Data Science Lab

Generative Adversarial Networks

Discriminator Optimal Transport

Group GANglions

Sebastiano SCARDERA — sebastiano.scardera@dauphine.eu
Constant BOURDREZ — constant.bourdrez@dauphine.eu
Mohamed Ali SRIR — mohamed-ali.srir@dauphine.eu

Submission Date : November 12, 2024

2024/2025 - Semester I

Contents

1 Discriminator optimal transport

Our work focuses on leveraging optimal transport methods to enhance image generation in GANs. Specifically, we implement the algorithms introduced in [3]. The main concept is to utilize the Discriminator D post-training to refine image generation. This approach aims to improve a generated image $\mathbf{y} = G(\mathbf{z})$ by directly adding a perturbation $\delta\mathbf{y}$ such that $D(\mathbf{y}+\delta\mathbf{y}) > D(\mathbf{y})$. Optimal transport theory ensures that the refined samples can be considered approximate samples from the target distribution p.

1.1 Optimal transport

In our framework, we have two probability distributions p and p_G defined on the image space. We consider the optimal cost to transport p_G to p the ℓ_2-Wasserstein distance:

$$W(p, p_G) = \min_{\pi \in \Pi(p,p_G)} \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\pi} \left[\|\mathbf{x}-\mathbf{y}\|_2\right] = \max_{\|\tilde{D}\|_L \leq 1} \mathbb{E}_{\mathbf{x}\sim p}\tilde{D}(\mathbf{x}) - \mathbb{E}_{\mathbf{y}\sim p_G}\tilde{D}(\mathbf{y}), \tag{1.1}$$

where $\Pi(p, p_G)$ is the set of all joint distributions with marginals p and p_G and $\tilde{D}$ is a 1-Lipschitz function. The second equality follows from the Kantorovich-Rubinstein duality. The following theorem plays a crucial role for the development of the algorithm.

Theorem 1.1. *Suppose π^* is a deterministic optimal transport plan between p_G and p, i.e. described by a deterministic function T, and D^* is the optimal solution of the dual problem. Then, it holds that $\|D^*\|_L = 1$ and*

$$T(\mathbf{y}) = argmin_{\mathbf{x}} \left\{\|\mathbf{x}-\mathbf{y}\| - D^*(\mathbf{x})\right\} \tag{1.2}$$

$$p(\mathbf{x}) = \int d\mathbf{y}\delta\left(\mathbf{x} - T(\mathbf{y})\right) p_G(\mathbf{y}). \tag{1.3}$$

In particular, T is the optimal transport map between p_G and p.

1.2 Optimal transport and GANs

Recall that D is the Discriminator of the GAN and let $K = \|D\|_L$ be the Lipschitz constant of D. The connection between OT and GAN's loss is given by the following theorem.

Theorem 1.2. *Let $\tilde{D} = D/K$ then*

$$\mathbb{E}_{\mathbf{x}\sim p}\left[\log D(\mathbf{x})\right] + \mathbb{E}_{\mathbf{y}\sim p_G}\left[\log(1 - D(\mathbf{y}))\right] \leq K\left(\mathbb{E}_{\mathbf{x}\sim p}\tilde{D}(\mathbf{x}) - \mathbb{E}_{\mathbf{y}\sim p_G}\tilde{D}(\mathbf{y})\right).$$

The last theorem implies that training the Discriminator maximizes the lower bound of the objective function in (1.1) in the dual form, so it can be regarded as approximately solving the problem described in (1.1). By exploiting this connection, we aim to recover the optimal transport map T to refine the distribution p_G to p.

For this purpose, we applied gradient descent on the target space and on the latent space to find the optimal transport map T.

To estimate the variation of the Lipschitz constant K, we used the following formulas:

$$K_{\text{eff}} = \max_{\mathbf{x},\mathbf{y}\sim p_G} \left\{\frac{|D(\mathbf{x}) - D(\mathbf{y})|}{\|\mathbf{x}-\mathbf{y}\|_2}\right\}, \quad k_{\text{eff}} = \max_{\mathbf{z},\mathbf{z}_y\sim p_Z} \left\{\frac{|D\circ G(\mathbf{z}) - D\circ G(\mathbf{z}_y)|}{\|\mathbf{z}-\mathbf{z}_y\|_2}\right\},$$

even if we observed better results by using the Lipschitz constant K_{eff} and k_{eff} equal to 1.

Starting from the generated image $\mathbf{y} = G(\mathbf{z}_y)$, we implemented the gradient descent algorithm on the image space and on the latent space to recover the argmin in (1.2). We used the Adam optimizer, and the corresponding pseudocodes can be found below.

Algorithm 1 Target space Discriminant Optimal Transport

Require: trained D, K_{eff}, sample y, stepsize ϵ

1: Initialize $x \leftarrow y$

2: **for** n_{trial} in range(N_{updates}) **do**

3: $\quad \mathbf{x} \leftarrow \mathbf{x} - \epsilon\nabla_x \left\{\|\mathbf{x}-\mathbf{y}+\delta\|_2 - \frac{1}{K_{\text{eff}}}D(\mathbf{x})\right\}$ ▷ δ to avoid overflow

4: **end for**

5: **return** x

Algorithm 2 Latent space Discriminant Optimal Transport

Require: trained D and G, k_{eff}, sample $\mathbf{z}_y$, stepsize ϵ

1: Initialize $\mathbf{z} \leftarrow \mathbf{z}_y$

2: **for** n_{trial} in range(N_{updates}) **do**

3: $\quad \mathbf{z} \leftarrow \mathbf{z} - \epsilon \nabla_z \left\{ \|G(\mathbf{z}) - G(\mathbf{z}_y) + \delta\|_2 - \frac{1}{k_{\text{eff}}} D(G(z)) \right\}$ ▷ δ to avoid overflow

4: **end for**

5: **return** $\mathbf{x} = G(\mathbf{z})$

1.3 Results on latent space

It is possible to view the optimal transport process during each iteration. This allows you to see the refinements that this method can propose:

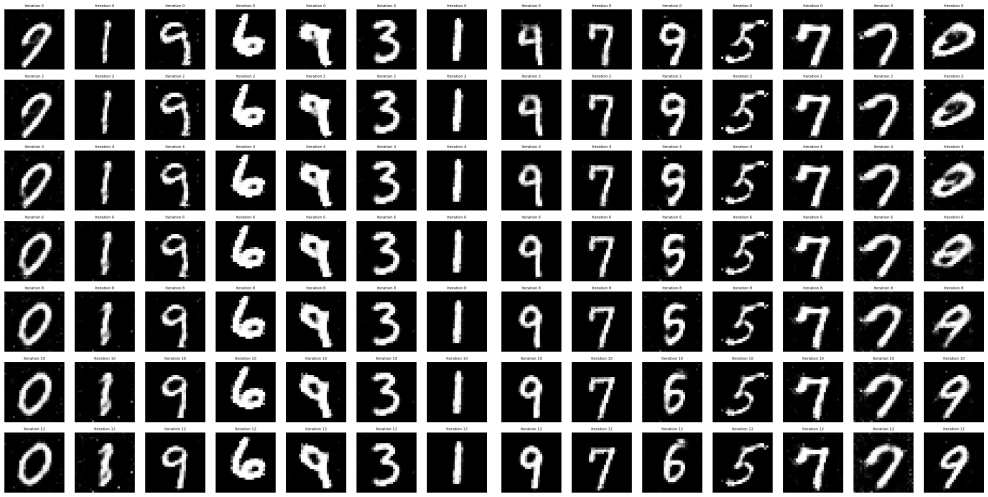


Figure 1: Evolution of the generated samples trough the optimal transport process

It is clear from this image that each number is becoming clearer and clearer, and that some numbers are being transformed into others to match the distribution learned by the discriminator. The next step was to fine-tune the OT parameters to get the best possible sample generation. We set the learning rate to **0.01**, which was the one used by Tanaka [3], and explored the influence of the number of iterations.

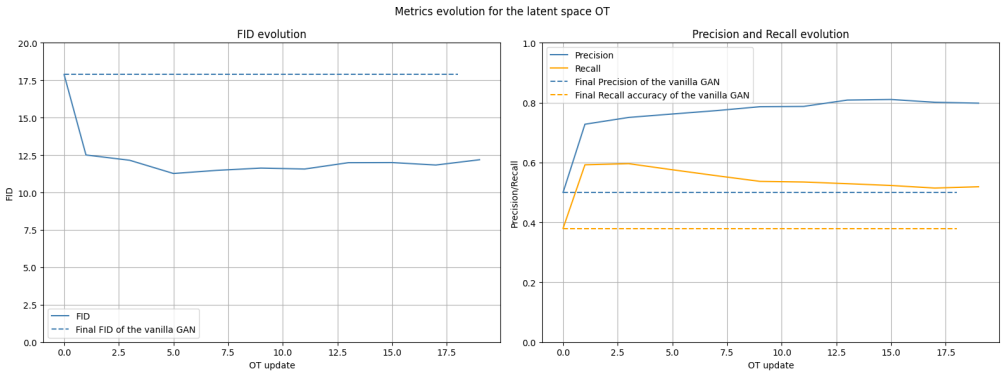


Figure 2: FID, Precision and Recall evolution during the OT process

The FID has a minimum about 5-6 iterations and then increases slightly, but the precision always increases, whereas the recall has a maximum of about 1 iteration. In order to get the best performance and to have both quality and diversity in our samples, we decided to choose 5 iterations.

1.4 Results on target space

The same method was then tried on the target space, where we modify directly in the image space. A strange phenomenon occurred during this experiment, where the numbers became blurred and replaced by others. Theoretically this should not

happen as we are supposed to stay close to the original image, but perhaps they were extremely close in the discriminator representation.

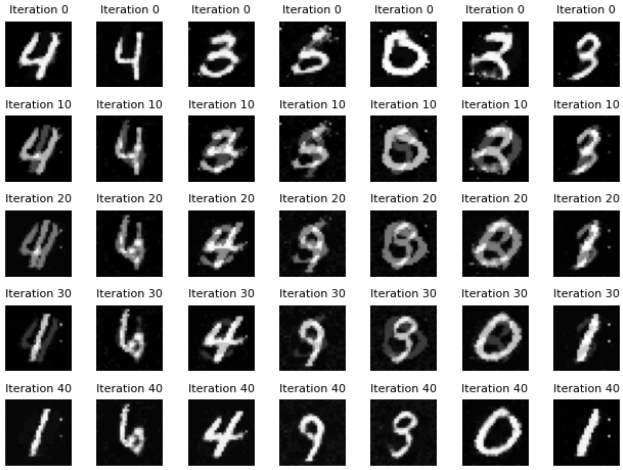


Figure 3: Optimal Transport process on target space

We also did the same tuning of the parameters. The curve does not look like the previous one because the image is degraded, each metric increases and then when the final number takes the place, they all decrease to a lower value than the vanilla GAN.

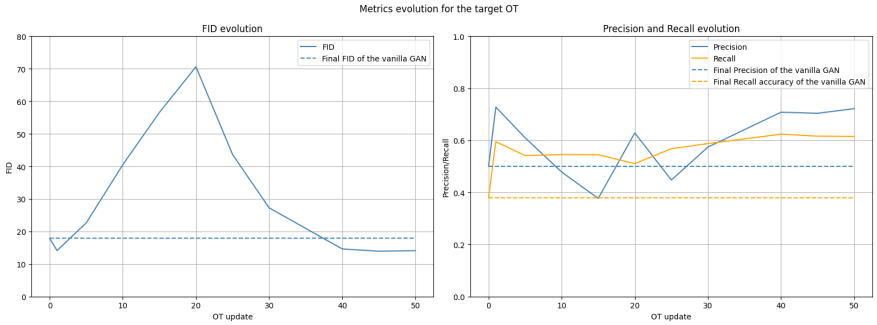


Figure 4: FID, Precision and Recall during the OT process on target space

1.5 Synthesis

Both methods improved every metrics with respect to the vanilla GAN, one can look at the final results below:

Table 1: Metrics comparison

Method	FID	Accuracy	Precision
Vanilla-GAN	49.87	0.15	0.47
Latent space OT	24.25	0.22	0.52
Target space OT	NA	NA	NA

2 Embedded OT method

2.1 Geometry of different representations

We explored how the geometry of the space change with different representation. In particular, we compared the original representation with ones given by VGG 5 and VGG 16 [2].

Different embeddings can significantly impact a method's effectiveness. For instance, when comparing the generated distribution to the real one, it becomes clear that the calculation of precision and recall is not invariant across representations.

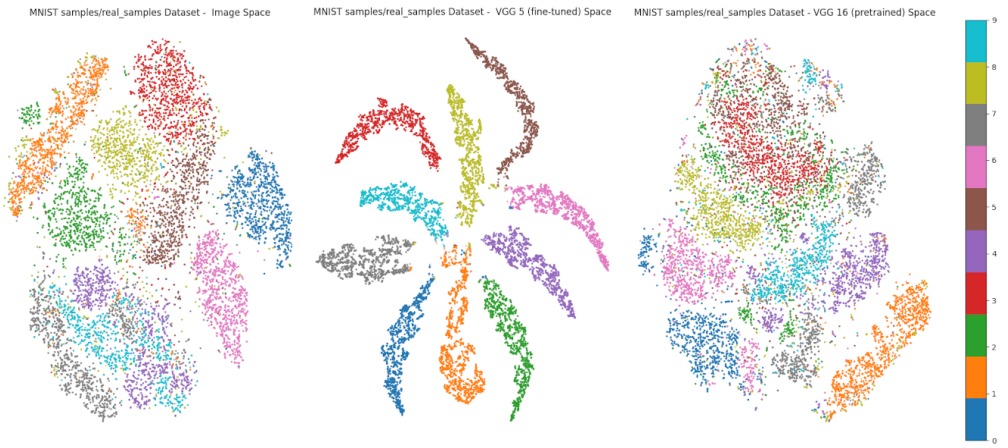


Figure 5: (Left to right) t-SNE [1] representation of MNIST dataset, VGG 5 embedding and VGG 16 embedding.

In the original space, we observe a precision of 0.78 and a recall of 0.28; however, in the VGG 5 space, precision rises to 0.84 and recall to 0.77. This shift suggests that relying on the VGG 5 representation might lead to an overly optimistic assessment of performance.

2.2 Embedded OT

As shown in the center image of Figure 5, the VGG 5 embedding reveals more distinctly separated clusters. Our idea is that using this representation could stabilize the DOT method, preventing the generated image from shifting clusters (and thus labels), phenomenon that occurs in the DOT on target space.

Denoting the VGG 5 embedding as E, we solve the following optimization problems:

Target space version:

$$\tilde{\mathbf{y}} = \text{argmin}_{\mathbf{x}} \left\{ \|E(\mathbf{x}) - E(\mathbf{y})\|_2 - D(\mathbf{x}) \right\}$$

Latent space version:

$$\tilde{\mathbf{z}} = \text{argmin}_{\mathbf{z}} \left\{ \|E(G(\mathbf{z})) - E(G(\mathbf{z}_y))\|_2 - D(G(\mathbf{z})) \right\}.$$

To tackle the problems above, we integrated the embedding in the DOT methods, getting the algorithms in Algorithm 3 and 4.

Algorithm 3 Target space Embedded Discriminant Optimal Transport

Require: trained D, K_{eff}, sample y, stepsize ϵ

1: Initialize $x \leftarrow y$
2: **for** n_{trial} in range(N_{updates}) **do**
3: $\quad \mathbf{x} \leftarrow \mathbf{x} - \epsilon \nabla_x \left\{ \|E(\mathbf{x}) - E(\mathbf{y}) + \delta\|_2 - \frac{1}{K_{\text{eff}}} D(\mathbf{x}) \right\}$ ▷ δ to avoid overflow
4: **end for**
5: **return** x

Algorithm 4 Latent space Embedded Discriminant Optimal Transport

Require: trained D and G, k_{eff}, sample $\mathbf{z}_y$, stepsize ϵ

1: Initialize $\mathbf{z} \leftarrow \mathbf{z}_y$
2: **for** n_{trial} in range(N_{updates}) **do**
3: $\quad \mathbf{z} \leftarrow \mathbf{z} - \epsilon \nabla_z \left\{ \|E(G(\mathbf{z})) - E(G(\mathbf{z}_y)) + \delta\|_2 - \frac{1}{k_{\text{eff}}} D(G(z)) \right\}$ ▷ δ to avoid overflow
4: **end for**
5: **return** $\mathbf{x} = G(\mathbf{z})$

2.3 Results

As we did in the original DOT methods we can view the optimal transport process during each iteration in Figures 6 and 7.

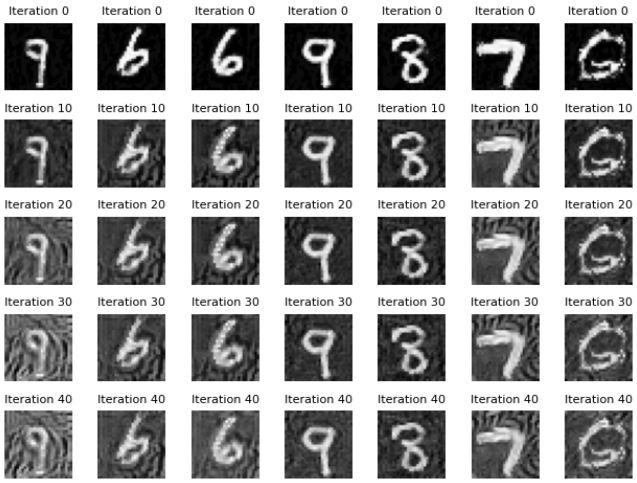


Figure 6: Target space Embedded DOT

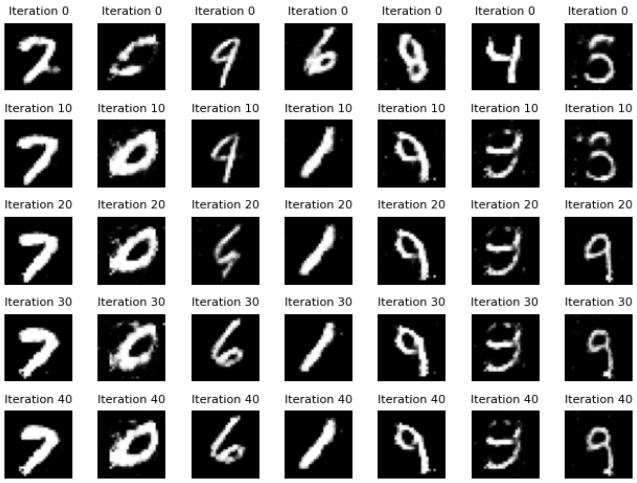
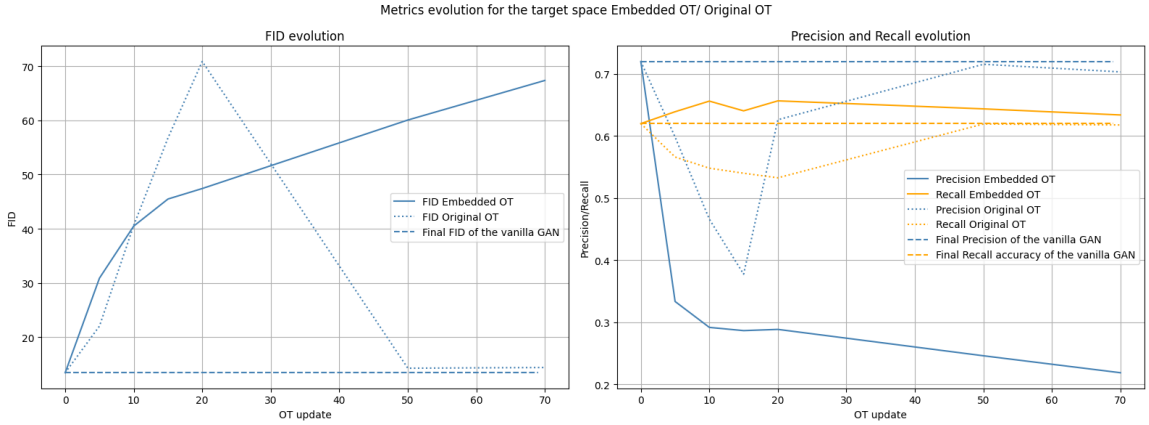
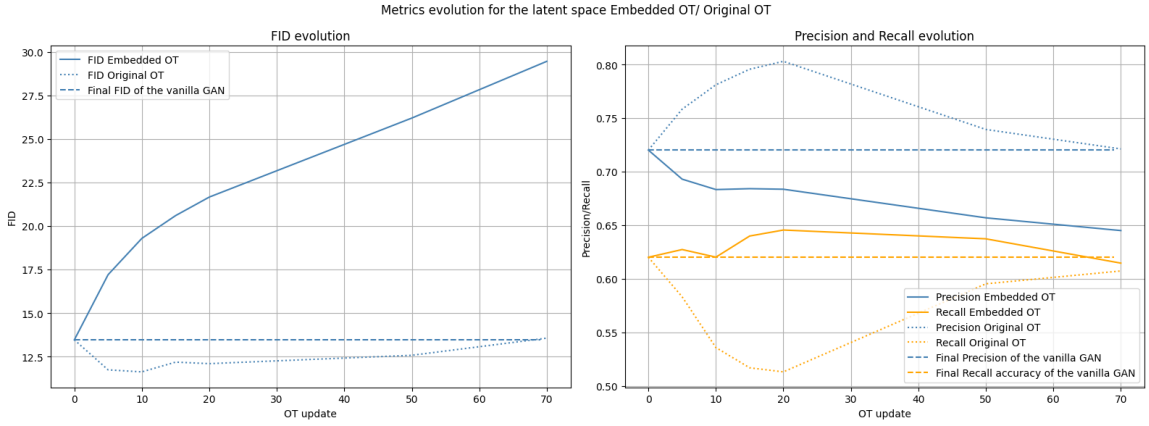


Figure 7: Latent space Embedded DOT

The reduction in overlapping within the target space is accompanied by a noticeable decline in image quality, as confirmed by the evolution of FID and precision metrics when compared with both the original DOT method and the Vanilla GAN. Notably, in both the target and latent versions, we observe improved recall with 20 steps. It is important to emphasize that these results are highly sensitive to the choice of embeddings. We believe that selecting optimal embeddings could further enhance model performance.





References

[1] Laurens van der Maaten and Geoffrey Hinton. "Visualizing Data using t-SNE". In: *Journal of Machine Learning Research* 9 (2008), pp. 2579–2605. URL: http://www.jmlr.org/papers/v9/vandermaaten08a.html.

[2] Karen Simonyan and Andrew Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2015. arXiv: 1409.1556 [cs.CV]. URL: https://arxiv.org/abs/1409.1556.

[3] Akinori Tanaka. "Discriminator optimal transport". In: *Advances in Neural Information Processing Systems* 32 (2019).