof.

Since $f \in \mathcal{F}$, we know that it is 1-Lipschitz. Furthermore, $g_\theta(z)$ is locally Lipschitz as a function of (θ, z). Therefore, $f(g_\theta(z))$ is locally Lipschitz on (θ, z) with constants $L(\theta, z)$ (the same ones as g). By Radamacher's Theorem, $f(g_\theta(z))$ has to be differentiable almost everywhere for (θ, z) jointly. Rewriting this, the set $A = \{(\theta, z) : f \circ g \text{ is not differentiable}\}$ has measure 0. By Fubini's Theorem, this implies that for almost every θ the section $A_\theta = \{z : (\theta, z) \in A\}$ has measure 0. Let's now fix a θ_0 such that the measure of A_{θ_0} is null (**such as when the right hand side of equation (5) is well defined**). For this θ_0 we have $\nabla_\theta f(g_\theta(z))|_{\theta_0}$ is well-defined for almost any z, and since $p(z)$ has a density, it is defined $p(z)$-a.e. By assumption 1 we know that

$$\mathbb{E}_{z\sim p(z)}[\|\nabla_\theta f(g_\theta(z))|_{\theta_0}\|] \le \mathbb{E}_{z\sim p(z)}[L(\theta_0, z)] < +\infty$$

so $\mathbb{E}_{z\sim p(z)}[\nabla_\theta f(g_\theta(z))|_{\theta_0}]$ is well-defined for almost every θ_0. Now, we can see

and since $\mathbb{E}_{z\sim p(z)}[2L(\theta_0, z)] < +\infty$ by assumption 1, we get by dominated convergence that Equation 6 converges to 0 as $\theta \to \theta_0$ so

$$\nabla_\theta \mathbb{E}_{z\sim p(z)}[f(g_\theta(z))] = \mathbb{E}_{z\sim p(z)}[\nabla_\theta f(g_\theta(z))]$$

for almost every θ, and in particular when the right hand side is well defined. Note that the mere existance of the left hand side (meaning the differentiability a.e. of $\mathbb{E}_{z\sim p(z)}[f(g_\theta(z))]$) had to be proven, which we just did. □

Fixing **f** assume that we have fixed $\boldsymbol{\theta}$:

Let $\boldsymbol{\theta}$, and let **f** be in $\boldsymbol{X}^*(\boldsymbol{\theta})$ …

A depends on f

$\boldsymbol{\theta_0}$ may therefore be different than $\boldsymbol{\theta}$

may therefore never be defined at $\boldsymbol{\theta}$…

True but this was needed for this specific $\boldsymbol{\theta}$…

WASSERSTEIN GAN THEOREM 3 FAILING CASE

Proposition 2 of our paper: a **counter-example where the formula never holds**

$\mu_\theta = \delta_\theta$

$\nu = \frac{1}{2}\delta_{y_1} + \frac{1}{2}\delta_{y_2}$

c-transform is never differentiable on the edge of Laguerre's cells.

θ_0 is always on the edge of these cells for ψ_0^*

OUR PROPOSITIONS

On the Gradient Formula for learning Generative Models with Regularized Optimal Transport Costs

Antoine Houdard, Arthur Leclaire, Nicolas Papadakis, Julien Rabin

Published: 15 Jul 2023, Last Modified: 15 Jul 2023 Accepted by TMLR Everyone Revisions BibTeX

- **Existence conditions** and **formulation of the gradient** in the **semi-discrete case** (Theorems 3 and 4)
- **Existence conditions** and **formulation of the gradient** for **regularized entropic optimal transport** (Theorem 5)
- **Existence conditions** and **formulation of the gradient** for **Sinkhorn divergence** (Theorem 6)

These formulations give a way to learn a generative model with stochastic gradient descent on the optimal transport cost

CORRESPONDING ALGORITHM

Dataset are **always finite**: semi-discrete case

In this case, we can approximate an **optimal potential** with **gradient ascent**

Algorithm:
Initialize parameters of generative model θ
Iterate:

compute optimal potential ψ_θ^* with gradient ascent
perform a batch step of stochastic gradient descent with formula from theorems

Available soon here: ahoudard / SDOT

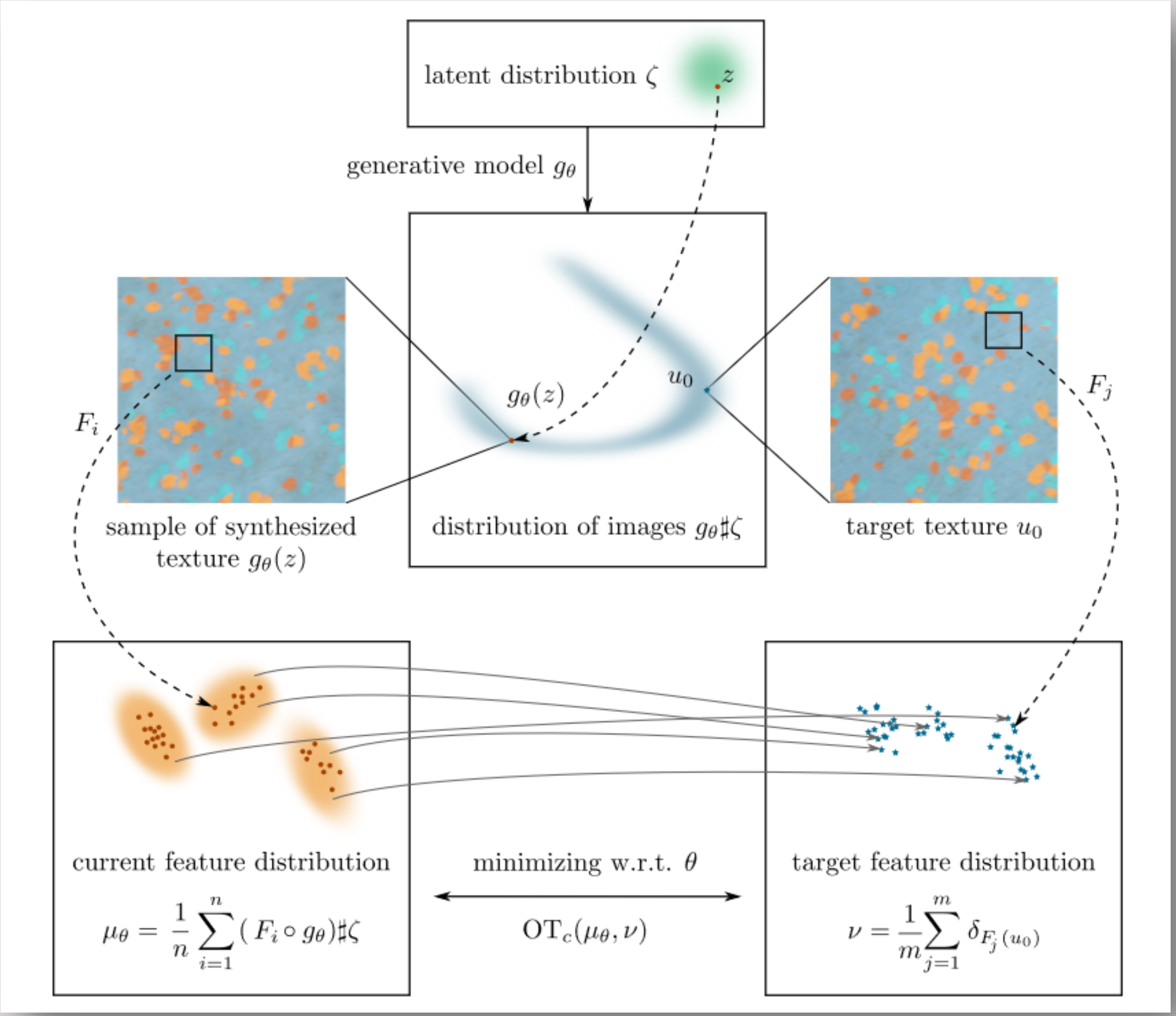
EXAMPLE ON MNIST

$\lambda = 0$ $\lambda = 0.025$ $\lambda = 0.05$ $\lambda = 0.075$ $\lambda = 0.1$

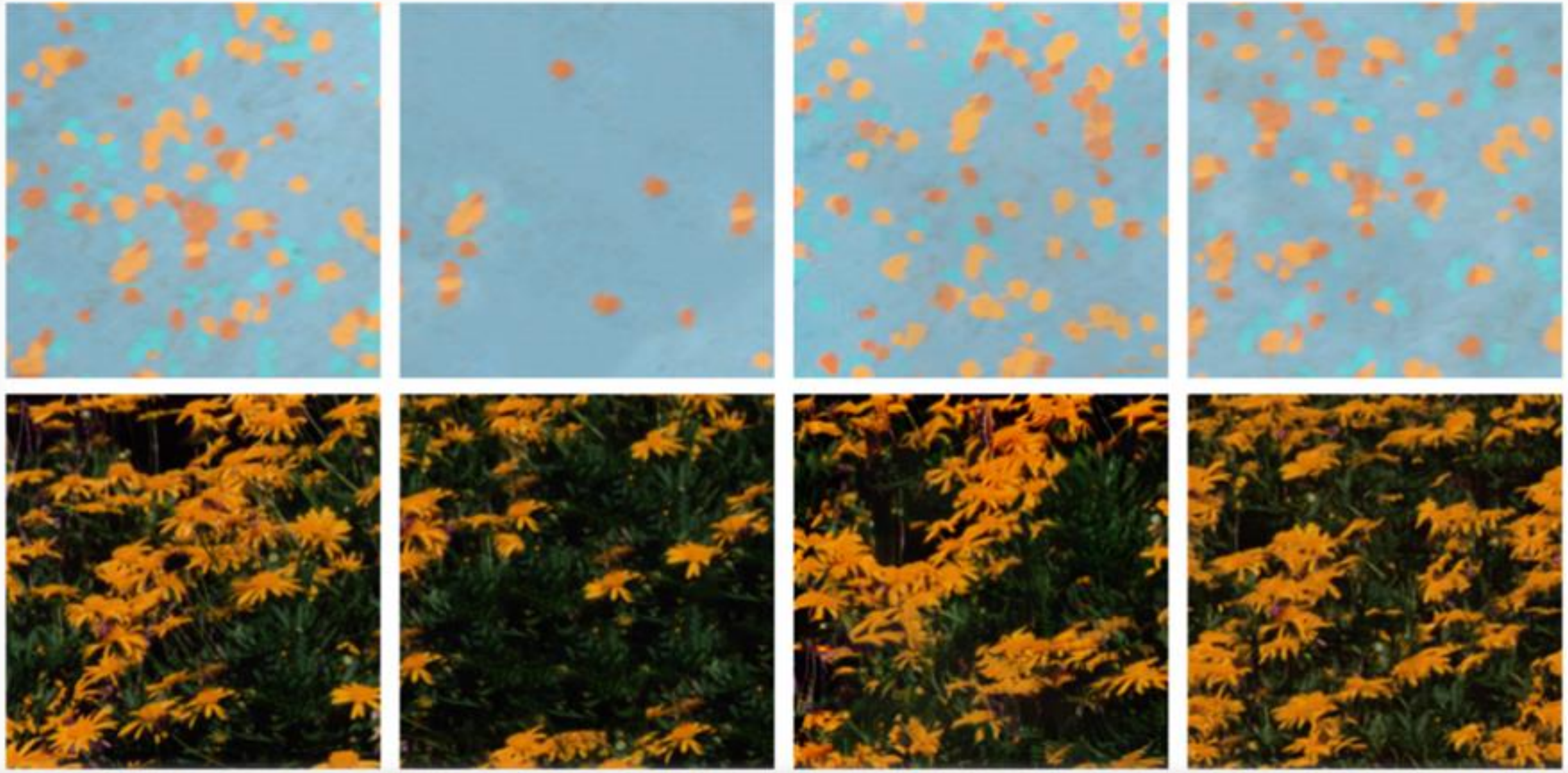
- ✓ **Semi-discrete case** allows to lean generative model **without discriminator network**
- ✓ proper **optimal transport formulation**
- ❖ when dataset is large, **dual potential is large accordingly** which impact performance
- ❖ need to compute an **optimal potential at eatch iteration**

APPLICATION: TEXTURE SYNTHESIS

Wasserstein generative models for patch-based texture synthesis, SSVM 2021
A generative model for texture synthesis based on optimal transport between feature distributions, JMIV 2022

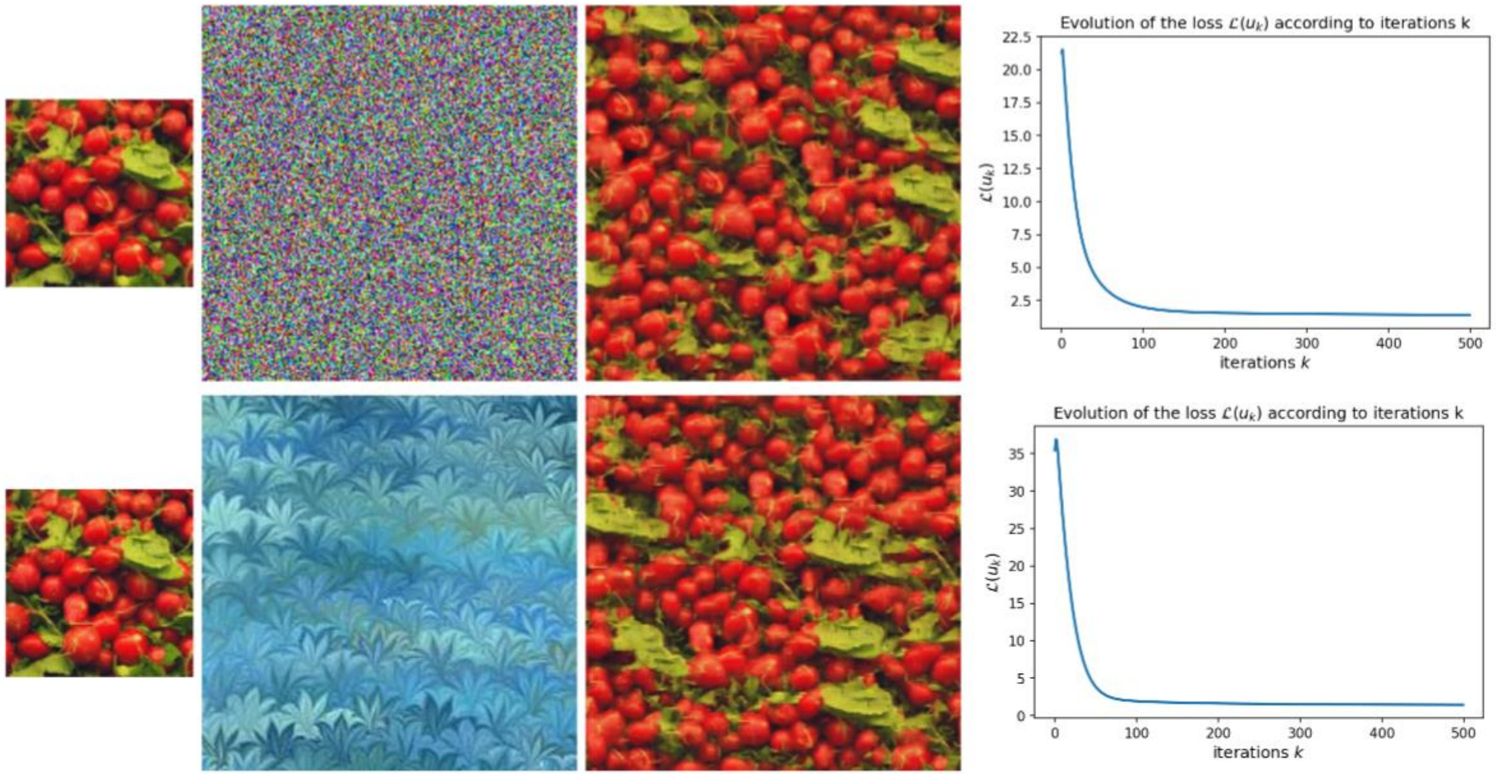


TEXTURE SYNTHESIS RESULTS



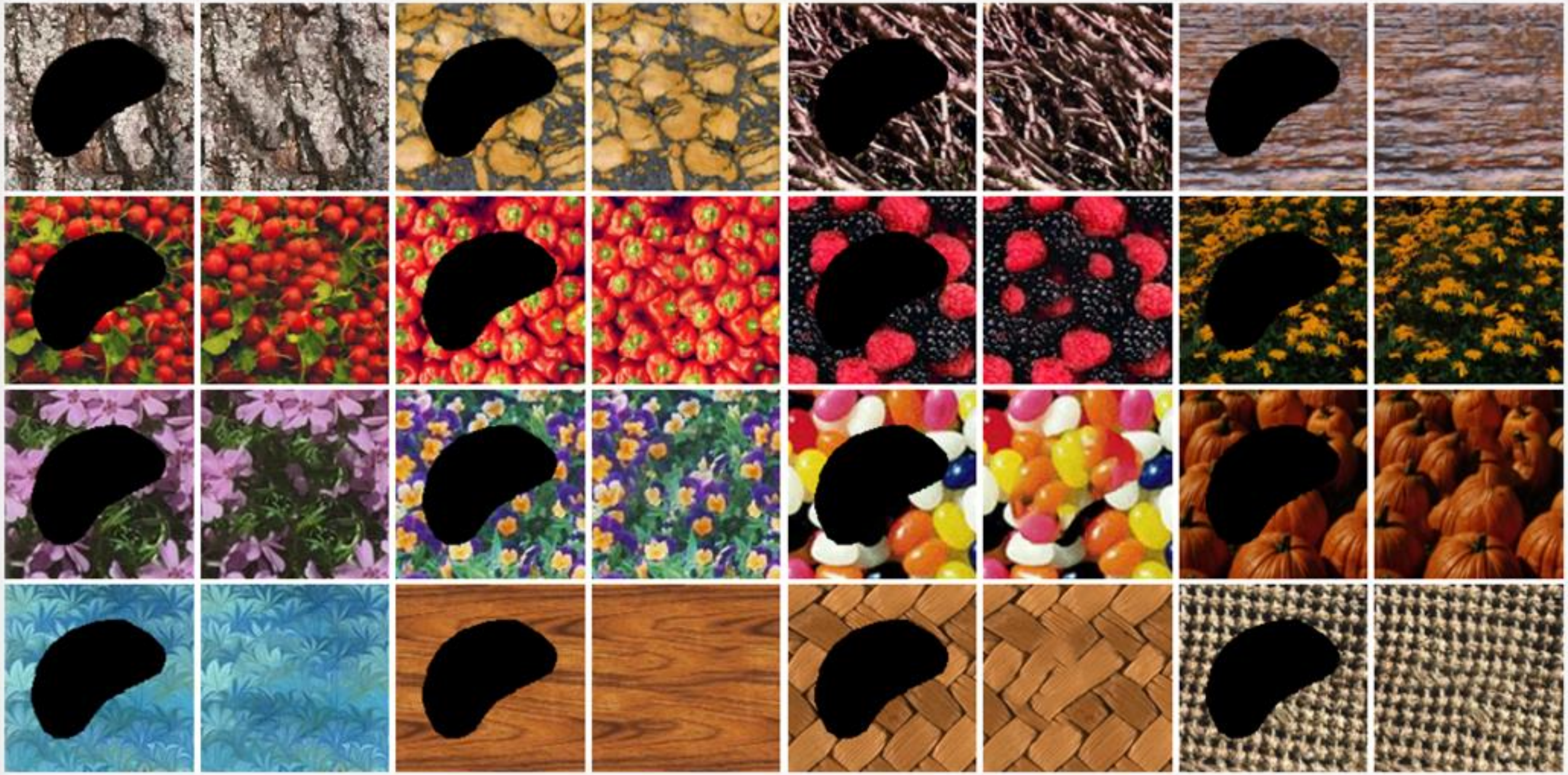
Example patchNN Gatys et al. Ours

CONVERGENCE ROBUSTNESS



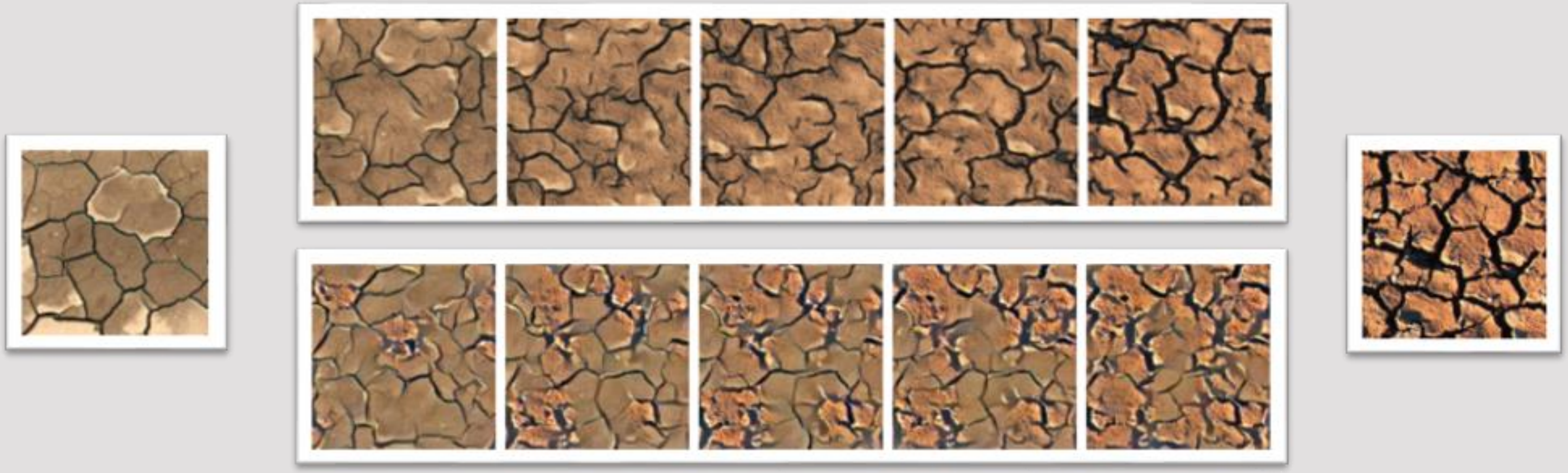
a. Sample b. Initialization c. Alg. 1 k=500 d. Loss

TEXTURE INPAINTING RESULTS



TEXTURE INTERPOLATION RESULTS

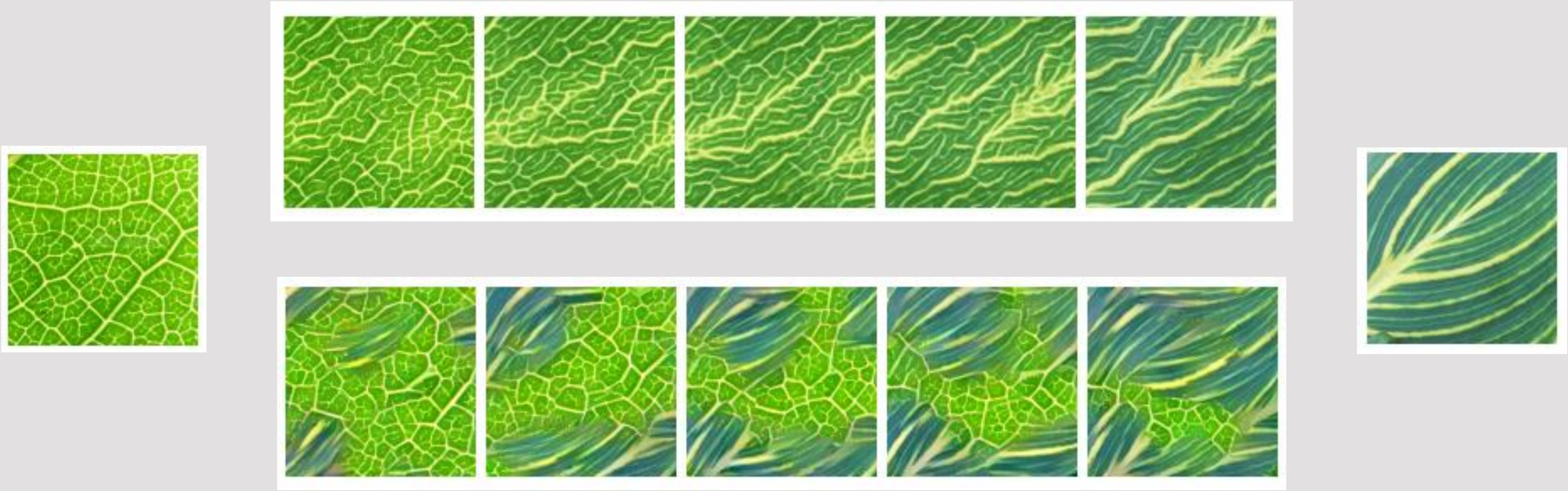
Our — Optimal transport barycenter between patch distributions



Gatys — Gram loss interpolation between VGG features

TEXTURE INTERPOLATION RESULTS

Our — Optimal transport barycenter between patch distributions



Gatys — Gram loss interpolation between VGG features

THANKS!

All papers available online (HAL or ArXiv)
Texture synthesis code github.com/ahoudard/GOTEX

Github ahoudard
Twitter @AntoineHou