Learning latent space representations and application to image generation

Aymeric Behaegel, Pierre Cornilleau,
Luka Lafaye de Micheaux

Vanilla GAN… WGAN… What now?

What is SAN?

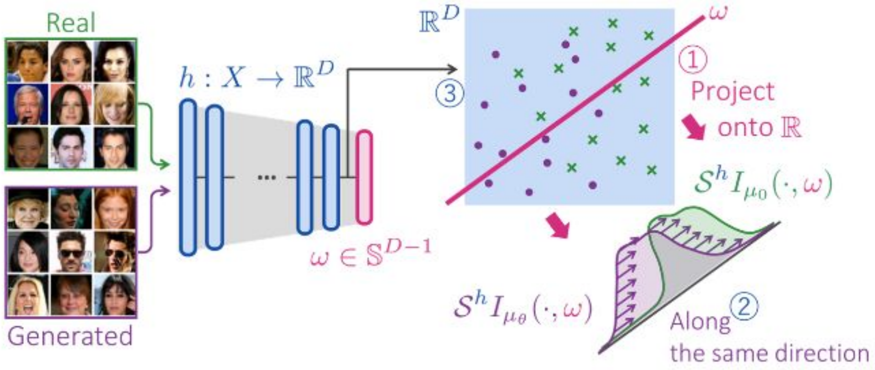
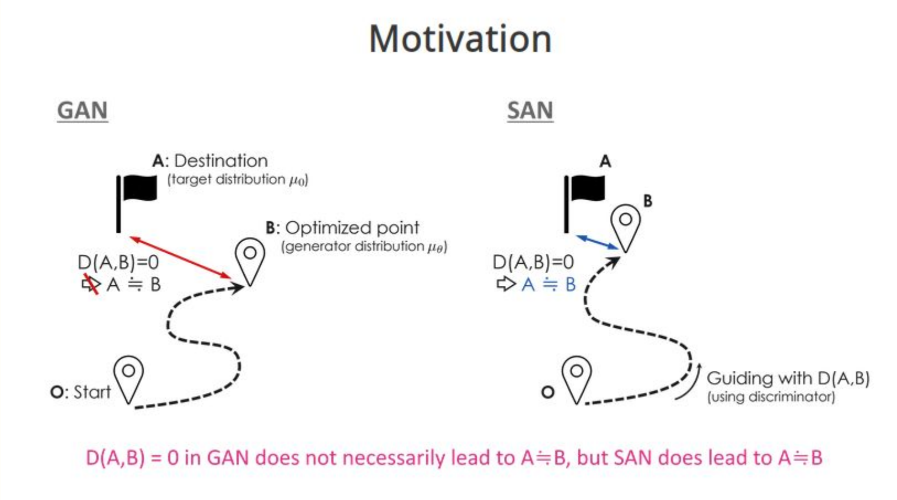


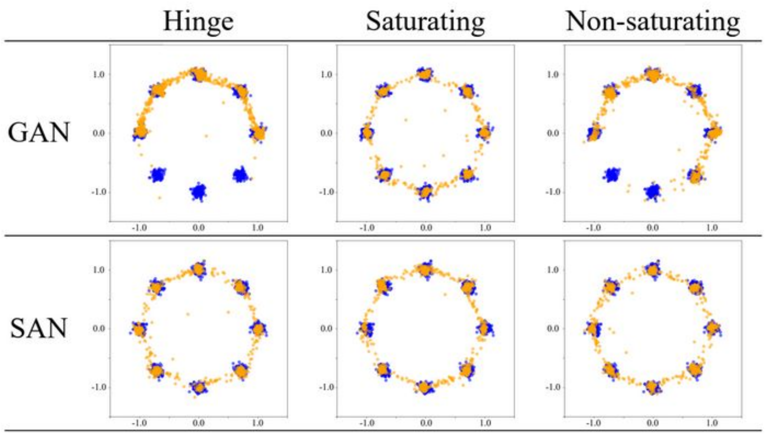
Figure 5: Illustration of direction optimality, separability and injectivity properties

$$\min_{\theta} \overbrace{\mathcal{J}_{\mathrm{W}}(\theta, \langle \omega^*, h \rangle)}^{\text{Wasserstein GAN loss}} \underset{\substack{\uparrow \\ \textit{Direction optimality on } \omega}}{\approx} \min_{\theta} FM_h^*(\mu_\theta, \mu_0) \underset{\substack{\uparrow \\ \textit{Separability on } h}}{\approx} \min_{\theta} \overbrace{\textit{max-}ASW_h(\mu_\theta, \mu_0)}^{\substack{\textit{Injectivity on } h \text{ ensures} \\ \textit{max-ASW} \text{ is a distance}}}$$

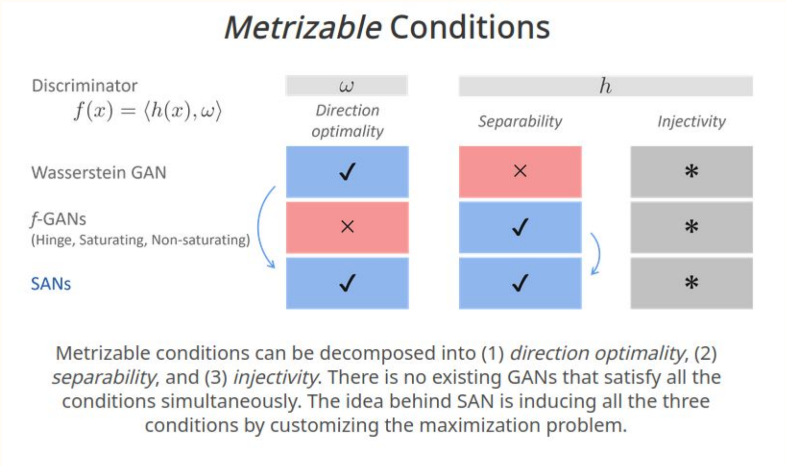
Figure 6: Under the previous properties, the Wassertein GAN loss is metrizable

Why SAN?





Comparison of the learned distributions between GAN and SAN with various objectives. In all cases, SANs cover all modes whereas mode collapse occurs in some GAN cases.



Metrizable Conditions

Discriminator $f(x) = \langle h(x), \omega \rangle$	ω: *Direction optimality*	h: *Separability*	h: *Injectivity*
Wasserstein GAN	✓	×	*
f-GANs (Hinge, Saturating, Non-saturating)	×	✓	*
SANs	✓	✓	*

Metrizable conditions can be decomposed into (1) *direction optimality*, (2) *separability*, and (3) *injectivity*. There is no existing GANs that satisfy all the conditions simultaneously. The idea behind SAN is inducing all the three conditions by customizing the maximization problem.

Gan to SAN

For generative modeling in GAN, we introduce the notion of a discriminator $f \in \mathcal{F}(X) \subset L^\infty(X, \mathbb{R})$. We formulate the GAN's optimization problem as a two-player game between the generator and discriminator with $\mathcal{V} : \mathcal{F}(X) \times \mathcal{P}(X) \to \mathbb{R}$ and $\mathcal{J} : \mathbb{R}^{D_\theta} \times \mathcal{F}(X) \to \mathbb{R}$, as follows:

$$\max_{f \in \mathcal{F}(X)} \mathcal{V}(f; \mu_\theta) \quad \text{and} \quad \min_{\theta \in \mathbb{R}^{D_\theta}} \mathcal{J}(\theta; f). \tag{1}$$

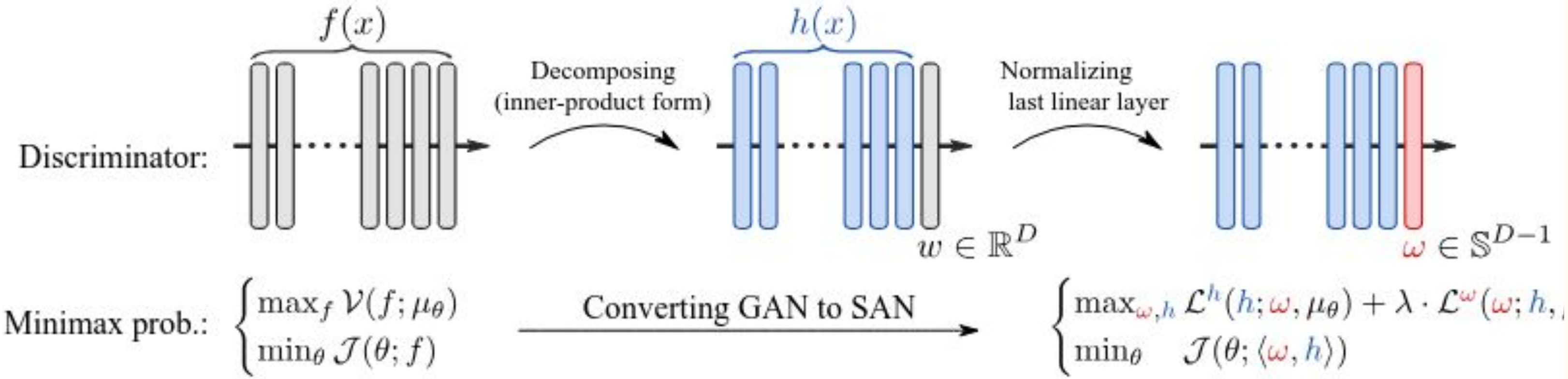
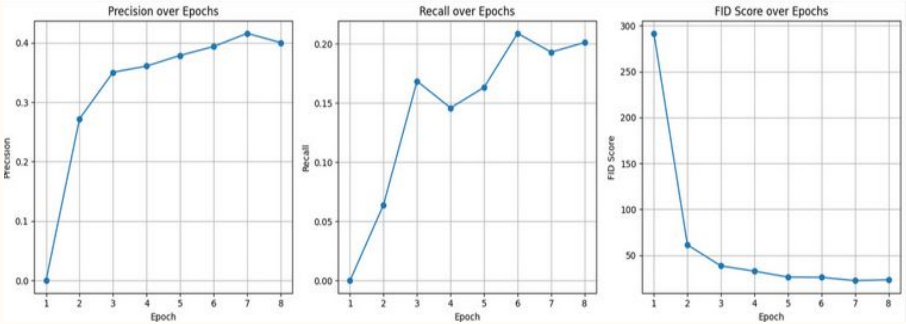


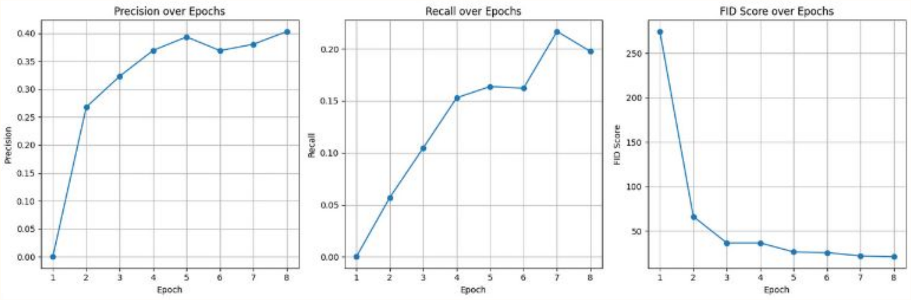
Figure 4: Converting GAN to SAN requires only simple modifications to discriminators.

Choice of λ=1 (SAN with BS=128)

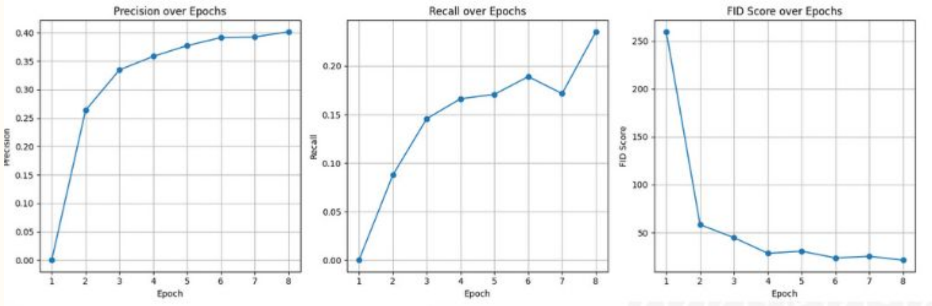
λ=1 → Final FID 20.54



λ=5 → Final FID 20.91

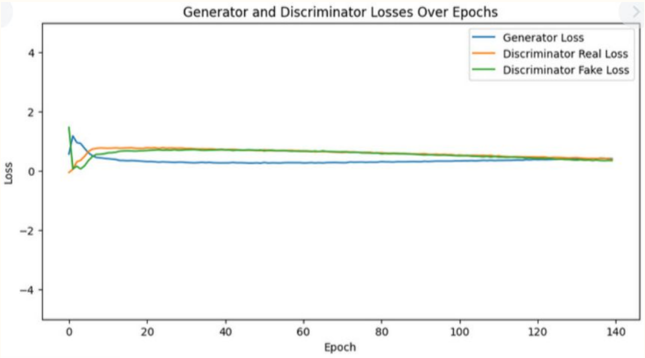


λ=20 → Final FID 21.29

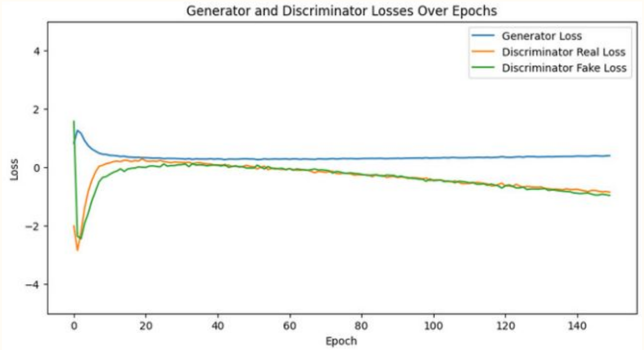


Choice of λ=1 (SAN with BS=128)

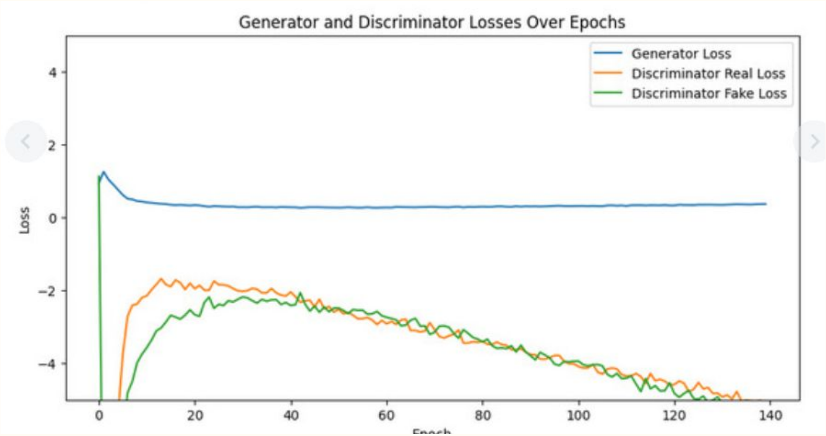
λ=1 → Final FID 20.54



λ=5 → Final FID 20.91

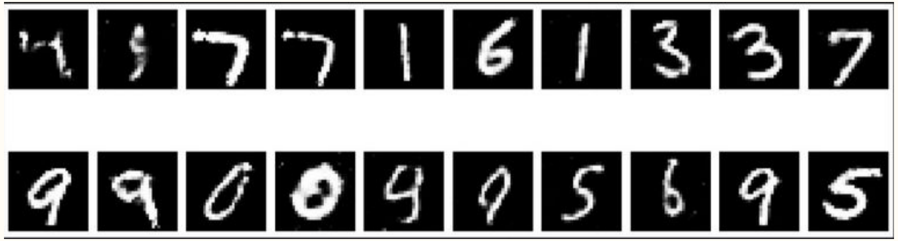


λ=20 → Final FID 21.29



Choice of λ=1 (SAN with BS=128)

λ=1 → Final FID 20.54

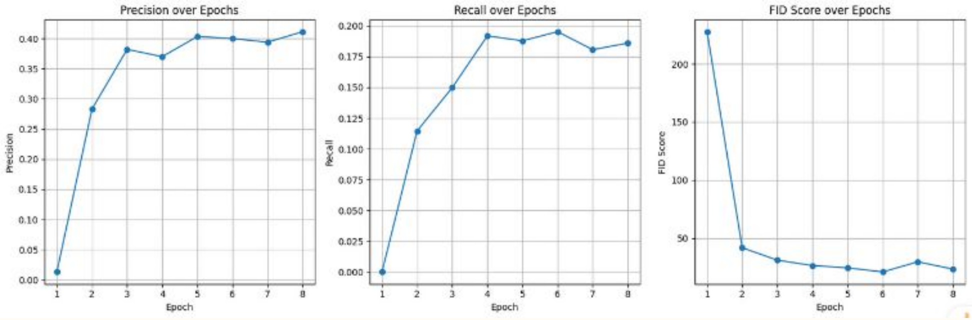


λ=20 → Final FID 21.29

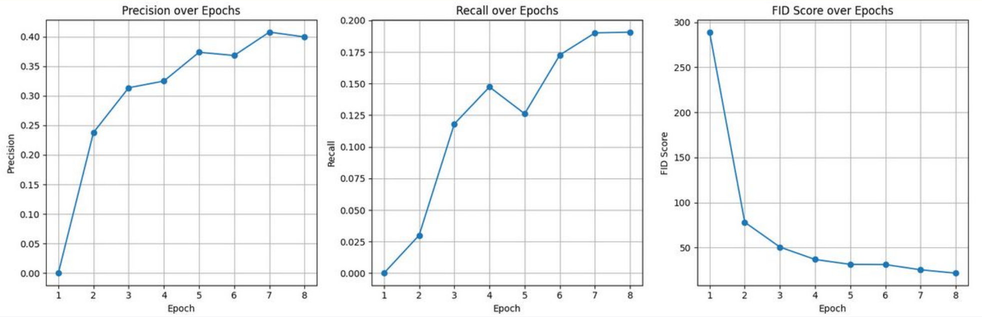
λ=5 → Final FID 20.91

Choice of BS=64 (SAN with λ=1)

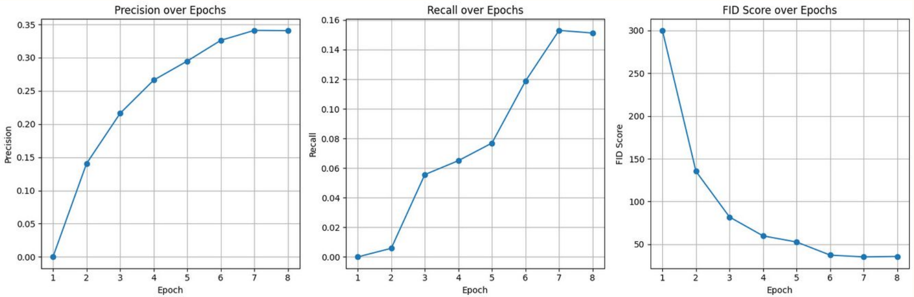
BS=64 → Final FID 19.04



BS=256 → Final FID 24.97

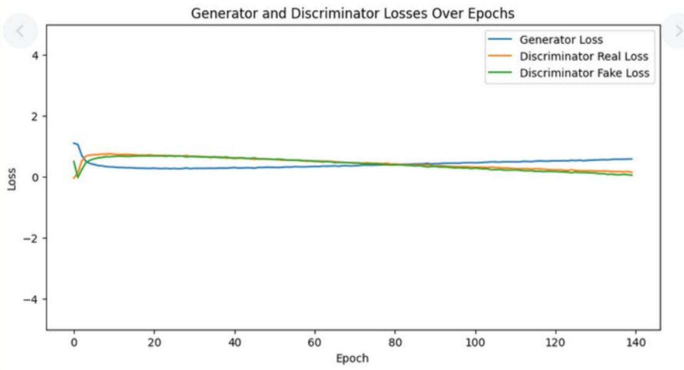


BS=512 → Final FID 36.19

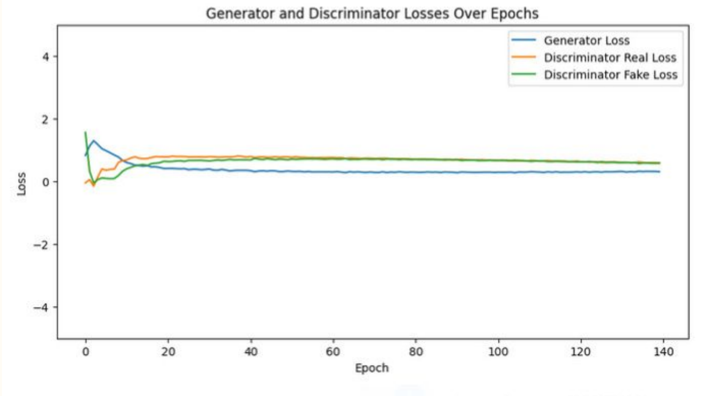


Choice of BS=64 (SAN with λ=1)

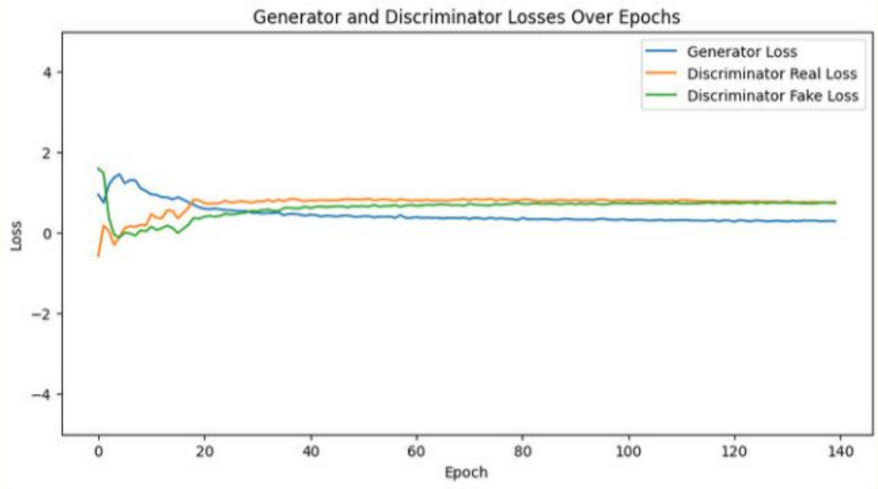
BS=64 → Final FID 19.04



BS=256 → Final FID 24.97



BS=512 → Final FID 36.19



Choice of BS=64 (SAN with λ=1)

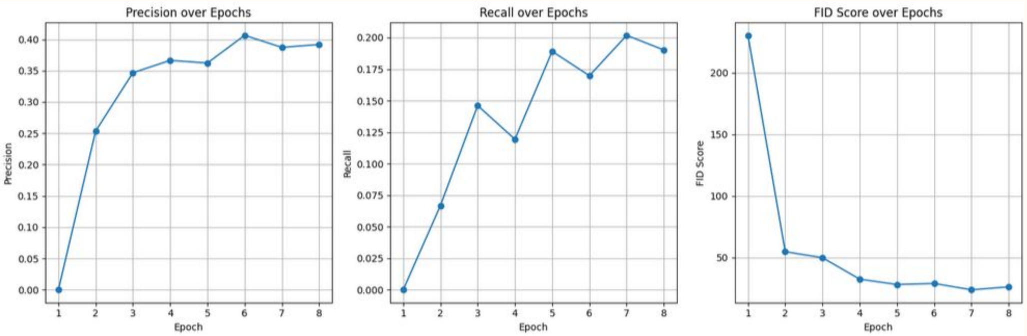
BS=64 → Final FID 19.04

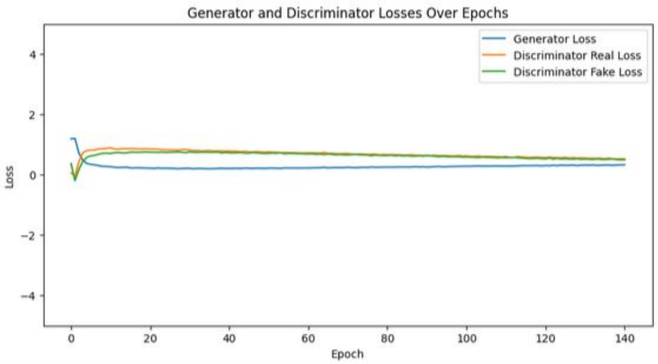
BS=512 → Final FID 36.19

BS=256 → Final FID 24.97

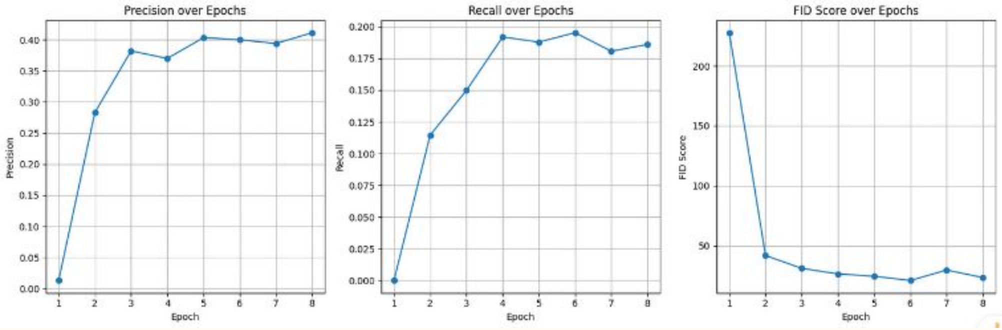
Choice of no dropout (SAN with λ=1, BS=64)

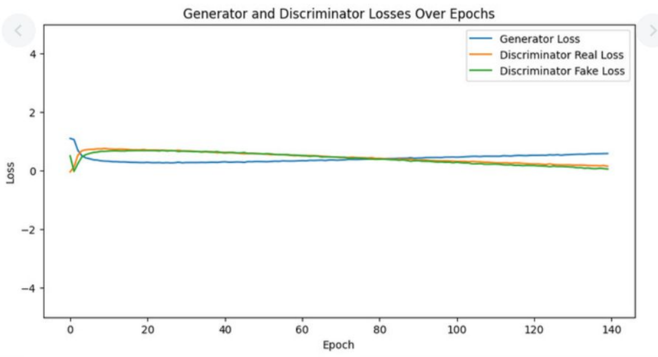
Dropout → Final FID 36.64





No dropout→ Final FID 19.04





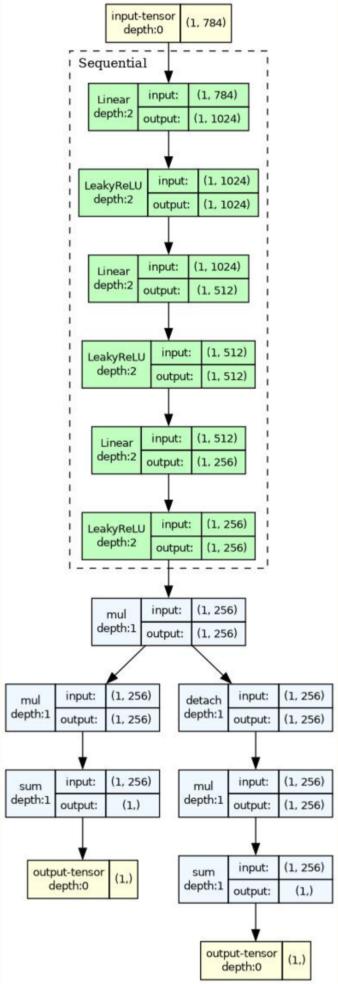
Final optimal SAN training

- No Dropout
- Batch size 64
- Lambda 1

→ FID = 19.04

\+ Discriminator rejection sampling (@gangineers)

→ Final Fid ≃ 17



Thanks for listening!
Questions?