Learning latent space representations and application to image generation

Aymeric Behaegel, Pierre Cornilleau,
Luka Lafaye de Micheaux

WGAN

Algorithm 1 WGAN, our proposed algorithm. All experiments in the paper used the default values $\alpha = 0.00005$, $c = 0.01$, $m = 64$, $n_{\text{critic}} = 5$.

Require: : α, the learning rate. c, the clipping parameter. m, the batch size. n_{critic}, the number of iterations of the critic per generator iteration.

Require: : w_0, initial critic parameters. θ_0, initial generator's parameters.

while θ has not converged **do**

  for $t = 0, ..., n_{\text{critic}}$ **do**

    Sample $\{x^{(i)}\}_{i=1}^m \sim \mathbb{P}_r$ a batch from the real data.

    Sample $\{z^{(i)}\}_{i=1}^m \sim p(z)$ a batch of prior samples.

    $g_w \leftarrow \nabla_w \left[\frac{1}{m}\sum_{i=1}^m f_w(x^{(i)}) - \frac{1}{m}\sum_{i=1}^m f_w(g_\theta(z^{(i)}))\right]$

    $w \leftarrow w + \alpha \cdot \text{RMSProp}(w, g_w)$

    $w \leftarrow \text{clip}(w, -c, c)$

  end for

  Sample $\{z^{(i)}\}_{i=1}^m \sim p(z)$ a batch of prior samples.

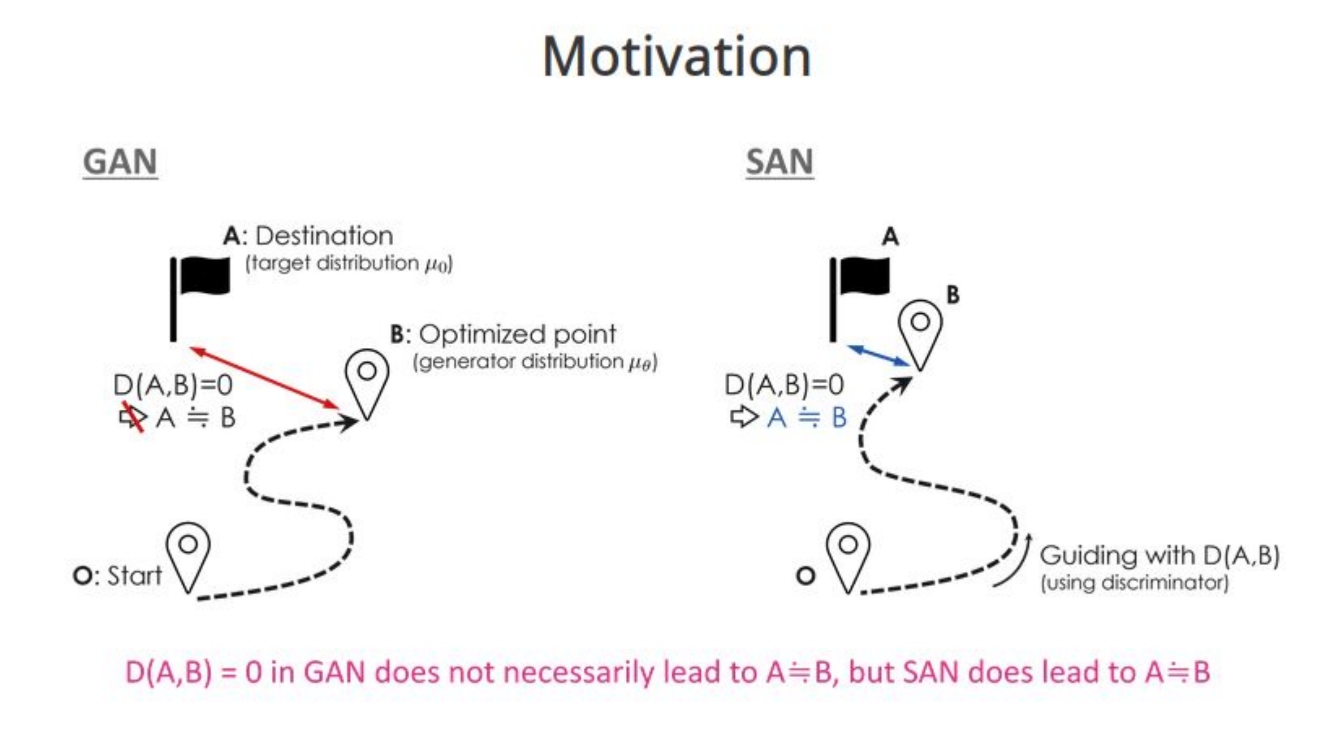
  $g_\theta \leftarrow -\nabla_\theta \frac{1}{m}\sum_{i=1}^m f_w(g_\theta(z^{(i)}))$

  $\theta \leftarrow \theta - \alpha \cdot \text{RMSProp}(\theta, g_\theta)$

end while

Vanilla GAN… WGAN… What now?

Why SAN?



Motivation
GAN
SAN
A: Destination
(target distribution μ_0)
B: Optimized point
(generator distribution μ_θ)
D(A,B)=0
A ≒ B
O: Start
A
B
D(A,B)=0
A ≒ B
O
Guiding with D(A,B)
(using discriminator)
D(A,B) = 0 in GAN does not necessarily lead to A≒B, but SAN does lead to A≒B

What is SAN?

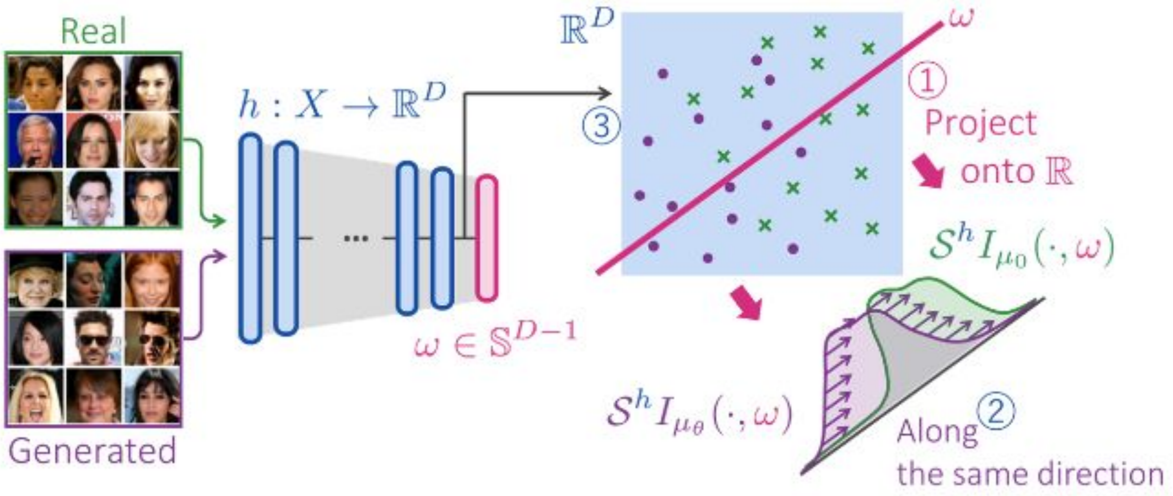


Figure 5: Illustration of direction optimality, separability and injectivity properties

Metrizability, maximum augmented sliced Wasserstein divergence

$$\min_{\theta \in \mathbb{R}^{D_\theta}} \mathcal{J}_{\mathrm{W}}(\theta; f) := -\mathbb{E}_{x \sim \mu_\theta}[f(x)]$$

$$\min_{\theta} \underbrace{\mathcal{J}_{\mathrm{W}}(\theta, \langle \omega^*, h \rangle)}_{\text{Wasserstein GAN loss}} \approx \min_{\theta} \underbrace{\textit{max-ASW}_h(\mu_\theta, \mu_0)}_{\substack{\textit{Injectivity} \text{ on } h \text{ ensures} \\ \textit{max-ASW} \text{ is a distance}}}$$

The metrizable conditions (direction optimality, separability, and injectivity) ensure that Wasserstein GAN loss evaluates the distance between data and generator distributions.

Gan to SAN

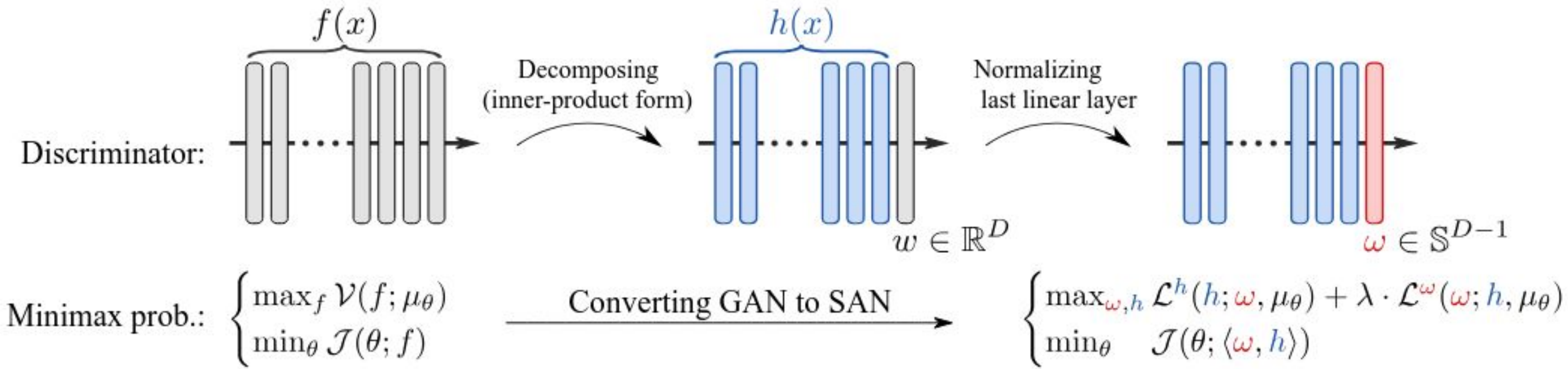
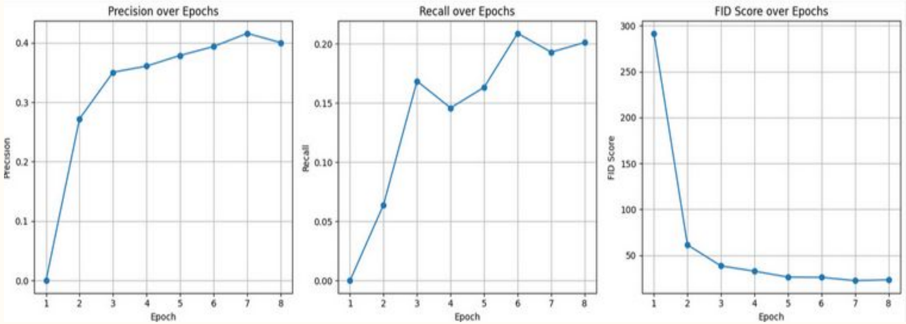


Figure 4: Converting GAN to SAN requires only simple modifications to discriminators.

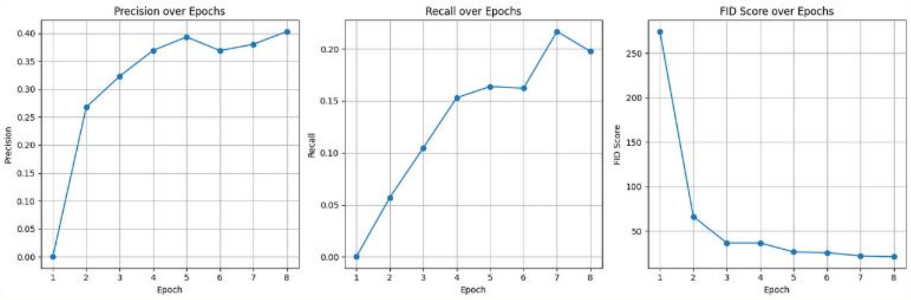
Results

Choice of λ=1 (SAN with BS=128)

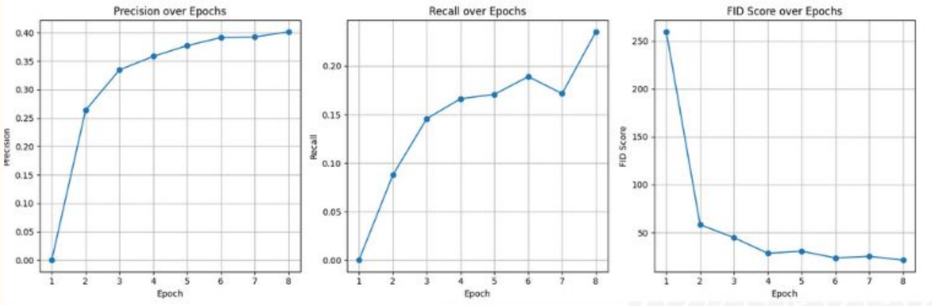
λ=1 → Final FID 20.54, P=40.0%, R=20.1%



λ=5 → Final FID 20.91, P=40.0%, R=19.7%

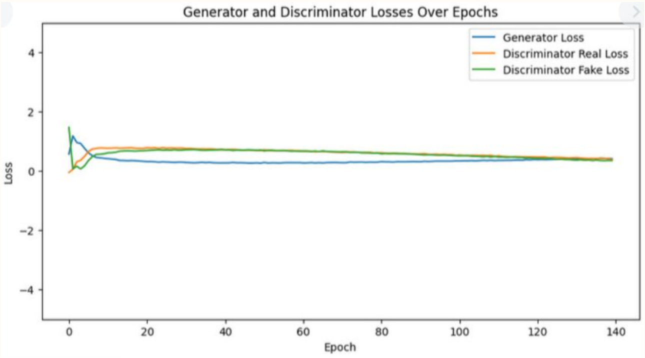


λ=20 → Final FID 21.29, P=40%, R=23.5%

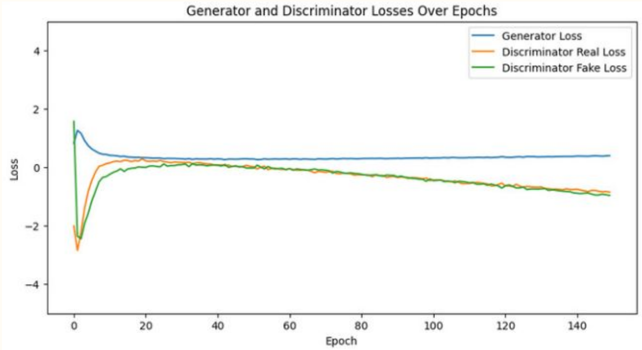


Choice of λ=1 (SAN with BS=128)

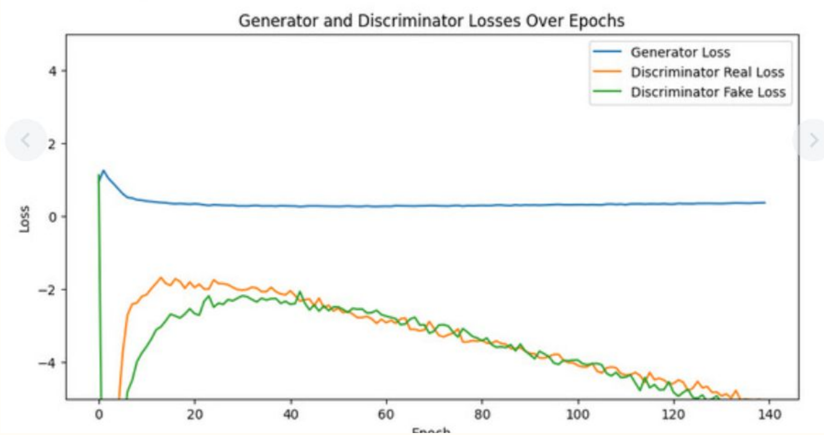
λ=1 → Final FID 20.54, P=40.0%, R=20.1%



λ=5 → Final FID 20.91, P=40.0%, R=19.7%

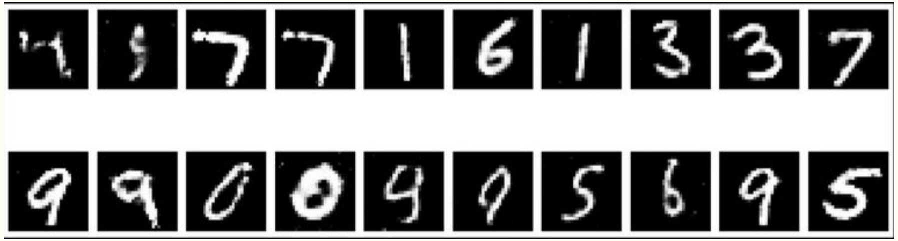


λ=20 → Final FID 21.29, P=40%, R=23.5%



Choice of λ=1 (SAN with BS=128)

λ=1 → Final FID 20.54, P=40.0%, R=20.1%

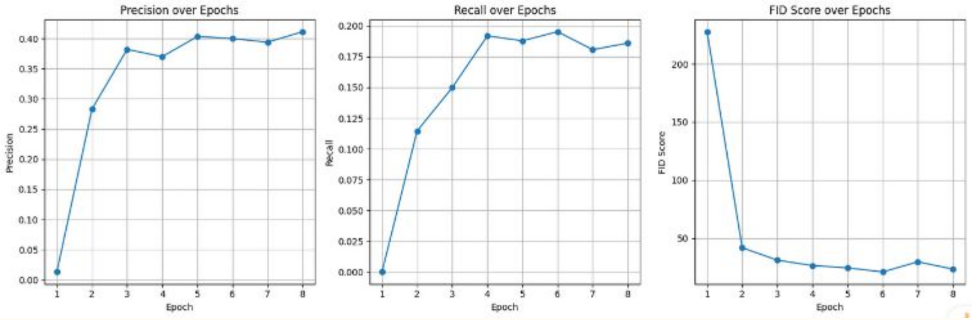


λ=5 → Final FID 20.91, P=40.0%, R=19.7%

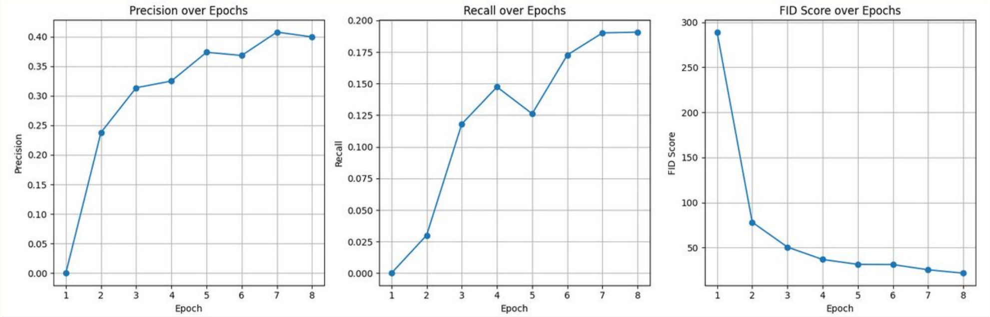
λ=20 → Final FID 21.29

Choice of BS=64 (SAN with λ=1)

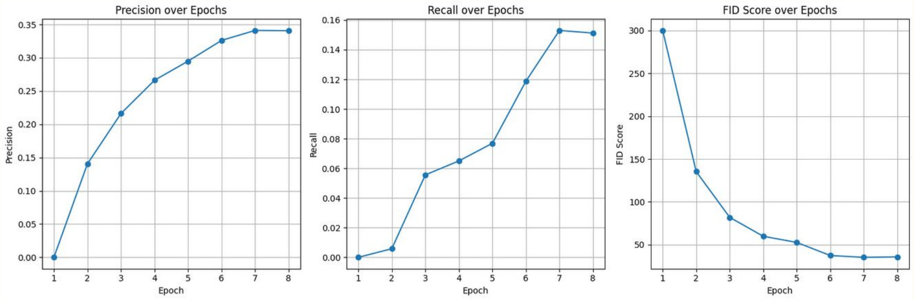
BS=64 → Final FID 19.04, P=41.1%, R=18.6%



BS=256 → Final FID 24.97, P=39.9%, R=19.0%

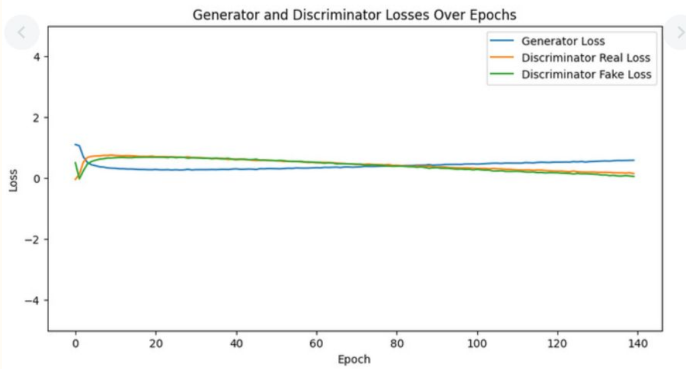


BS=512 → Final FID 36.19, P=34.0%, R=15.1%

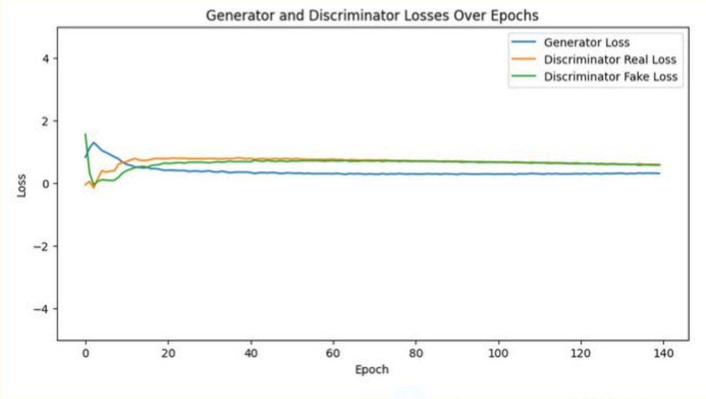


Choice of BS=64 (SAN with λ=1)

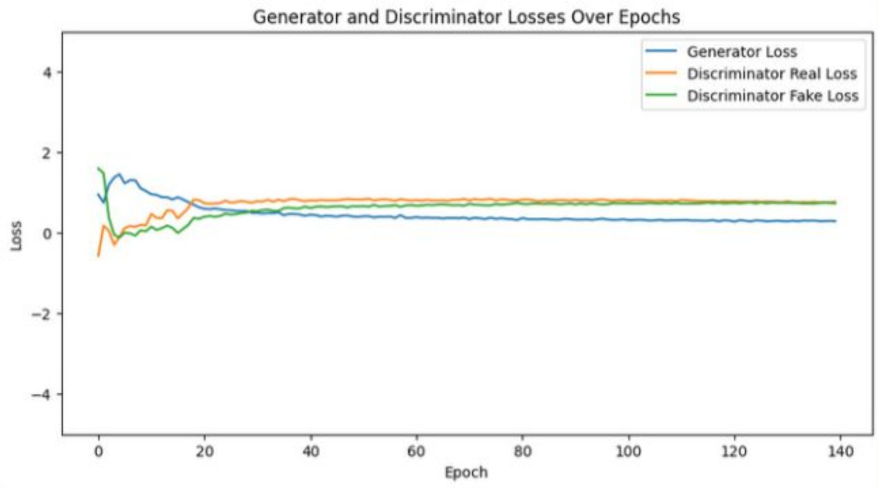
BS=64 → Final FID 19.04, P=41.1%, R=18.6%



BS=256 → Final FID 24.97, P=39.9%, R=19.0%



BS=512 → Final FID 36.19, P=34.0%, R=15.1%



Choice of BS=64 (SAN with λ=1)

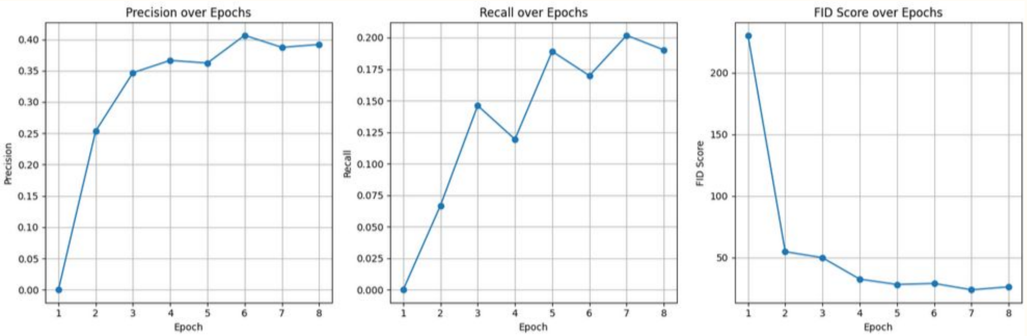
BS=64 → Final FID 19.04, P=41.1%, R=18.6%

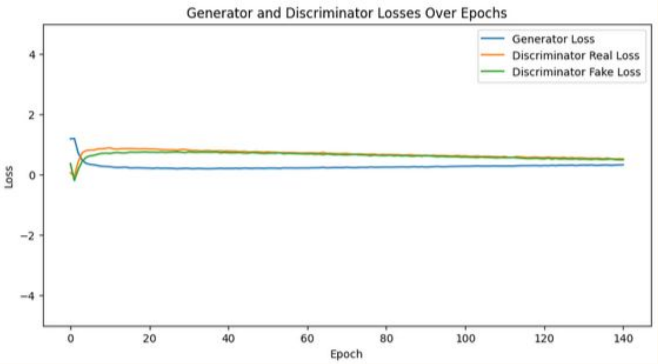
BS=256 → Final FID 24.97, P=39.9%, R=19.0%

BS=512 → Final FID 36.19, P=34.0%, R=15.1%

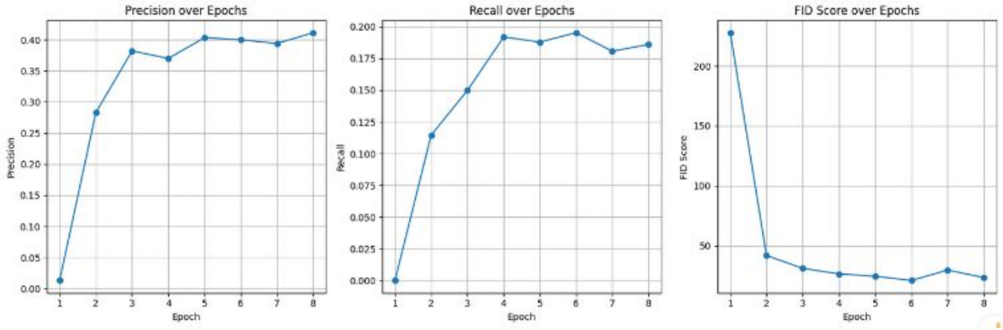
Choice of no dropout (SAN with λ=1, BS=64)

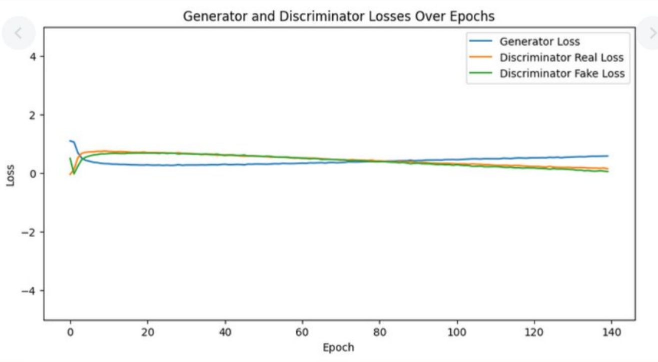
Dropout → Final FID 36.64





No dropout→ Final FID 19.04





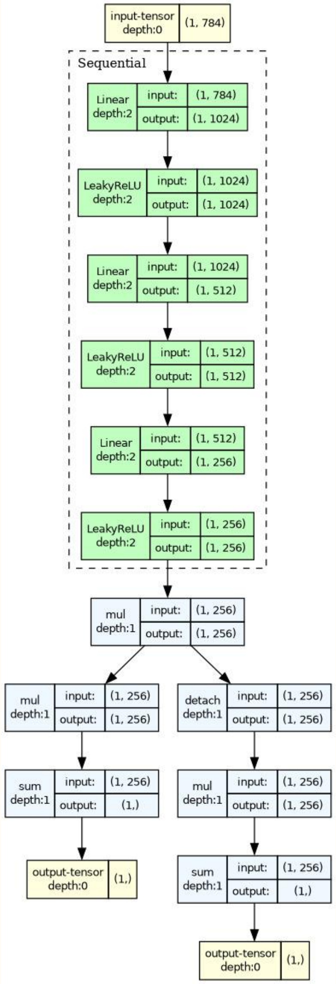
Final optimal SAN training

- No Dropout
- Batch size 64
- Lambda 1

→ FID = 19.04

\+ Discriminator rejection sampling (@gangineers)

→ Final Fid ≃ 17



Thanks for listening!
Questions?