Project : Deep Optimal Transport for Image Restoration

Mohammed Dahbani
Mahmoud Mokrane

UE F : COMPI

25/03/2025

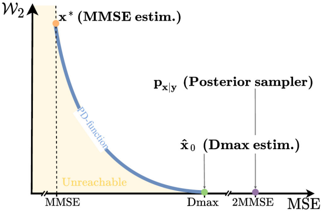
Overview

1. Optimal Transport : quick paper Recap
   - What is OT ?
   - How is it used in practice ?
2. Method Recap
   - What is Deep Optimal Transport (DOT) ?
   - How does DOT work ?
   - Pipeline
3. Question
4. Implementation Strategy
   - Dataset Preparation
   - VAE
5. Experiments & results
   - Experiment
   - Qualitative & quantitative analysis (paired images)
   - Qualitative & quantitative analysis (Unpaired images)

What is OT ?

Optimal transport (OT) is a mathematical framework for finding the most efficient way to move one distribution of mass (probability, resources, ...) to another while minimizing a given cost.

The $\mathcal{W}_2$ distance is linked to the optimal transport problem, as it measures the minimum work required to transform one distribution into another.



This paper defines an estimator that achieves the lowest possible **distortion** (measured with **MSE**) while maintaining great **perceptual quality** (measured with **Wasserstein-2** ($\mathcal{W}_2$) **distance**).

How is it used in practice ?

Goal : We are trying to find the estimator $\hat{\boldsymbol{x}}_0$ that achieves perfect perceptual quality and minimum distortion.

- $\boldsymbol{x}$ is the original image.
- $\boldsymbol{x}^*$ achieves minimum MSE ($MSE(\boldsymbol{x}, \boldsymbol{x}^*) = \boldsymbol{D}_{min}$).
- $\hat{\boldsymbol{x}}_0$ achieves minimum MSE **and** satisfies $\mathcal{W}_2(p_{\boldsymbol{x}}, p_{\hat{\boldsymbol{x}}_0}) = 0$.

Main result

The joint distribution $p_{\hat{\boldsymbol{x}}_0, \boldsymbol{x}^*}$ is an optimal transport plan between $\boldsymbol{x}^*$ and $\boldsymbol{x}$.
Finding this plan gives us access to $\hat{\boldsymbol{x}}_0$ with :

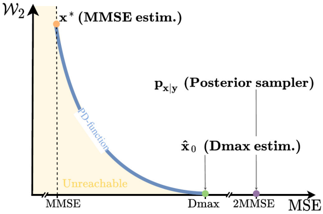
$$\hat{\boldsymbol{x}}_0 = T_{p_{\boldsymbol{x}^*} \longrightarrow p_{\boldsymbol{x}}}(\boldsymbol{x}^*)$$

What is Deep Optimal Transport (DOT) ?

DOT refines model outputs using **optimal transport** in latent space. It balances the **perception-distortion tradeoff** *(without retraining)* to map outputs closer to natural images.

To control the tradeoff between perception and distortion, we introduce an interpolation parameter α :

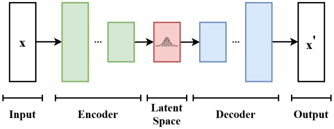
$$\hat{x}_P = (1-\alpha)\hat{x}_0 + \alpha x^*$$



How does DOT work ?

Why Use Latent Representations ?

- OT in pixel space is **expensive** and lacks structure.
- A structured space enables **better learning & generalization**.
- Solution : Map images to a **latent space** $\mathcal{S}$ for efficient OT computation.



VAE architecture

Mathematical Formulation :

$$z_x \sim \mathcal{N}(\mu_x, \Sigma_x), \quad z_{\hat{x}} \sim \mathcal{N}(\mu_{\hat{x}}, \Sigma_{\hat{x}})$$

Pipeline

1- Training Phase (Offline)

- Select **10 restored images** and **10 clean images**.
- Encode them into **latent space** via a **VAE**.
- Compute **mean and covariance matrices**.
- Apply the **optimal transport transformation** :

$$\mathbf{T}^{\mathrm{MVG}}_{p_{x_1}\to p_{x_2}}(x_1) = \Sigma_{x_1}^{-\frac{1}{2}} \left(\Sigma_{x_1}^{\frac{1}{2}} \Sigma_{x_2} \Sigma_{x_1}^{\frac{1}{2}} \right)^{\frac{1}{2}} \Sigma_{x_1}^{-\frac{1}{2}} \cdot (x_1 - \mu_{x_1}) + \mu_{x_2}$$

2- Inference Phase (Test Time)

- **Encode** a new restored image.
- Apply the **optimal transport (OT)** operator.
- **Decode** into the final enhanced image.

Question :

What effect does the variation of α (Perception-Distortion trade-off) have on model results, particularly in terms of fine textures and detail ?

Dataset Preparation

Training Phase

- **Step 1 :** Downloaded images from the web.
- **Step 2 :** Selected **10 clean images**.
- **Step 3 :** Added **Gaussian noise**.
- **Step 4 :** Restored images using Gaussian Blur.

Inference Phase

- **Step 1 :** Choose a **test image**.
- **Step 2 :** Applied the same pipeline : Clean→Noisy→Restored

Choice of Pretrained VAE Model

Why the Pretrained VAE from StabilityAI ?

For the image restoration task, we decided to use the pretrained VAE model, ***sd-vae-ft-ema***, sourced from **Hugging Face**.

Key Reasons for Selection :

- Optimized for better reconstruction of faces and aesthetic features.
- Fine-tuned with **EMA weights** for enhanced stability and sharper results.

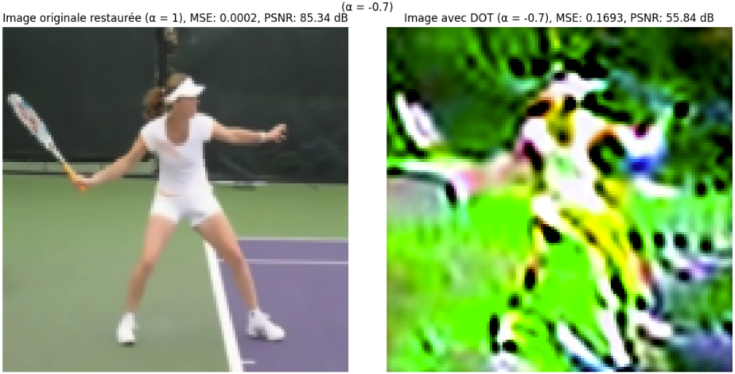
Model Specifics :

- Uses a **L1 + LPIPS loss** for perceptual quality enhancement.
- The **ft-MSE** version smoothens outputs, leveraging **MSE + 0.1 LPIPS**.

Test with Different Alpha Values (Paired images)

- **Alpha Variations :** Alphas tested : {-0.7, -0.5, -0.3, 0.0, 0.3, 0.5, 0.7, 1.0, 1.3, 1.5}.
- **Objective :** Analyze the impact of different alpha values on the image restoration quality.
- **Testing :** For each alpha, apply the OT transformation and compare the result visually.

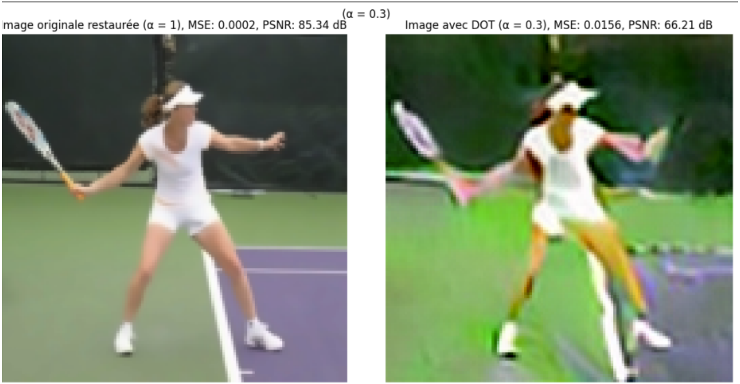
$\alpha = -0.7$


(α = -0.7)
Image originale restaurée (α = 1), MSE: 0.0002, PSNR: 85.34 dB
Image avec DOT (α = -0.7), MSE: 0.1693, PSNR: 55.84 dB

$\alpha = 0$


(α = 0.0)
Image originale restaurée (α = 1), MSE: 0.0002, PSNR: 85.34 dB
Image avec DOT (α = 0.0), MSE: 0.0414, PSNR: 61.96 dB

$\alpha = 0.3$


(α = 0.3)
mage originale restaurée (α = 1), MSE: 0.0002, PSNR: 85.34 dB
Image avec DOT (α = 0.3), MSE: 0.0156, PSNR: 66.21 dB

$\alpha = 0.5$


(α = 0.5)
nage originale restaurée (α = 1), MSE: 0.0002, PSNR: 85.34 dB
Image avec DOT (α = 0.5), MSE: 0.0074, PSNR: 69.45 dB

$\alpha = 0.7$


(α = 0.7)
Image originale restaurée (α = 1), MSE: 0.0002, PSNR: 85.34 dB
Image avec DOT (α = 0.7), MSE: 0.0034, PSNR: 72.86 dB

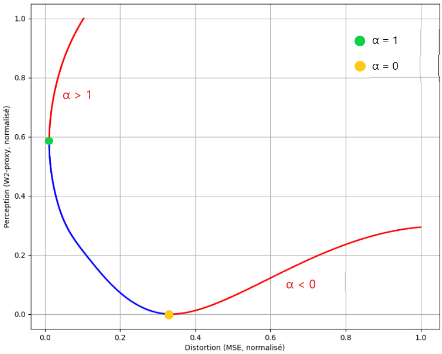
$\alpha = 1$


(α = 1.0)
Image originale restaurée (α = 1), MSE: 0.0002, PSNR: 85.34 dB
Image avec DOT (α = 1.0), MSE: 0.0015, PSNR: 76.48 dB

$\alpha = 1.3$


(α = 1.3)
mage originale restaurée (α = 1), MSE: 0.0002, PSNR: 85.34 dB
Image avec DOT (α = 1.3), MSE: 0.0035, PSNR: 72.66 dB

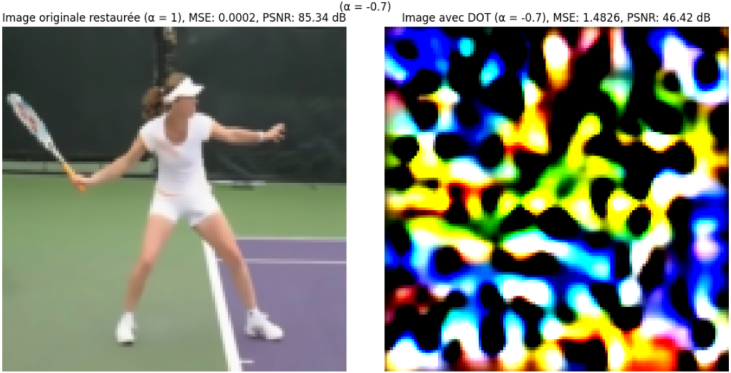
Perception-Distortion Curve



Insights :

- Alpha values closer to 0.7 may give the most balanced restoration.
- Higher positive alpha values produce smoother results, but sometimes at the cost of detail.
- Negative alpha values might result in over-sharpening or excessive cleaning.

$\alpha = -0.7$


(α = -0.7)
Image originale restaurée (α = 1), MSE: 0.0002, PSNR: 85.34 dB
Image avec DOT (α = -0.7), MSE: 1.4826, PSNR: 46.42 dB

$\alpha = 1$

Image originale restaurée (α = 1), MSE: 0.0002, PSNR: 85.34 dB

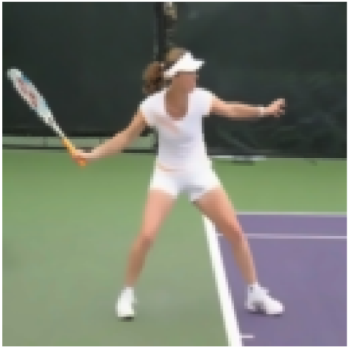
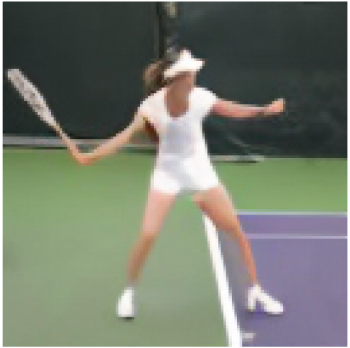
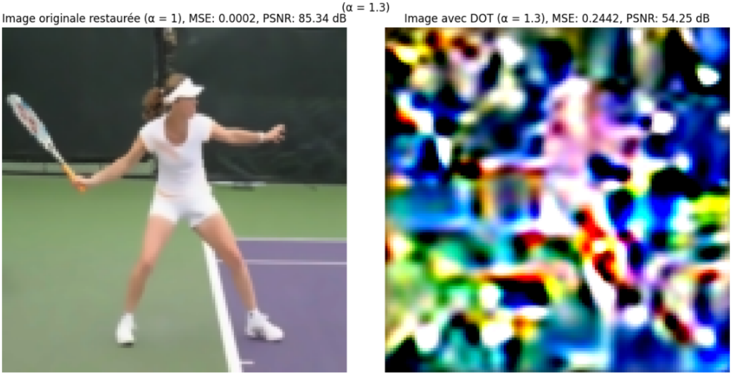


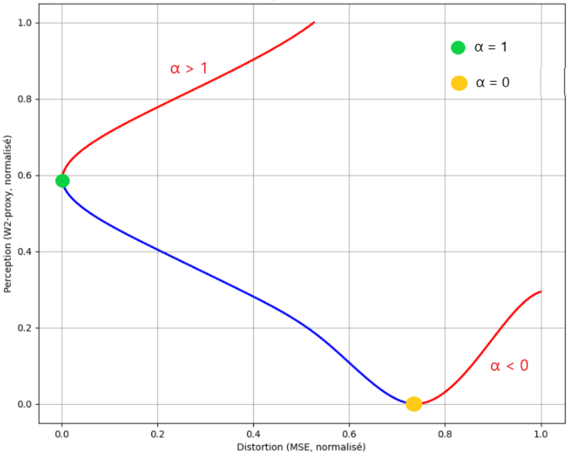
Image avec DOT (α = 1.0), MSE: 0.0015, PSNR: 76.48 dB



$\alpha = 1.3$


(α = 1.3)
Image originale restaurée (α = 1), MSE: 0.0002, PSNR: 85.34 dB
Image avec DOT (α = 1.3), MSE: 0.2442, PSNR: 54.25 dB

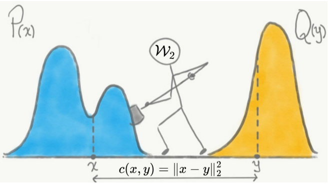
Perception-Distortion Curve



Thank you ! Questions ?

What is Optimal Transport ?

Optimal transport (OT) is a mathematical framework for finding the most efficient way to move one distribution of mass (probability, resources, ...) to another while minimizing a given cost.



Optimal Transport Problem

$$\min_{\pi \in \Pi(P,Q)} \int c(x,y) d\pi(x,y)$$

P, Q : probability distributions on X, Y
$\Pi(P,Q)$: set of couplings with marginals P, Q
$c(x,y)$: transport cost from x to y