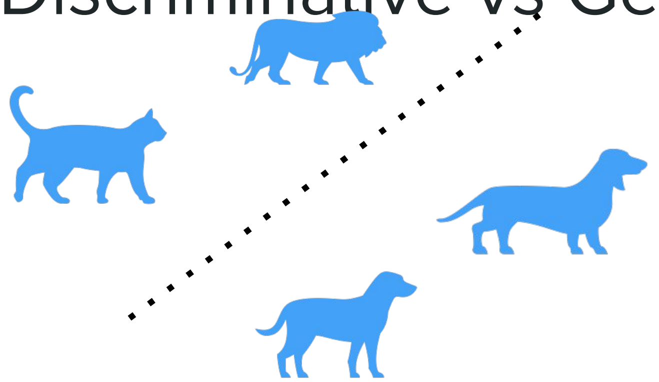
Generative Adversarial Networks

Konrad Kording

neuromatch academy

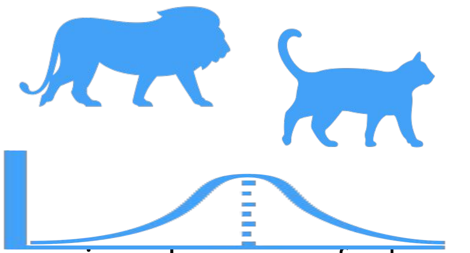
Discriminative vs Generative models



features class

$$X \longrightarrow Y$$

$$p(Y|X)$$



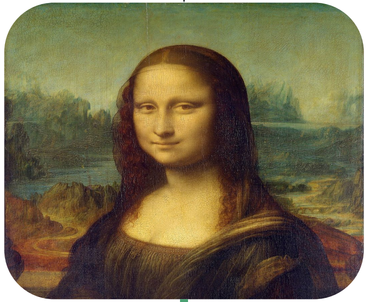
noise, class features

$$\xi, Y \longrightarrow X$$

$$p(X|Y)$$

Generator

Look here is a nice image

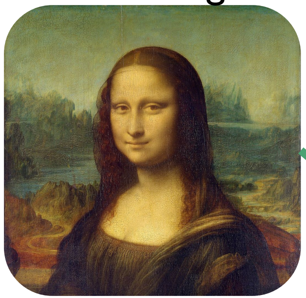


Discriminator

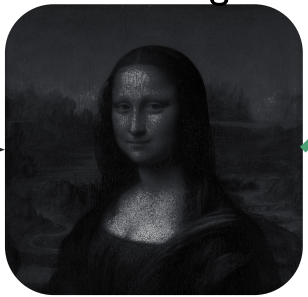
so fake

Generator constructs an image. **Their reward function is to not get caught.**

Real Image



Fake Image

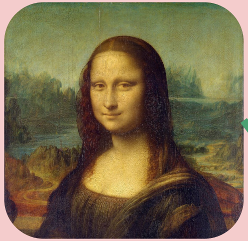


Generator

Discriminator

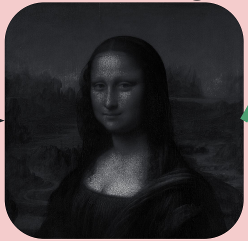
The Discriminator

Real Image



The discriminator sees a mix of real and fake images.

Fake Image



Generator

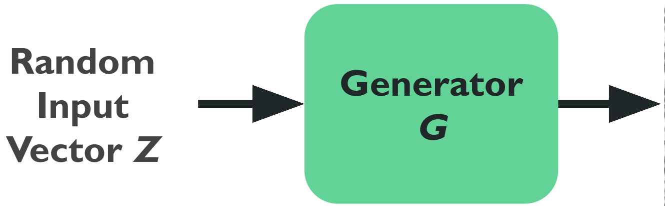
Discriminator

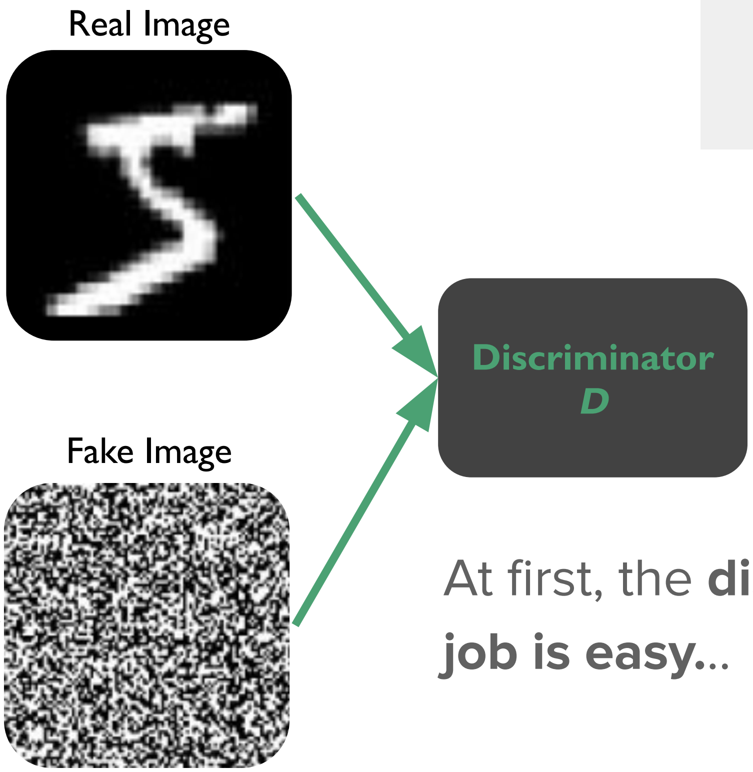
P(Real)=.9

Reward for identification of fake vs real.

The Model

- The generator *G* takes an input *Z* to generate some fake image
- The discriminator has to tell the difference between fake and real images

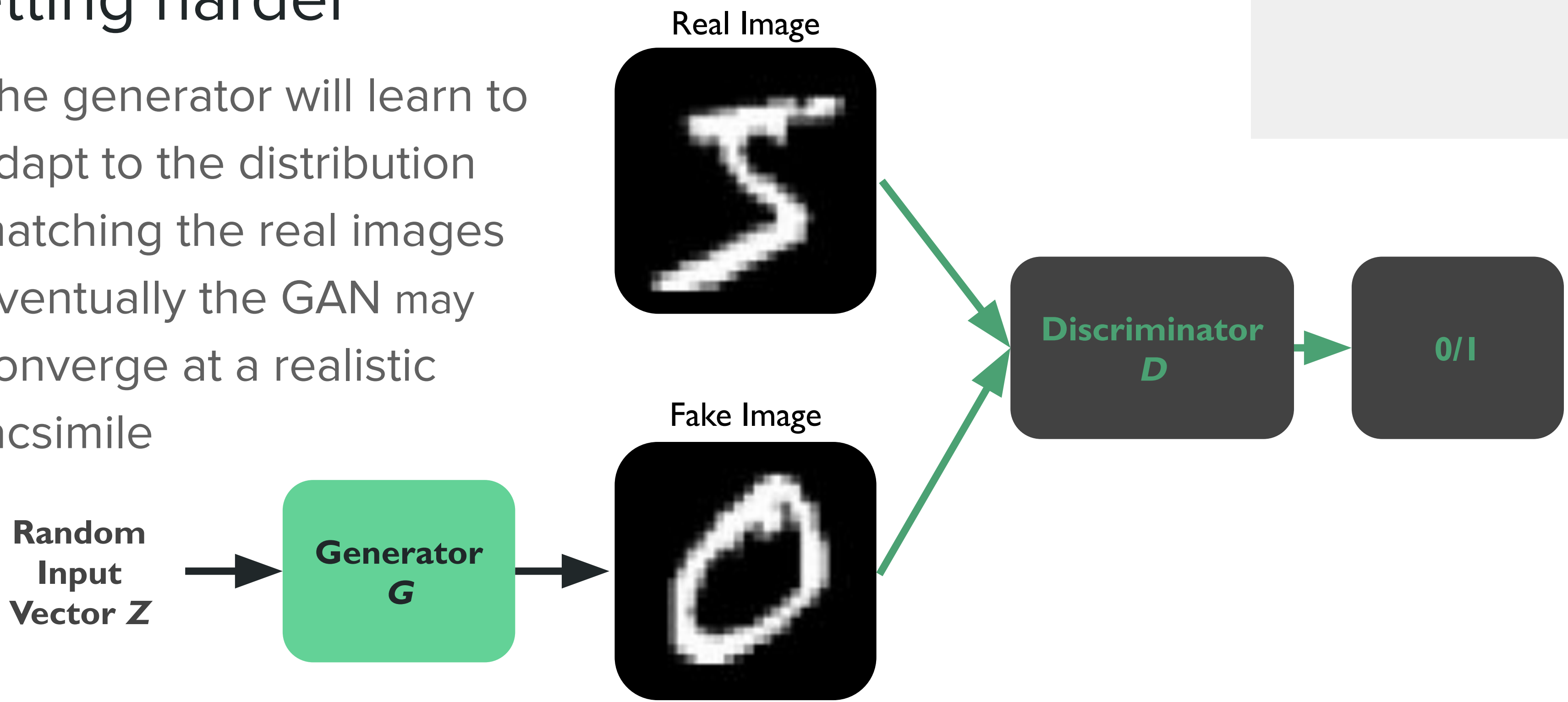




At first, the **discriminator's job is easy...**

Getting harder

- The generator will learn to adapt to the distribution matching the real images
- Eventually the GAN may converge at a realistic facsimile



The Discriminator Loss Function

Real y= 1, Fake y= 0

$$J_D = -\frac{1}{m}\sum_{i=1}^{m} y_i \log D(x_i) + (1 - y_i)\log\left(1 - D(x_i)\right)$$

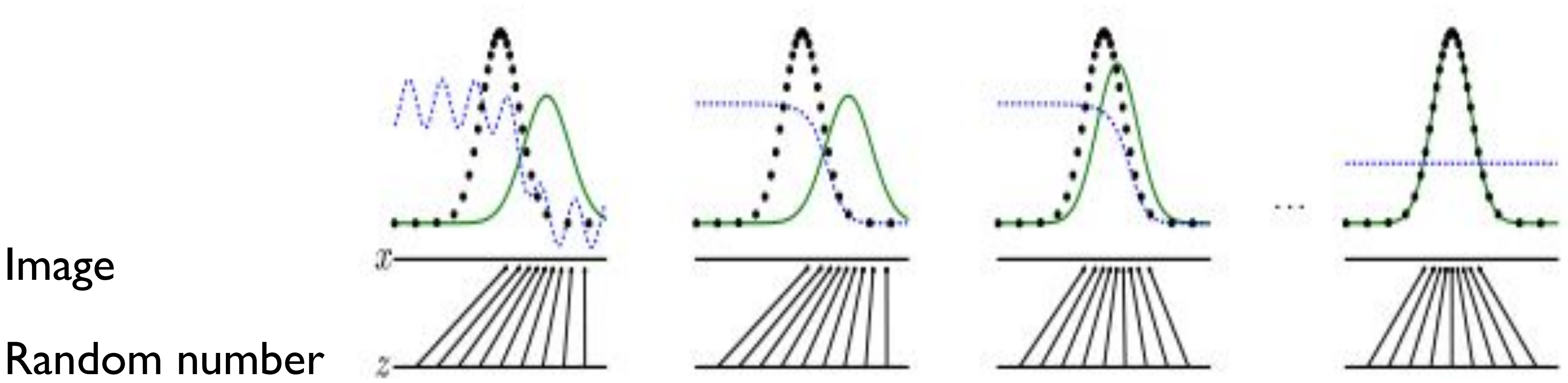
The Generator Loss Function

Can *G* avoid getting caught? How well did it do at fooling *D*?

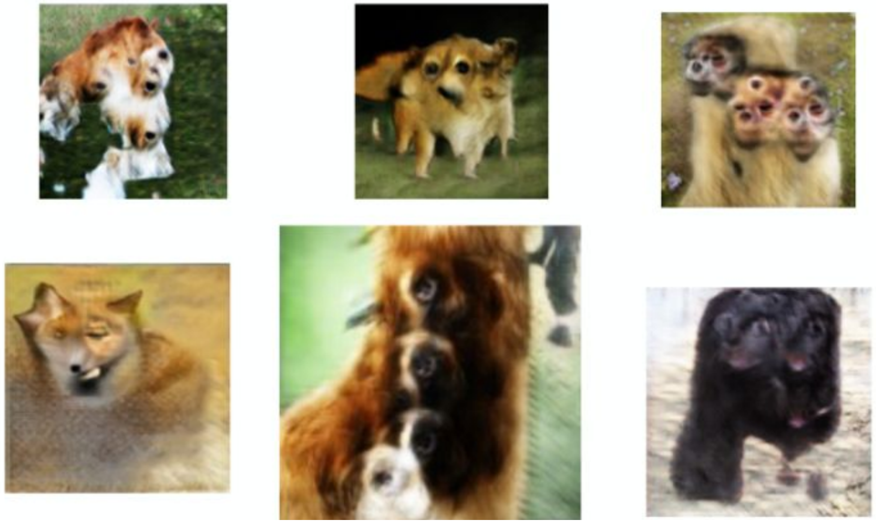
$$J_G = -J_D = \frac{1}{m}\sum_{i=1}^{m} y_i \log D(x_i) + (1 - y_i)\log(1 - D(x_i))$$

This loss function has problems, though...

GAN generator learning idea

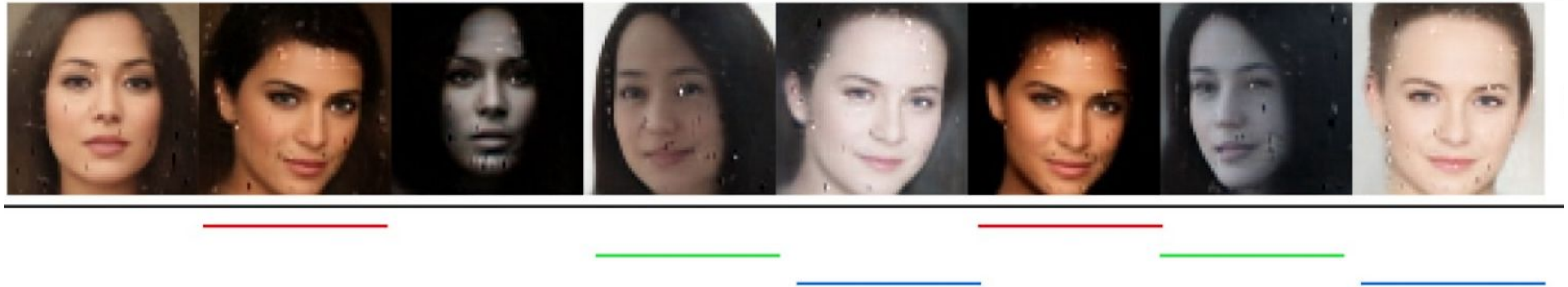


Typical GAN failure modes: Bad Images

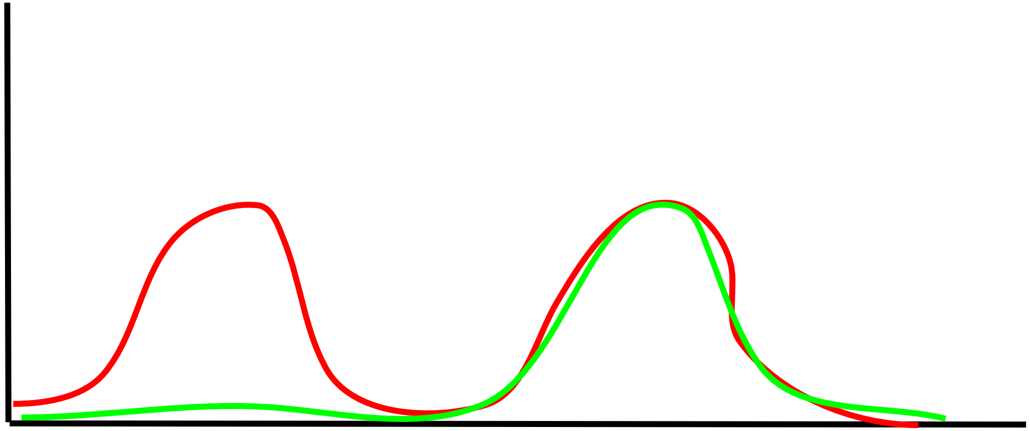


What is going on here?

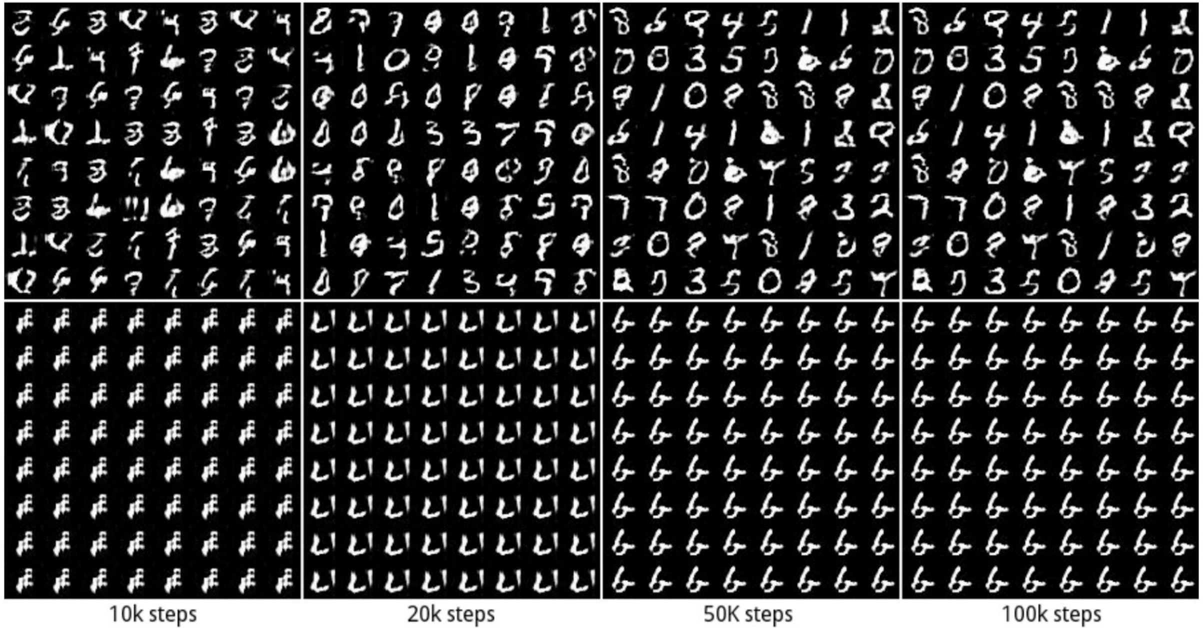
Mode collapse



Mode collapse



Mode collapse



Let us get an intuition about mode collapse

How to improve? Many ideas:

GAN Type	Key Take-Away
GAN	The original (JSD divergence)
WGAN	EM distance objective
Improved WGAN	No weight clipping on WGAN
LSGAN	L2 loss objective
RWGAN	Relaxed WGAN framework
McGAN	Mean/covariance minimization objective
GMMN	Maximum mean discrepancy objective
MMD GAN	Adversarial kernel to GMMN
Cramer GAN	Cramer distance
Fisher GAN	Chi-square objective
EBGAN	Autoencoder instead of discriminator
BEGAN	WGAN and EBGAN merged objectives
MAGAN	Dynamic margin on hinge loss from EBGAN

https://towardsdatascience.com/gan-objective-functions-gans-and-their-variations-ad77340bce3c

The math is often just a thin veneer to hide what we do not understand

But maybe read

On How Well Generative Adversarial Networks Learn Densities: Nonparametric and Parametric Results

Tengyuan Liang*[1]

[1]University of Chicago, Booth School of Business

Did they succeed?

Did anyone find the magical well-working formulation?

Measuring how good a GAN does

Take inception network

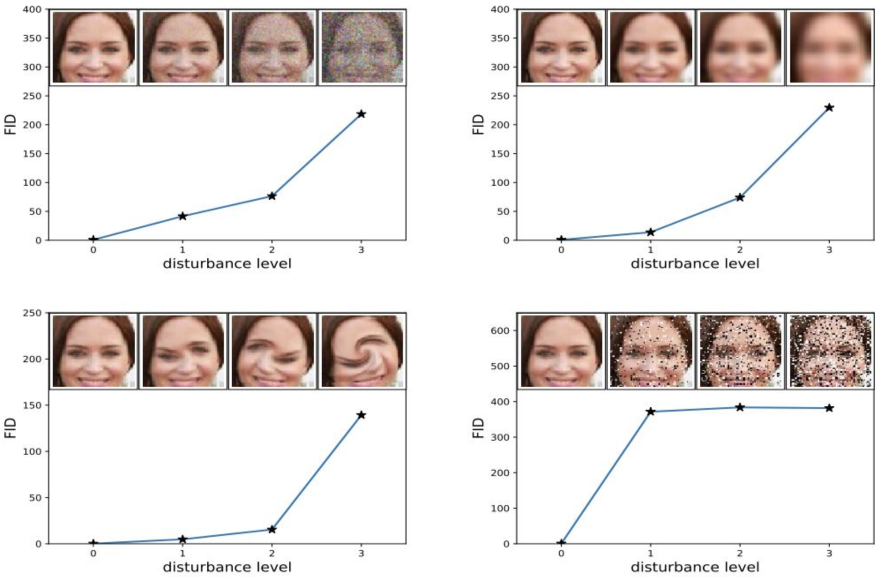
Use FC 2048 width layer

Probability distributions should be matched

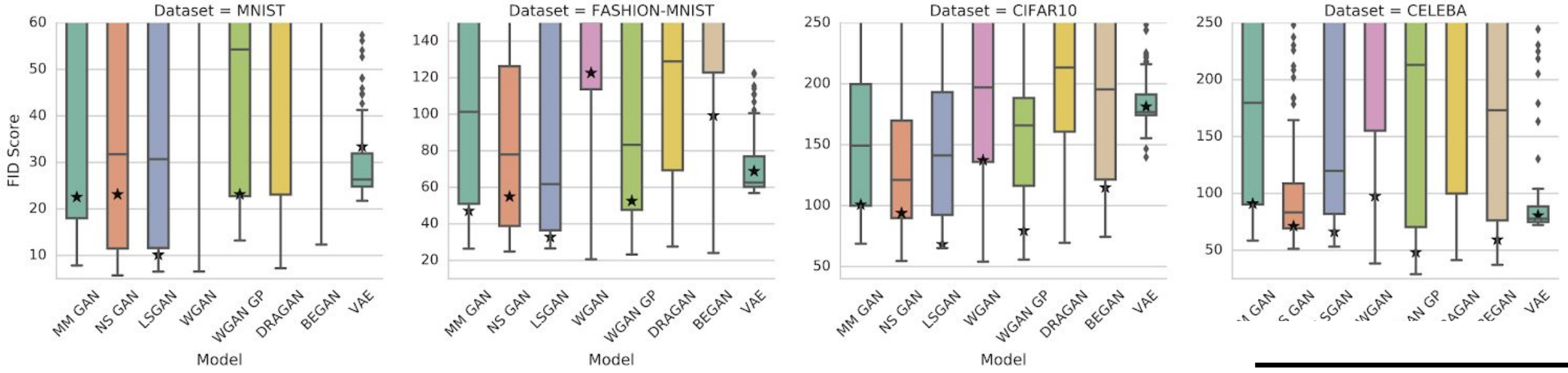
Frechet inception distance

$$\mathrm{FID} = ||\mu_r - \mu_g||^2 + \mathrm{Tr}(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{1/2}).$$

Kinda makes sense



Surprisingly all our innovations seem pretty useless



Are GANs Created Equal? A Large-Scale Study

Mario Lucic* Karol Kurach* Marcin Michalski Olivier Bousquet Sylvain Gelly
Google Brain