Learning latent space representations, and application to image generation

Aymeric Behaegel, Pierre Cornilleau, Luka Lafaye

Introduction

Introduced in 2014 by Ian Goodfellow et al., Generative Adversarial Networks (GANs) are a framework for training generative models through an adversarial process. This setup consists of two neural networks, a generator G and a discriminator D, that are trained simultaneously in a minimax game.

1 Baseline approach: vanilla GAN

1.1 Introductory settings

Given a real distribution P (of images from database MNIST here), we aim to generate samples x_g obeying a certain distribution $\hat{P}_g$ as (visually) close as possible to P.

In the original (logistic) approach, the generator G and the discriminator D are trained simultaneously to solve the following min-max problem:

$$\min_G \max_D \mathbb{E}_{x_r \sim P}[\log(D(x))] + \mathbb{E}_{x_g \sim \hat{P}_G}[\log(1 - D(x))],$$

where $x_g = G(z)$, $G : \mathcal{Z} \to \mathcal{X}$ is the generator function, and $\mathcal{Z}$ the latent space [1]

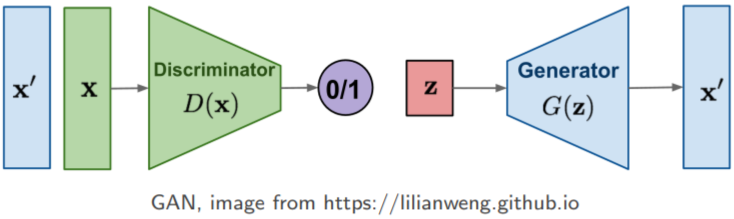

Figure 1: Generator and discriminator associated to a GAN

1.2 Results

1.2.1 First implementation

We first used a simple three-layer structure for the Discriminator, all being linear.

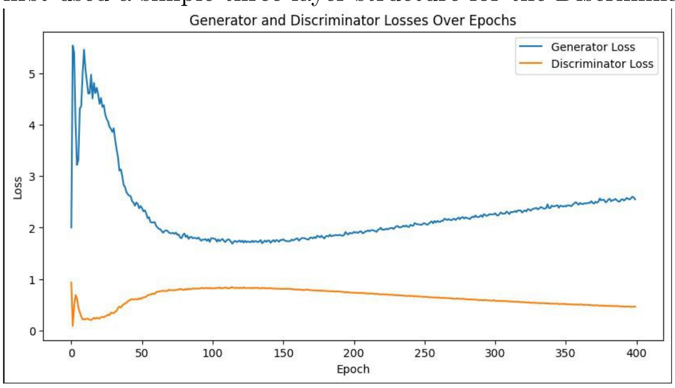

Figure 2: Generator and Discriminator Losses. Overfitting phenomena.

Figure 3: Some characters generated by the vanilla GAN. Lack of diversity.

[1] As a consequence, $\hat{P}_g$ is the pushforward of a measure on $\mathcal{Z}$.

1.2.2 Architecture of the Discriminator

As explicitly mentioned inside this project's guidelines, the architecture of the Generator will be fixed throughout this note. However, we would like to change the Generator architecture to get a better computation of the inner maximum.

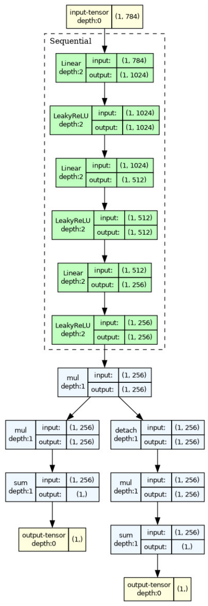

Figure 4: Chosen architecture for the Discriminator

After some work on its structure (see fig. 4), we obtained considerably better results, compared to the ones with the previous architecture.

Method	FID	Precision	Recall
1	44.12	0.56	0.18
2	22.08	0.40	0.20

Figure 5: Results of the different GAN methods

2 An intermediary model: Wasserstein GAN

2.1 Introduction

The *Earth-Mover* distance or Wasserstein-1 distance is defined

$$W_1\left(P_r, P_g\right) = \inf_{\gamma \in \Pi\left(P_r, P_g\right)} \mathbb{E}_{(x,y)\sim\gamma}[||x-y||],$$

where $\Pi\left(\mathbb{P}_r, \mathbb{P}_g\right)$ denotes the set of all joint distributions $\gamma(x, y)$ whose marginals are respectively $\mathbb{P}_r$ and $\mathbb{P}_g$. As the Fréchet Inception Distance (FID)/Wasserstein-2 distance, often used as a metric for GAN problems, it shares many properties. Defined as

$$W_2\left(P_r, P_g\right) = \inf_{\gamma \in \Pi\left(\mathbb{P}_r, \mathbb{P}_g\right)} \mathbb{E}_{(x,y)\sim\gamma}[||x-y||^2]^{1/2},$$

the FID is also a distance on certain probability measures on $\mathcal{X}$ [2], such that $W_1 \leq W_2$ [2].

Interestingly, the Kantorovich-Rubinstein duality tells us that

$$W_1\left(P_r, P_g\right) = \sup_{[f]_{\mathrm{Lip}} \leq 1} \left(\mathbb{E}_{x\sim Pr}[f(x)] - \mathbb{E}_{x\sim Pg}[f(x)]\right)$$

where the supremum is over all the 1-Lipschitz functions $f : \mathcal{X} \to \mathbb{R}$. Hence the minimization problem

$$\inf_{P_g} W_1(P_r, P_g)$$

becomes, similarly to the vanilla GAN, a minimax problem:

$$\inf_{P_g} W_1(P_r, P_g) = \inf_{P_g} \sup_{[f]_{\mathrm{Lip}} \leq 1} \left(\mathbb{E}_{x\sim Pr}[f(x)] - \mathbb{E}_{x\sim Pg}[f(x)]\right).$$

Apart from this adversarial structure (as $x = G(z)$, the infimum can actually be indexed by the generator G, whereas f is the Discriminator), this last problem also have some *metrisability* property : solving it ensures that the generator's distribution becomes closer to the target distribution in a measurable way.

[2] We also have an inequality of the form $W_2 \leq CW_1$ as long as $\mathcal{X}$ is a bounded set of some $\mathbb{R}^{D_x}$.

2.2 Results

Vanilla wGAN Similarly to the vanilla GAN algorithm, this first version shows poor FID results. We only reach a FID of 97, along with bad precision and recall figures.

Figure 6: Some characters generated by the vanilla wGAN

Loss As already presented last week, Hinge loss allows us to get better results. This problem amounts to solve

$$\inf_{P_g} \sup_{[f]_{\text{Lip}} \leq 1} \left(\mathbb{E}_{x \sim Pr}[\max(0, 1 + f(x))] + \mathbb{E}_{x \sim Pg}[\max(0, 1 - f(x))] \right),$$

meaning that the functions f are cut off inside the expectations. We reach a FID of 27 this time, and as low as 25 with the same improvement of the structure of the Discriminator as described in Section 1, for the vanilla GAN.

3 Sliced Adversarial Networks

3.1 Introduction

Following [3], we here present a novel approach to enhancing Generative Adversarial Networks (GANs) by introducing the Slicing Adversarial Network (SAN) framework. This framework aims to make GANs more metrizable, meaning that it helps ensure that the generator's distribution becomes closer to the target distribution in a measurable way (as for Wasserstein GAN), without requiring the ideal discriminator.

3.2 Theoretical points

SAN-ify your GANs for further improvement!

Simple modifications to discriminators bring us metrizable discriminator

SAN-ify: Converting GAN to Slicing Adversarial Network (SAN) ✓ Discriminator architecture ✓ Discriminator objective (maximization problem) $\min_\theta \mathcal{J}_{\text{GAN}}(\theta; f)$ and $\max_{f \in \mathcal{F}(X,\mathbb{R})} \mathcal{V}_{\text{GAN}}(f; \theta) \gtrsim$

Motivation (our questions) Optimal discriminator

$$\begin{aligned} &\min_\theta \mathcal{D}\left(\mu_0, \mu_\theta\right) \qquad \text{Too hard in practice!} \\ \iff &\min_\theta \mathcal{J}_{\text{GAN}}(\theta; \hat{f}) \text{ with } \hat{f} = \underset{f \in \mathcal{F}(X,\mathbb{R})}{\text{argm}} \mathcal{V}_{\text{GAN}}(f; \theta) \end{aligned}$$

Discriminator is not guaranteed to make the two distributions closer without optimality assumption, which is too strong to impose

Relax the condition Metrizable discriminator

$$\begin{aligned} &\min_\theta \tilde{\mathcal{D}}_f\left(\mu_0, \mu_\theta\right) \qquad \text{(Inducing a certain distance)} \\ \iff &\min_\theta \mathcal{J}_{\text{GAN}}(\theta; f) \qquad \text{with } f \overset{?}{\in} \mathcal{F}_{\text{Met}}\left(\mu_0, \mu_\theta\right) \end{aligned}$$

Questions 1. What are sufficient conditions for "metrizability"? 2. How to induce "metrizability" of discriminators?

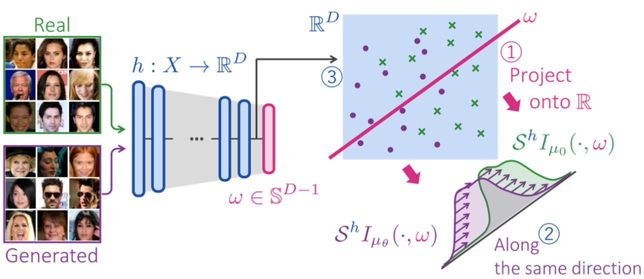

Figure 7: Illustration of direction optimality, separability and injectivity properties

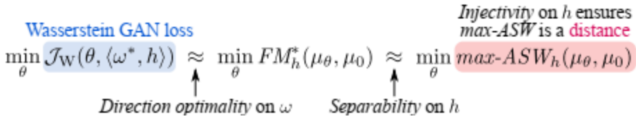

Figure 8: Under the previous properties, the Wassertein GAN loss is metrizable

Answer to Question 1 Metrizable conditions Under the decomposition of a discriminator into Normalized last linear layer and Neural function

Intuitive visualization

Transform, Slice, then Move samples in $\mathbb{R}$ with minimization problem!

Answer to Question 2 Method: SAN-ify Enforcing direction optimality on the last layer $\boldsymbol{\omega}$ by applying Wasserstein-ish loss to it Any GAN maximization loss do NOT induce all the condition simultaneously

3.3 Definition of the loss

TODO

$$\mathcal{V}^{\mathrm{SAN}}(\omega, h; \mu_\theta) := \underbrace{\mathcal{V}\left(\left\langle \omega^-, h \right\rangle; \mu_\theta\right)}_{\mathcal{L}^h(h;\omega,\mu_\theta)} + \lambda \cdot \underbrace{d_{(\omega, h^-}\left(\tilde{\mu}_0^{J_J \circ f}, \tilde{\mu}_\theta^{r\mathcal{J} \circ f}\right)}_{\mathcal{L}^\omega(\omega;h,\mu\theta)}$$

3.4 Results

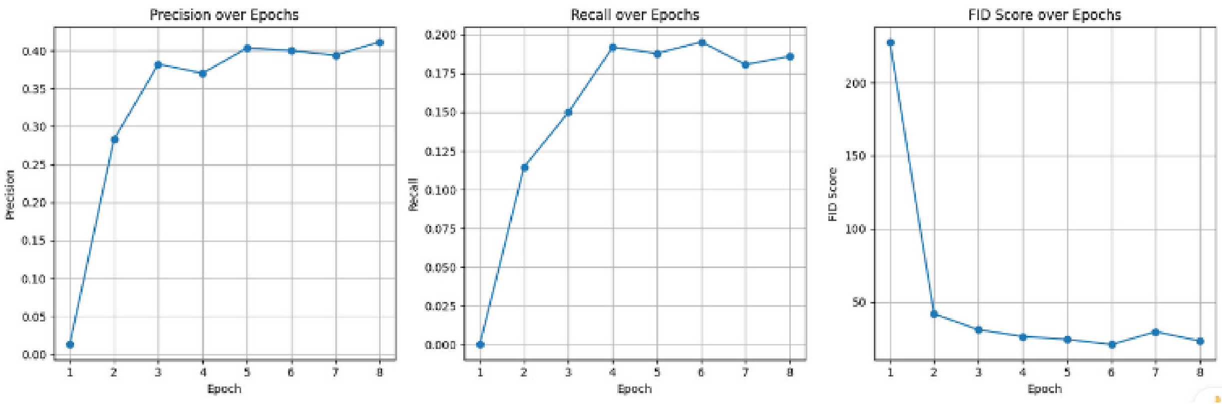

Figure 9: With a batch size of 64, we reach a FID of 19.04

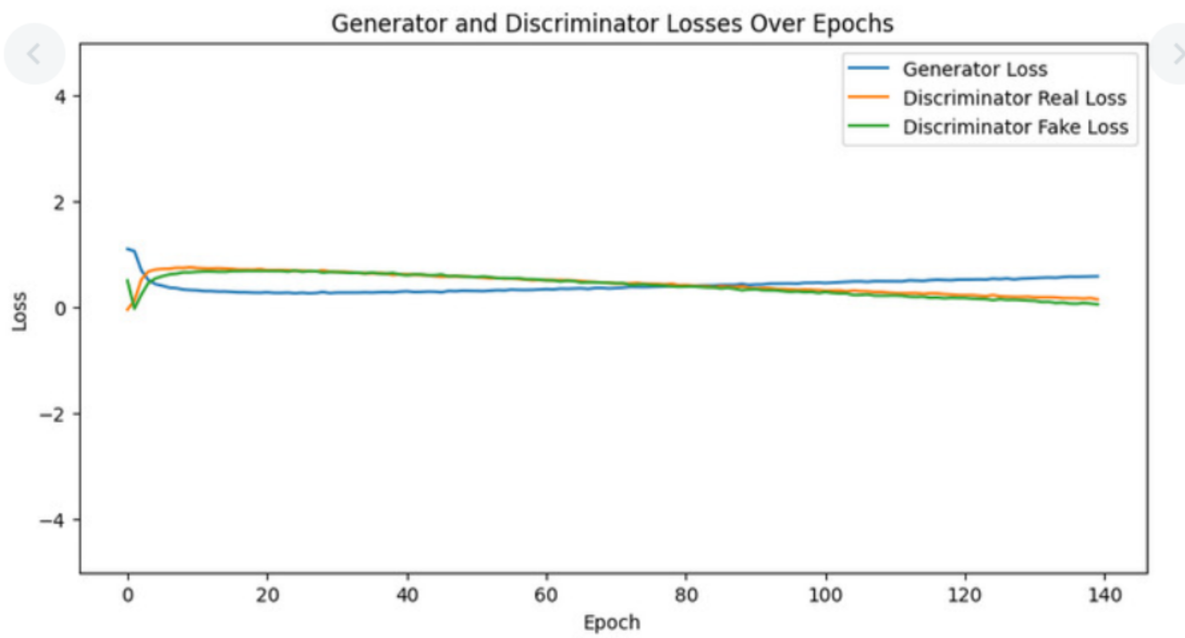

Figure 10: Generator and Discriminator Losses with a batch size of 64.

Conclusion

This research addresses fundamental limitations of traditional GANs by ensuring that the discriminator more effectively guides the generator towards the target distribution. The SAN framework provides a promising, theoretically motivated, and empirically supported method for improving GAN performance across a range of tasks.

References

[1] Arjovsky M., Chintalah S., Bottou L.. *Wasserstein GAN*, https://arxiv.org/abs/1701.07875, 2017.

[2] Santambrogio F.. *The Wassertein distances*, https://math.univ-lyon1.fr/~santambrogio/Wp.pdf, 2011.

[3] Takida Y., Imaizumi M., Shibuya T., Lai C.-H., Uesaka T., Murata N. and Mitsufuji Y.. *SAN: Inducing Metrizability of GAN with Discriminative Normalized Linear Layer*, International Conference on Learning Representations 2024, https://iclr.cc/virtual/2024/poster/18212.