

Differentiable EM and applications with GMMOT

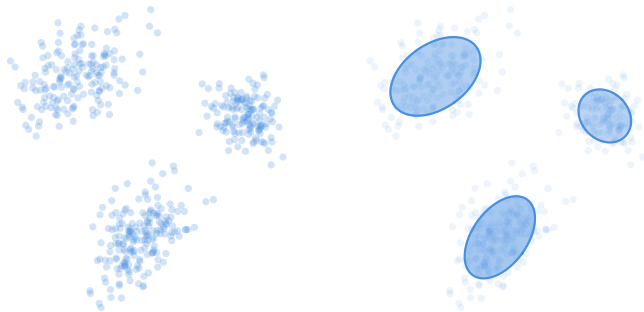
Samuel Boité, Eloi Tanguy, Julie Delon,
Agnès Desolneux, Rémi Flamary

JDS 2026 – 1st June 2026

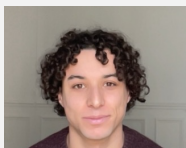


Differentiable Expectation-Maximization and (toy) applications with Gaussian Mixture OT

[Boi+25]



$$\text{Data } X \in (\mathbb{R}^d)^n \xrightarrow{\nabla ?} \text{EM}(X) : \text{GMM}$$



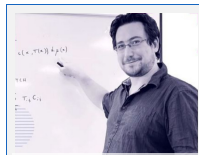
Samuel Boit e



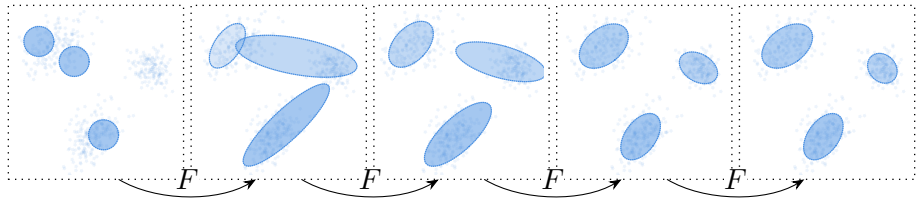
Julie Delon



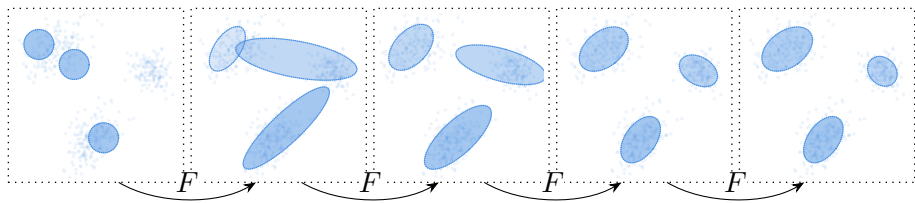
Agn es Desolneux



R emi Flamary



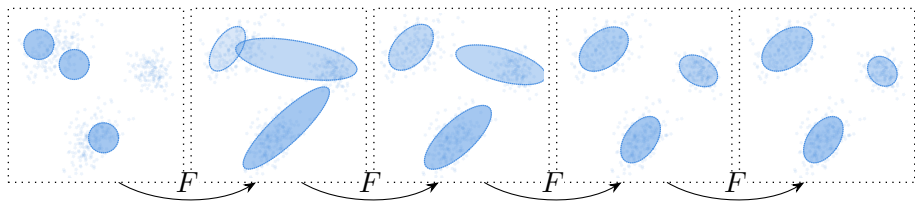
$$\theta_t = F(\theta_{t-1}, X) = F_X^t(\theta_0).$$



$$\theta_t = F(\theta_{t-1}, X) = F_X^t(\theta_0).$$

Differentiation methods:

- Automatic: $\frac{\partial \theta_T}{\partial X} = \frac{\partial}{\partial X} [F_X^T(\theta_0)]$.

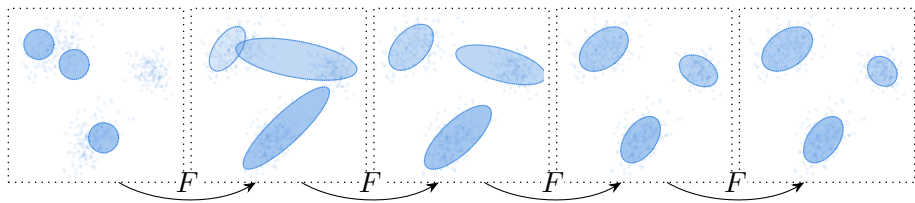


$$\theta_t = F(\theta_{t-1}, X) = F_X^t(\theta_0).$$

Differentiation methods:

- Automatic: $\frac{\partial \theta_T}{\partial X} = \frac{\partial}{\partial X} [F_X^T(\theta_0)]$.

$$F(\theta_\infty, X) = \theta_\infty.$$

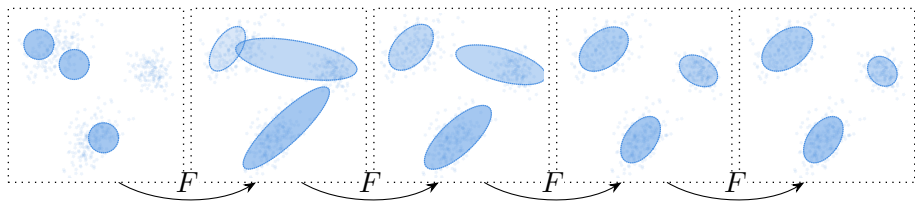


$$\theta_t = F(\theta_{t-1}, X) = F_X^t(\theta_0).$$

Differentiation methods:

- Automatic: $\frac{\partial \theta_T}{\partial X} = \frac{\partial}{\partial X} [F_X^T(\theta_0)]$.

$$\frac{\partial}{\partial X} [F(\theta_\infty, X)] = \frac{\partial \theta_\infty}{\partial X}.$$

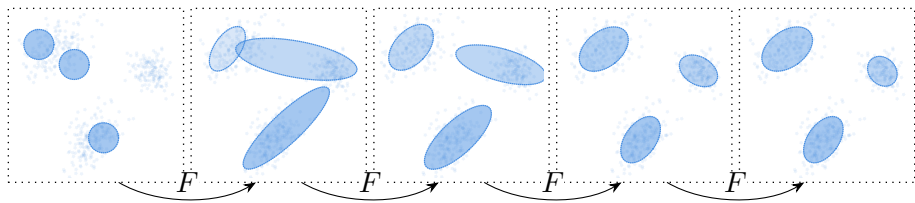


$$\theta_t = F(\theta_{t-1}, X) = F_X^t(\theta_0).$$

Differentiation methods:

- Automatic: $\frac{\partial \theta_T}{\partial X} = \frac{\partial}{\partial X} [F_X^T(\theta_0)]$.

$$\frac{\partial F}{\partial \theta}(\theta_\infty, X) \frac{\partial \theta_\infty}{\partial X} + \frac{\partial F}{\partial X}(\theta_\infty, X) = \frac{\partial \theta_\infty}{\partial X}.$$

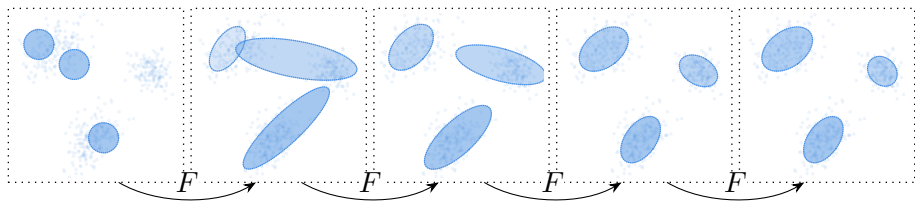


$$\theta_t = F(\theta_{t-1}, X) = F_X^t(\theta_0).$$

Differentiation methods:

- Automatic: $\frac{\partial \theta_T}{\partial X} = \frac{\partial}{\partial X} [F_X^T(\theta_0)]$.

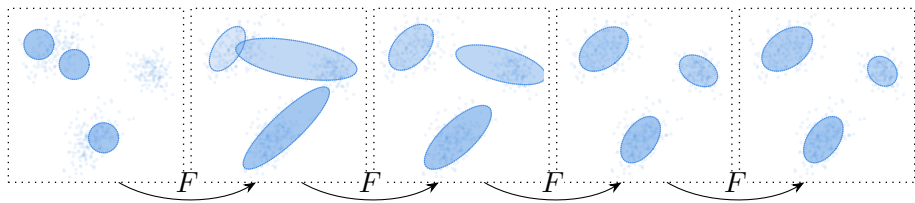
$$\frac{\partial \theta_\infty}{\partial X} = \left(I - \frac{\partial F}{\partial \theta}(\theta_\infty, X) \right)^{-1} \frac{\partial F}{\partial X}(\theta_\infty, X).$$



$$\theta_t = F(\theta_{t-1}, X) = F_X^t(\theta_0).$$

Differentiation methods:

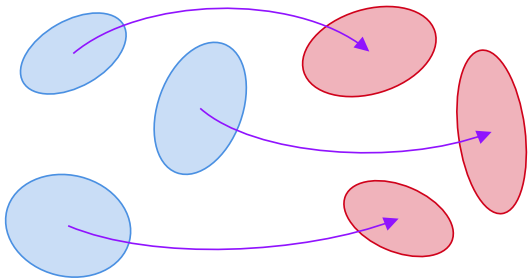
- Automatic: $\frac{\partial \theta_T}{\partial X} = \frac{\partial}{\partial X} [F_X^T(\theta_0)]$.
- Implicit: $\frac{\partial \theta_T}{\partial X} \approx \left(I - \frac{\partial F}{\partial \theta}(\theta_T, X) \right)^{-1} \frac{\partial F}{\partial X}(\theta_T, X)$.

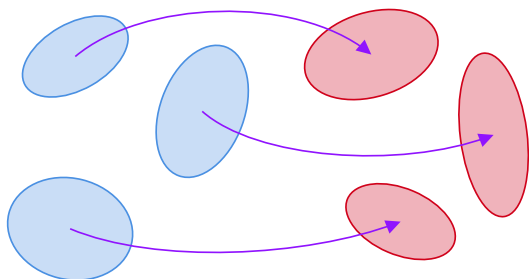


$$\theta_t = F(\theta_{t-1}, X) = F_X^t(\theta_0).$$

Differentiation methods:

- Automatic: $\frac{\partial \theta_T}{\partial X} = \frac{\partial}{\partial X} [F_X^T(\theta_0)]$.
- Implicit: $\frac{\partial \theta_T}{\partial X} \approx \left(I - \frac{\partial F}{\partial \theta}(\theta_T, X) \right)^{-1} \frac{\partial F}{\partial X}(\theta_T, X)$.
- One-Step: $\frac{\partial \theta_T}{\partial X} \approx \frac{\partial F}{\partial X}(\theta_T, X)$.



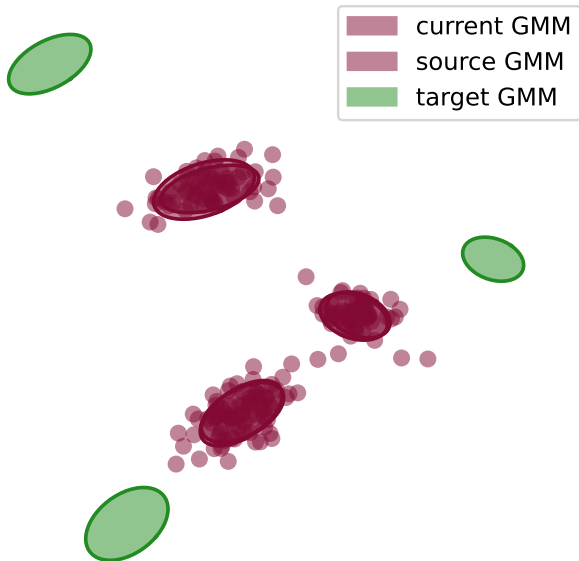


Mixture-Wasserstein Distance [DD20]

$$\begin{aligned}
 \text{MW}_2^2(\mu_0, \mu_1) &= \min_{P \in \Pi(w_0, w_1)} \sum_{k, \ell} P_{k\ell} W_2^2(\mu_{0,k}, \mu_{1,\ell}) \\
 &= \min_{\pi \in \Pi(\mu_0, \mu_1) \cap \text{GMM}} \int_{\mathbb{R}^{2d}} \|x_1 - x_2\|_2^2 d\pi(x_1, x_2).
 \end{aligned}$$

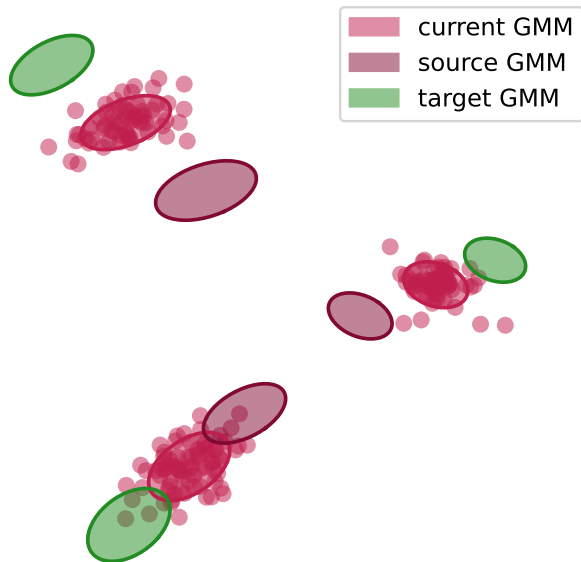
Gradient Descent on $X \mapsto \text{MW}_2^2(\mu(F_X^T(\theta_0)), \nu)$.

EM Flow Iteration 001/300



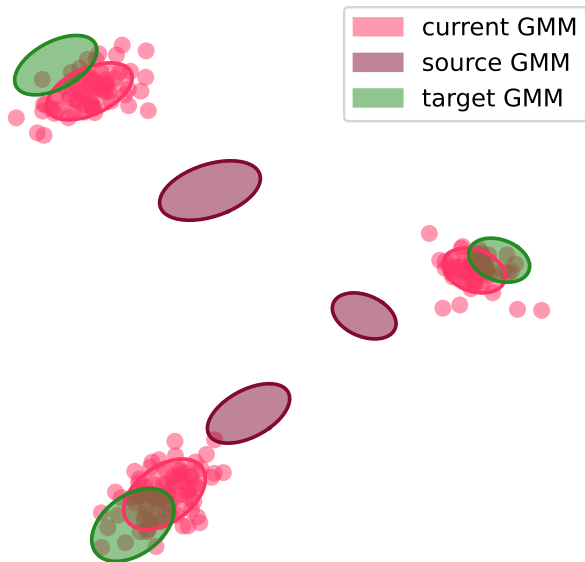
Gradient Descent on $X \mapsto \text{MW}_2^2(\mu(F_X^T(\theta_0)), \nu)$.

EM Flow Iteration 075/300



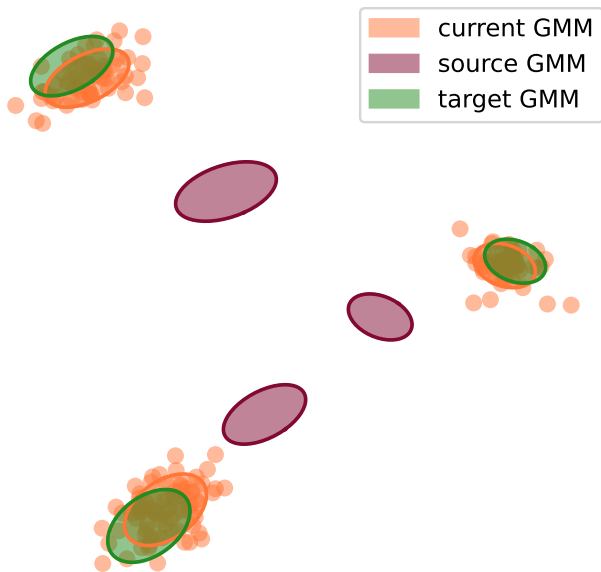
Gradient Descent on $X \mapsto \text{MW}_2^2(\mu(F_X^T(\theta_0)), \nu)$.

EM Flow Iteration 150/300



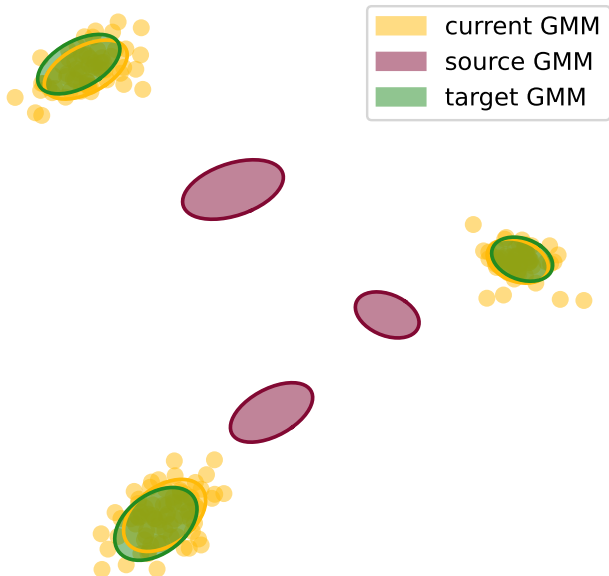
Gradient Descent on $X \mapsto \text{MW}_2^2(\mu(F_X^T(\theta_0)), \nu)$.

EM Flow Iteration 225/300

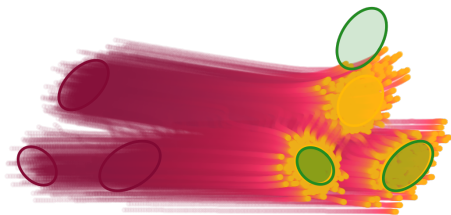


Gradient Descent on $X \mapsto \text{MW}_2^2(\mu(F_X^T(\theta_0)), \nu)$.

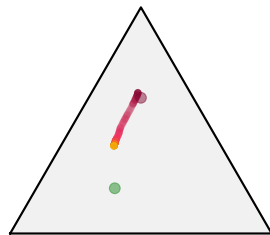
EM Flow Iteration 300/300



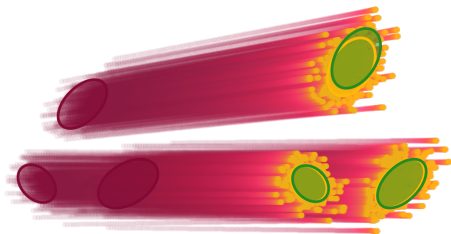
Gradient Descent on $X \mapsto \text{MW}_2^2(\mu(F_X^T(\theta_0)), \nu)$.



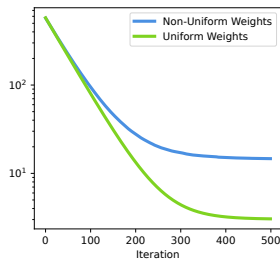
Particle flow (non-uniform).



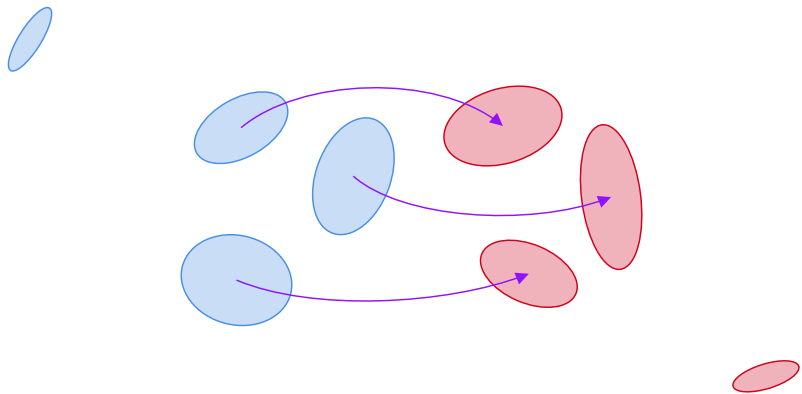
Weight evolution (non-uniform).

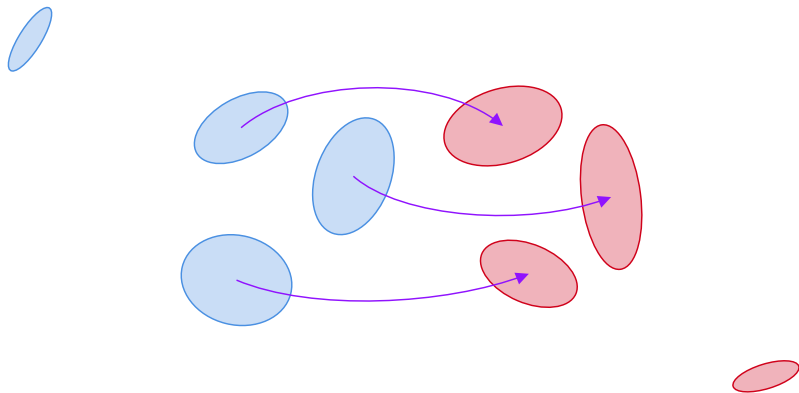


Particle flow (uniform).



Energy evolutions.



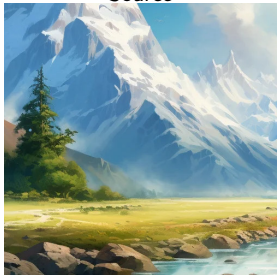


Unbalanced Mixture-Wasserstein Distance [DD20]

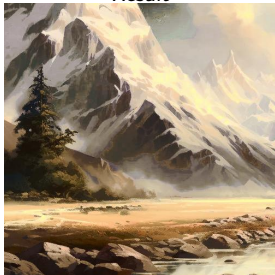
$$\text{MW}_2^2(\mu_0, \mu_1; \lambda_0, \lambda_1) = \min_{P \in \mathbb{R}_+^{K_0 \times K_1}} \sum_{k, \ell} P_{k\ell} W_2^2(\mu_{0,k}, \mu_{1,\ell}) + \lambda_0 \text{KL}(P\mathbf{1} | w_0) + \lambda_1 \text{KL}(P^\top \mathbf{1} | w_1).$$

Gradient Descent on $X \in (\mathbb{R}^3)^{h \times w} \mapsto \text{MW}_2^2 \left(\mu \left(F_X^T(\theta) \right), \mu \left(F_Y^T(\theta_0) \right) \right)$.

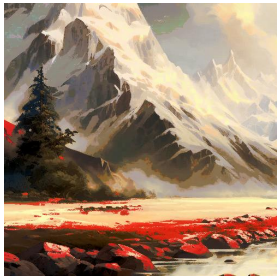
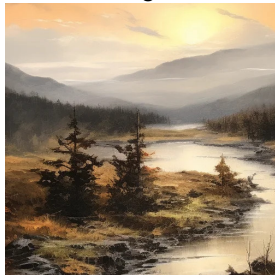
Source



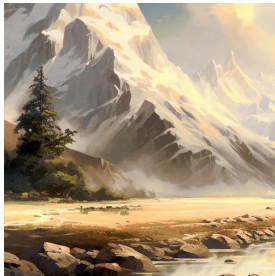
Result



Target



Balanced result

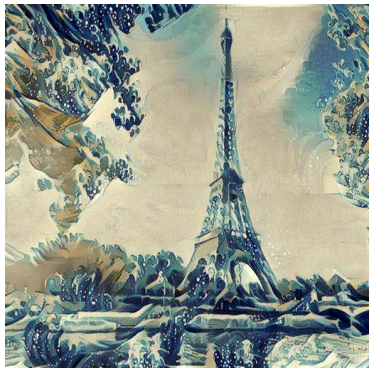
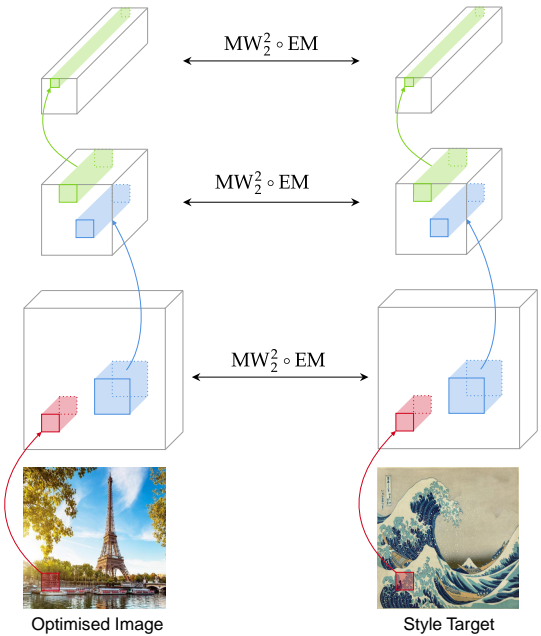


Unbalanced result



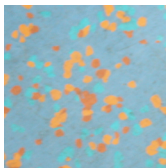
Corrupted target

Application: Style Transfer

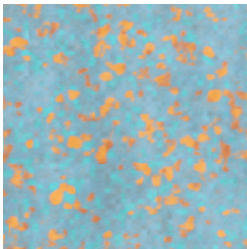
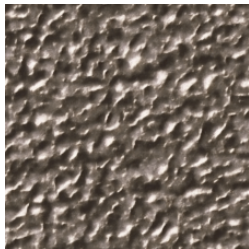


$$\min_X \sum_{\ell=1}^L \lambda_{\ell} \text{MW}_2^2 \left(F^T \circ \text{Patches}_{p \times p} \circ \text{Downscale}_{s_{\ell}}(X), \right. \\ \left. F^T \circ \text{Patches}_{p \times p} \circ \text{Downscale}_{s_{\ell}}(Y) \right).$$

Reference



Generated



A vibrant, colorful space-themed illustration. The background is a deep blue and purple starry sky. On the left, a large blue planet with a prominent ring system is visible. In the center, a multi-tiered, colorful rocket ship or space station is depicted, featuring various platforms, a tall spire, and glowing elements. A yellow star character with a smiling face is positioned to the right of the central structure. The word "Thanks!" is written in a large, white, bold font with a black outline, centered over the scene.

Thanks!

Bibliography

- [Boï+25] Samuel Boïté, Eloi Tanguy, Julie Delon, Agnès Desolneux, and Rémi Flamary. *Differentiable Expectation-Maximisation and Applications to Gaussian Mixture Model Optimal Transport*. 2025. arXiv: 2509.02109 [cs.LG] (cit. on p. 2).
- [DD20] Julie Delon and Agnes Desolneux. “A Wasserstein-type distance in the space of Gaussian mixture models”. In: *SIAM Journal on Imaging Sciences* 13.2 (2020), pp. 936–970 (cit. on pp. 11, 12, 19, 20).

$$F(\theta, X) = (w', m', \Sigma') :$$

$$\gamma_{ik}(\theta) = \frac{w_k g_{m_k, \Sigma_k}(x_i)}{\sum_{\ell=1}^K w_\ell g_{m_\ell, \Sigma_\ell}(x_i)},$$

$$w'_k = \frac{1}{n} \sum_{i=1}^n \gamma_{ik}(\theta),$$

$$m'_k = \frac{\sum_{i=1}^n \gamma_{ik}(\theta) x_i}{\sum_{j=1}^K \gamma_{jk}(\theta)},$$

$$\Sigma'_k = \frac{\sum_{i=1}^n \gamma_{ik}(\theta) (x_i - m'_k)(x_i - m'_k)^\top}{\sum_{j=1}^K \gamma_{jk}(\theta)}.$$