Differentiable EM and applications with GMMOT

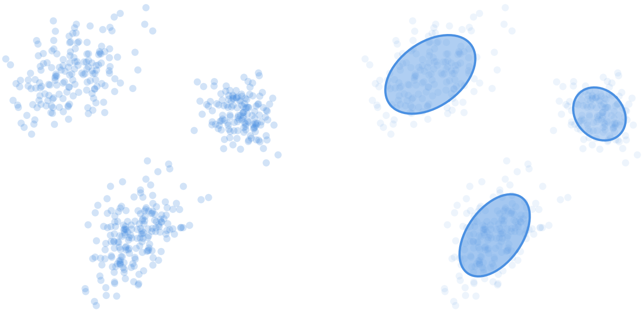
Samuel Boïté, Eloi Tanguy, Julie Delon,
Agnès Desolneux, Rémi Flamary

JDS 2026 – 1st June 2026

LABORATOIRE MAP5

Université Paris Cité

Differentiable Expectation-Maximization and (toy) applications with Gaussian Mixture OT [Boï+25]

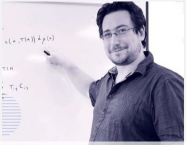


Data $X \in (\mathbb{R}^d)^n \xmapsto{\nabla ?} \mathrm{EM}(X)$: GMM

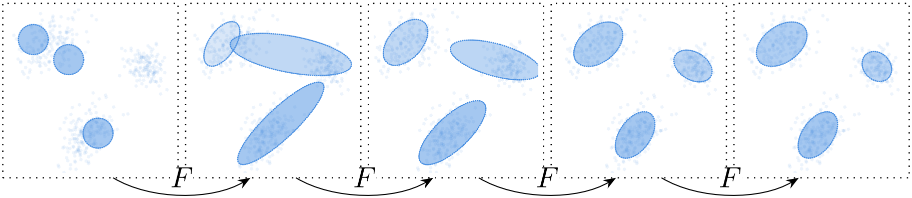
Samuel Boïté

Julie Delon

Agnès Desolneux

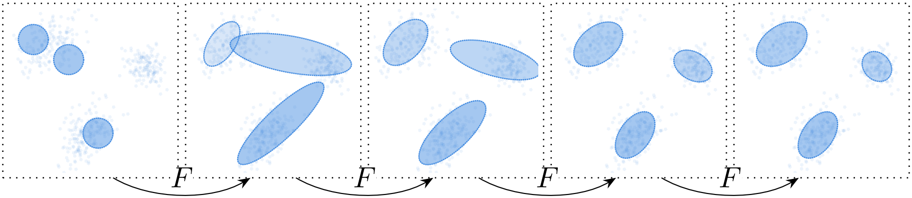


Rémi Flamary



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

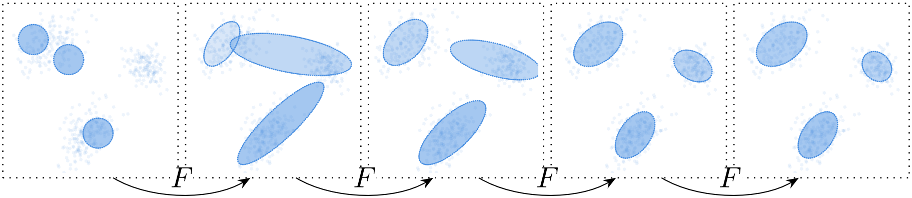
Differentiation of EM as a Fixed-Point Algorithm



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

Differentiation methods:

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

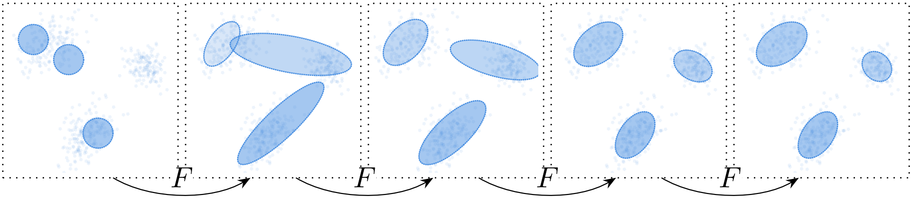


$$\theta_t = F(\theta_{t-1}, X) = F_X^t(\theta_0).$$

Differentiation methods:

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

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

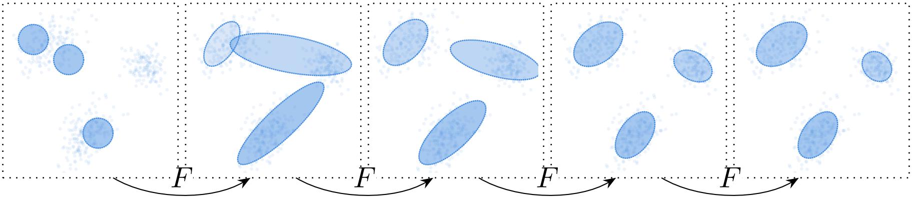


$$\theta_t = F(\theta_{t-1}, X) = F_X^t(\theta_0).$$

Differentiation methods:

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

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

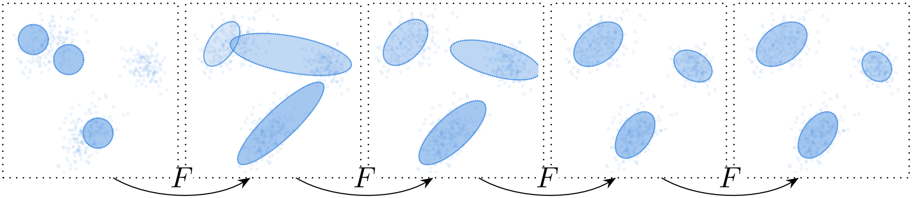


$$\theta_t = F(\theta_{t-1}, X) = F_X^t(\theta_0).$$

Differentiation methods:

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

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

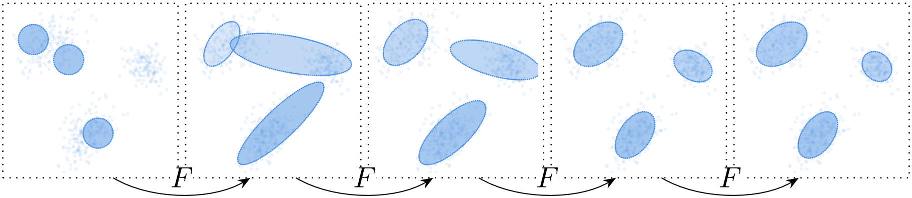


$$\theta_t = F(\theta_{t-1}, X) = F_X^t(\theta_0).$$

Differentiation methods:

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

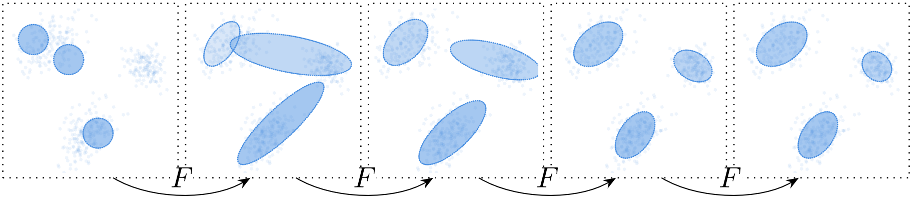
$$\frac{\partial \theta_\infty}{\partial X} = \left(I - \frac{\partial F}{\partial \theta}(\theta_\infty, X)\right)^{-1} \frac{\partial F}{\partial X}(\theta_\infty, X).$$



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

Differentiation methods:

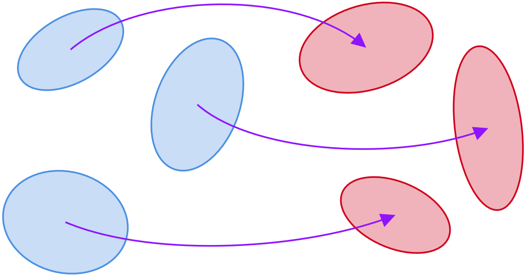
- Automatic: $\dfrac{\partial \theta_T}{\partial X} = \dfrac{\partial}{\partial X}\left[F_X^T(\theta_0)\right].$
- Implicit: $\dfrac{\partial \theta_T}{\partial X} \approx \left(I - \dfrac{\partial F}{\partial \theta}(\theta_T, X)\right)^{-1} \dfrac{\partial F}{\partial X}(\theta_T, X).$

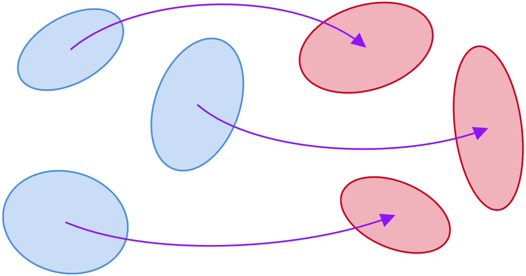


$$\theta_t = F(\theta_{t-1}, X) = F_X^t(\theta_0).$$

Differentiation methods:

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



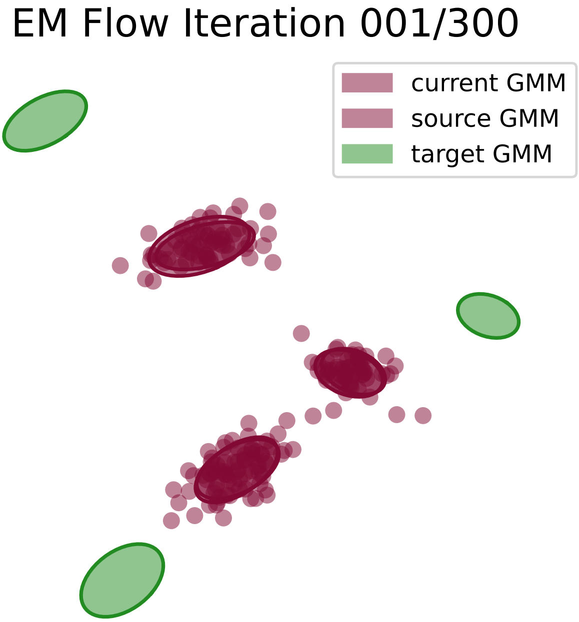


Mixture-Wasserstein Distance [DD20]

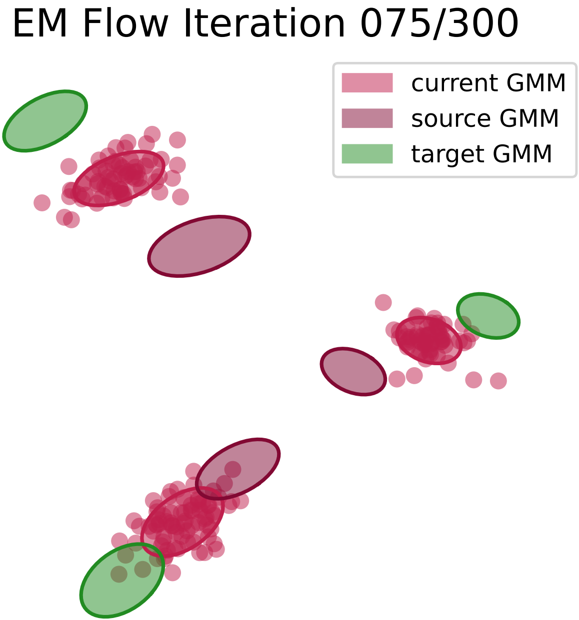
$$\mathrm{MW}_2^2(\mu_0,\mu_1) = \min_{P\in\Pi(w_0,w_1)} \sum_{k,\ell} P_{k\ell}\mathrm{W}_2^2(\mu_{0,k},\mu_{1,\ell})$$

$$= \min_{\pi\in\Pi(\mu_0,\mu_1)\cap\mathrm{GMM}} \int_{\mathbb{R}^{2d}} \|x_1 - x_2\|_2^2 \mathrm{d}\pi(x_1,x_2).$$

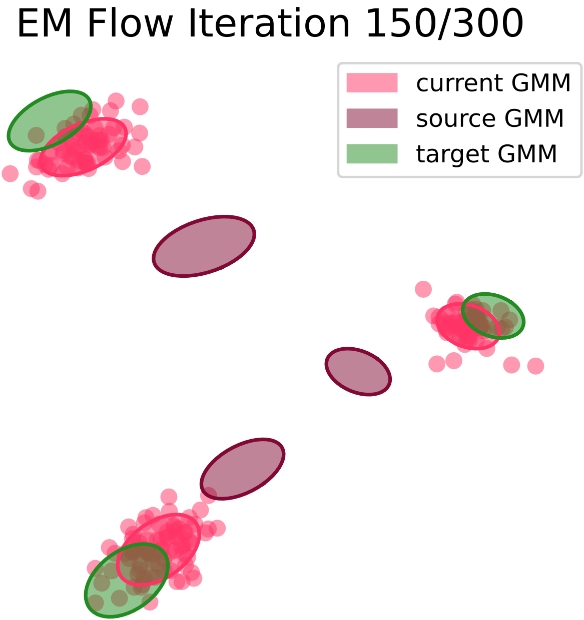
Gradient Descent on $X \longmapsto \mathrm{MW}_2^2(\mu(F_X^T(\theta_0)), \nu)$.



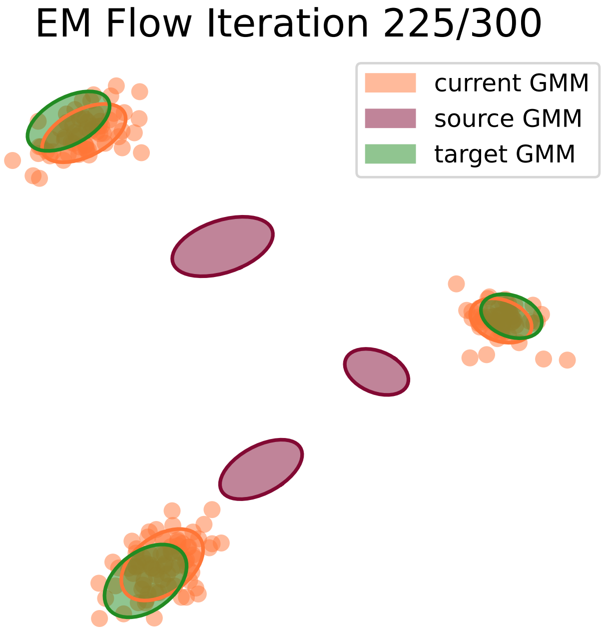
Gradient Descent on $X \longmapsto \mathrm{MW}_2^2(\mu(F_X^T(\theta_0)), \nu)$.



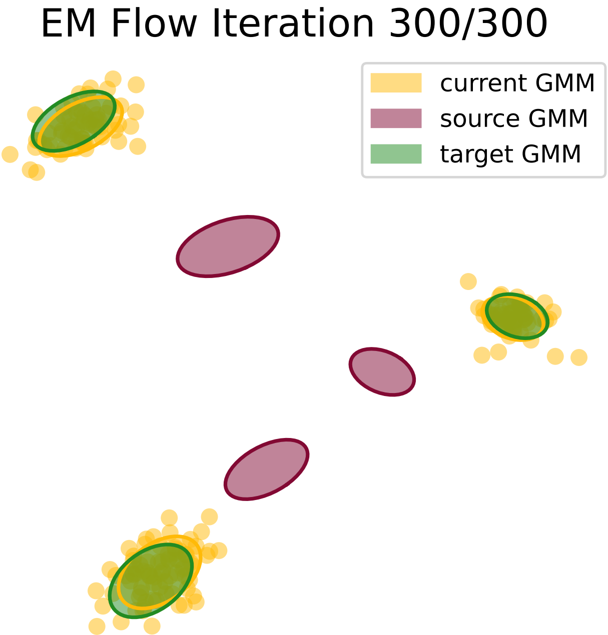
Gradient Descent on $X \longmapsto \mathrm{MW}_2^2(\mu(F_X^T(\theta_0)), \nu)$.



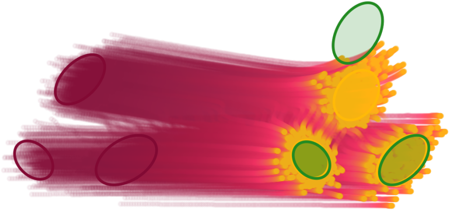
Gradient Descent on $X \longmapsto \mathrm{MW}_2^2(\mu(F_X^T(\theta_0)), \nu)$.



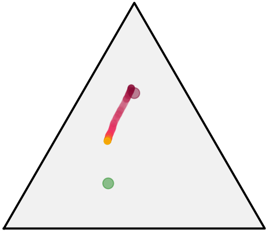
Gradient Descent on $X \longmapsto \mathrm{MW}_2^2(\mu(F_X^T(\theta_0)), \nu)$.



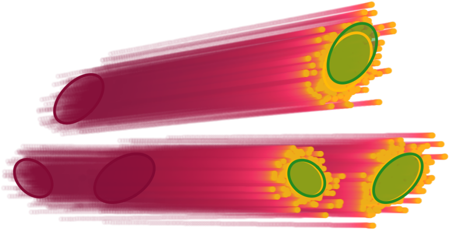
Gradient Descent on $X \longmapsto \mathrm{MW}_2^2(\mu(F_X^T(\theta_0)), \nu)$.



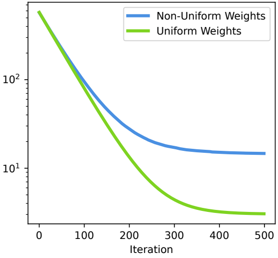
Particle flow (non-uniform).



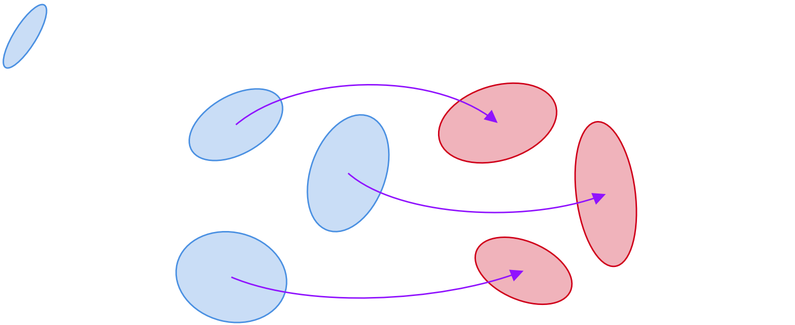
Weight evolution (non-uniform).

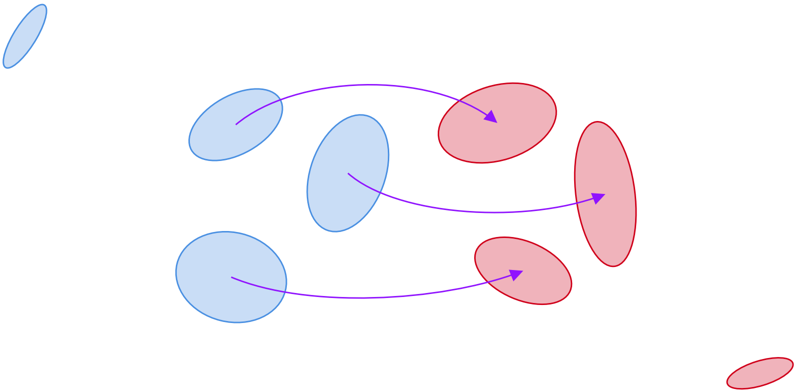


Particle flow (uniform).



Energy evolutions.





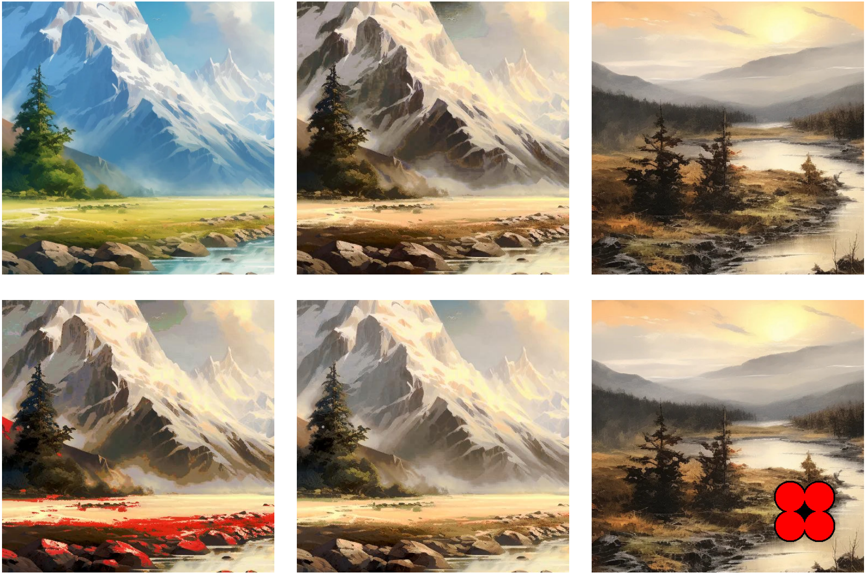
Unbalanced Mixture-Wasserstein Distance [DD20]

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

Application: Colour Transfer

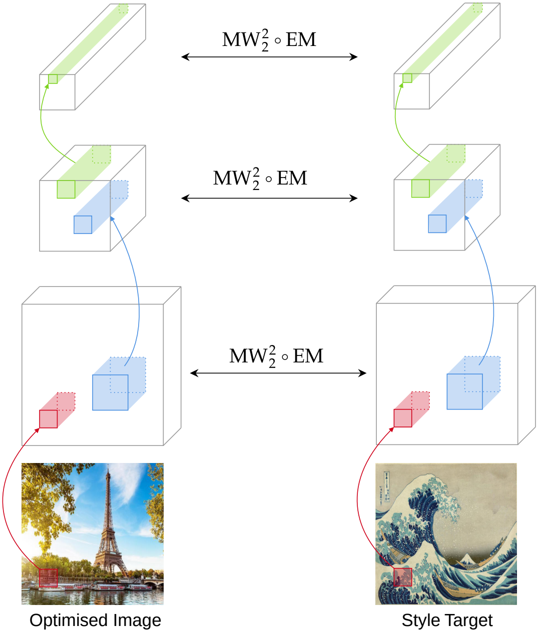
Gradient Descent on $X \in (\mathbb{R}^3)^{h \times w} \longmapsto \mathrm{MW}_2^2\left(\mu\left(F_X^T(\theta)\right), \mu\left(F_Y^T(\theta_0)\right)\right)$.

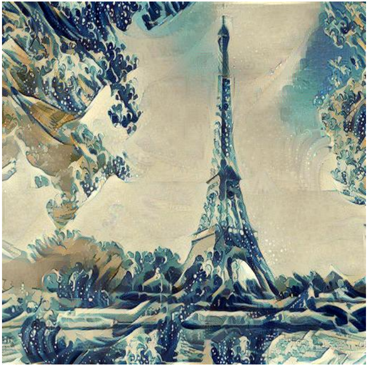
Source | Result | Target



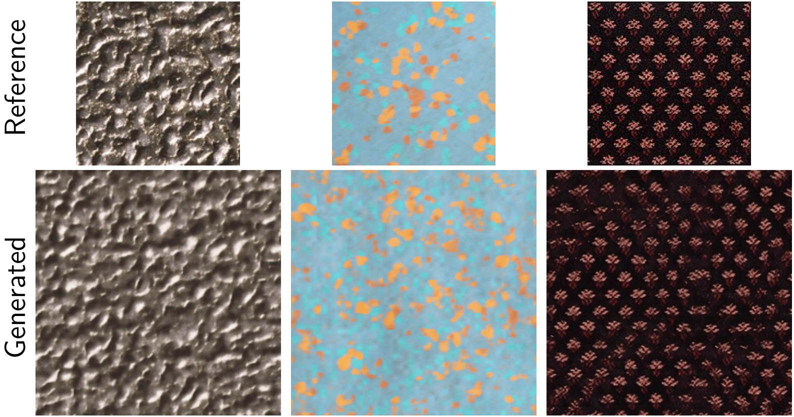
Balanced result | Unbalanced result | Corrupted target

Application: Style Transfer





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




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).

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

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

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

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

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