

Deep Generative Models

Lecture 11

Roman Isachenko



Spring, 2022

Recap of previous lecture

WGAN objective

$$\min_{\theta} W(\pi || p) = \min_{\theta} \max_{\phi \in \Phi} [\mathbb{E}_{\pi(x)} f(x, \phi) - \mathbb{E}_{p(z)} f(G(z, \theta), \phi)].$$

- ▶ Function f in WGAN is usually called *critic*.
- ▶ If parameters ϕ lie in a compact set $\Phi \in [-0.01, 0.01]^d$ then $f(x, \phi)$ will be K -Lipschitz continuous function.

Gradient penalty

$$W(\pi || p) = \underbrace{\mathbb{E}_{\pi(x)} f(x) - \mathbb{E}_{p(x)} f(x)}_{\text{original critic loss}} + \lambda \underbrace{\mathbb{E}_{U[0,1]} \left[(\|\nabla_{\hat{x}} f(\hat{x})\|_2 - 1)^2 \right]}_{\text{gradient penalty}}.$$

Samples $\hat{x}_t = t\mathbf{y} + (1-t)\mathbf{z}$ with $t \in [0, 1]$ are uniformly sampled along straight lines between pairs of points: \mathbf{y} from the data distribution $\pi(\mathbf{x})$ and \mathbf{z} from the generator distribution $p(\mathbf{x}|\theta)$.

Arjovsky M., Chintala S., Bottou L. Wasserstein GAN, 2017

Gulrajani I. et al. Improved Training of Wasserstein GANs, 2017

Recap of previous lecture

$$f(\mathbf{x}, \phi) = \mathbf{W}_{K+1} \sigma_K (\mathbf{W}_K \sigma_{K-1} (\dots \sigma_1 (\mathbf{W}_1 \mathbf{x}) \dots)).$$

- ▶ σ_k is a pointwise nonlinearities. We assume that $\|\sigma_k\|_L = 1$ (it holds for ReLU).
- ▶ $\mathbf{g}(\mathbf{x}) = \mathbf{W}\mathbf{x}$ is a linear transformation ($\nabla \mathbf{g}(\mathbf{x}) = \mathbf{W}$).

$$\|\mathbf{g}\|_L = \sup_{\mathbf{x}} \|\nabla \mathbf{g}(\mathbf{x})\|_2 = \|\mathbf{W}\|_2.$$

Critic spectral norm

$$\|f\|_L \leq \|\mathbf{W}_{K+1}\|_2 \cdot \prod_{k=1}^K \|\sigma_k\|_L \cdot \|\mathbf{W}_k\|_2 = \prod_{k=1}^{K+1} \|\mathbf{W}_k\|_2.$$

Spectral Normalization GAN

If we replace the weights in the critic $f(\mathbf{x}, \phi)$ by $\mathbf{W}_k^{SN} = \mathbf{W}_k / \|\mathbf{W}_k\|_2$, we will get $\|f\|_L \leq 1$.

Power iteration approximates the value of $\|\mathbf{W}\|_2$.

Recap of previous lecture

f-divergence minimization

$$D_f(\pi || p) = \mathbb{E}_{p(\mathbf{x})} f\left(\frac{\pi(\mathbf{x})}{p(\mathbf{x})}\right) \rightarrow \min_p .$$

Here $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ is a convex, lower semicontinuous function satisfying $f(1) = 0$.

Variational divergence estimation

$$D_f(\pi || p) \geq \sup_{T \in \mathcal{T}} [\mathbb{E}_\pi T(\mathbf{x}) - \mathbb{E}_p f^*(T(\mathbf{x}))],$$

Fenchel conjugate

$$f^*(t) = \sup_{u \in \text{dom}_f} (ut - f(u)), \quad f(u) = \sup_{t \in \text{dom}_{f^*}} (ut - f^*(t))$$

Note: To evaluate lower bound we only need samples from $\pi(\mathbf{x})$ and $p(\mathbf{x})$. Hence, we could fit implicit generative model.

Outline

1. Evaluation of likelihood-free models

Inception score

Frechet Inception Distance

Precision-Recall

2. Evolution of GANs

Outline

1. Evaluation of likelihood-free models

Inception score

Frechet Inception Distance

Precision-Recall

2. Evolution of GANs

Evaluation of likelihood-free models

How to evaluate generative models?

Likelihood-based models

- ▶ Split data to train/val/test.
- ▶ Fit model on the train part.
- ▶ Tune hyperparameters on the validation part.
- ▶ Evaluate generalization by reporting likelihoods on the test set.

Not all models have tractable likelihoods

- ▶ VAE: compare ELBO values.
- ▶ GAN: ???

Evaluation of likelihood-free models

Let's take some pretrained image classification model to get the conditional label distribution $p(y|x)$ (e.g. ImageNet classifier).

What do we want from samples?

- ▶ Sharpness



The conditional distribution $p(y|x)$ should have low entropy (each image x should have distinctly recognizable object).

- ▶ Diversity



The marginal distribution $p(y) = \int p(y|x)p(x)dx$ should have high entropy (there should be as many classes generated as possible).

Evaluation of likelihood-free models

What do we want from samples?

- ▶ **Sharpness.** The conditional distribution $p(y|x)$ should have low entropy (each image x should have distinctly recognizable object).
- ▶ **Diversity.** The marginal distribution $p(y) = \int p(y|x)p(x)dx$ should have high entropy (there should be as many classes generated as possible).

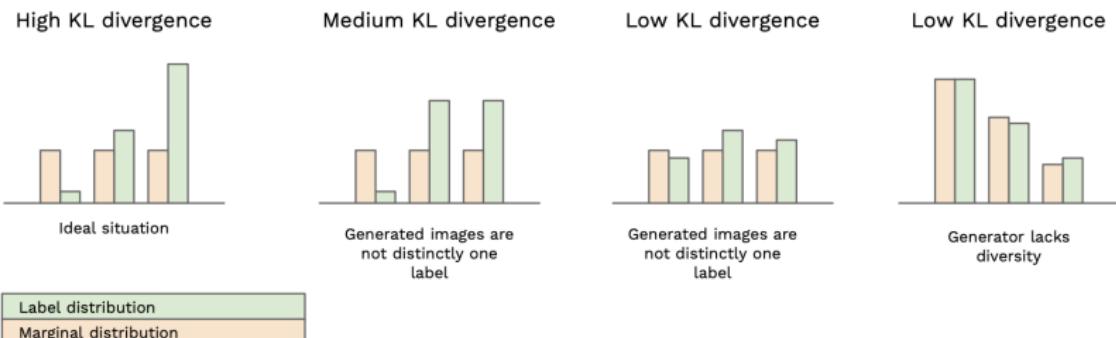


image credit: <https://medium.com/octavian-ai/a-simple-explanation-of-the-inception-score-372dff6a8c7a>

Outline

1. Evaluation of likelihood-free models

Inception score

Frechet Inception Distance

Precision-Recall

2. Evolution of GANs

Evaluation of likelihood-free models

What do we want from samples?

- ▶ Sharpness \Rightarrow low $H(y|\mathbf{x}) = - \sum_y \int_{\mathbf{x}} p(y, \mathbf{x}) \log p(y|\mathbf{x}) d\mathbf{x}$.
- ▶ Diversity \Rightarrow high $H(y) = - \sum_y p(y) \log p(y)$.

Inception Score

$$\begin{aligned} IS &= \exp(H(y) - H(y|\mathbf{x})) \\ &= \exp \left(- \sum_y p(y) \log p(y) + \sum_y \int_{\mathbf{x}} p(y, \mathbf{x}) \log p(y|\mathbf{x}) d\mathbf{x} \right) \\ &= \exp \left(\sum_y \int_{\mathbf{x}} p(y, \mathbf{x}) \log \frac{p(y|\mathbf{x})}{p(y)} d\mathbf{x} \right) \\ &= \exp \left(\mathbb{E}_{\mathbf{x}} \sum_y p(y|\mathbf{x}) \log \frac{p(y|\mathbf{x})}{p(y)} \right) = \exp(\mathbb{E}_{\mathbf{x}} KL(p(y|\mathbf{x}) || p(y))) \end{aligned}$$

Outline

1. Evaluation of likelihood-free models

Inception score

Frechet Inception Distance

Precision-Recall

2. Evolution of GANs

Evaluation of likelihood-free models

Theorem (informal)

If $\pi(\mathbf{x})$ and $p(\mathbf{x}|\theta)$ has moment generation functions then

$$\pi(\mathbf{x}) = p(\mathbf{x}|\theta) \Leftrightarrow \mathbb{E}_\pi \mathbf{x}^k = \mathbb{E}_p \mathbf{x}^k, \quad \forall k \geq 1.$$

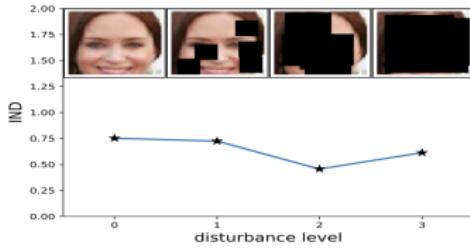
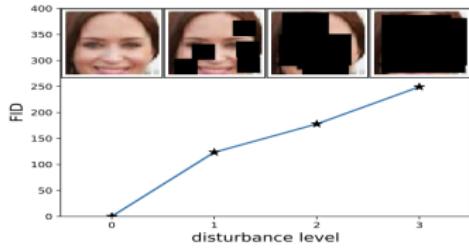
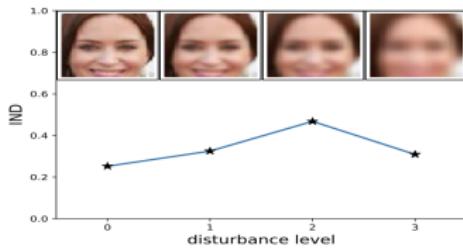
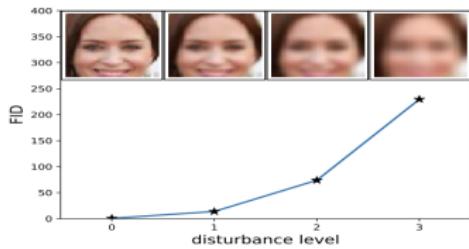
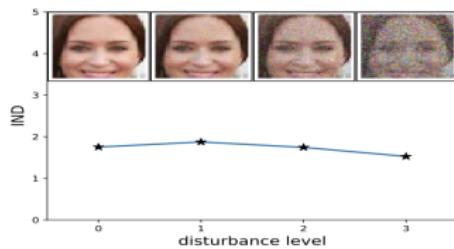
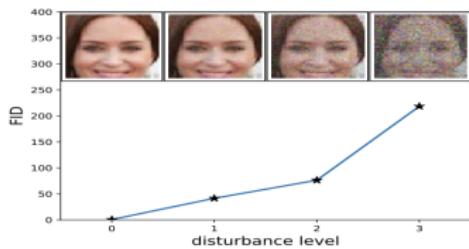
This is intractable to calculate all moments.

Frechet Inception Distance

$$FID(\pi, p) = \|\mathbf{m}_\pi - \mathbf{m}_p\|_2^2 + \text{Tr} \left(\boldsymbol{\Sigma}_\pi + \boldsymbol{\Sigma}_p - 2\sqrt{\boldsymbol{\Sigma}_\pi \boldsymbol{\Sigma}_p} \right)$$

- ▶ Representations are the outputs of the intermediate layer from the pretrained classification model.
- ▶ $\mathbf{m}_\pi, \boldsymbol{\Sigma}_\pi$ are the mean vector and the covariance matrix of feature representations for samples from $\pi(\mathbf{x})$
- ▶ $\mathbf{m}_p, \boldsymbol{\Sigma}_p$ are the mean vector and the covariance matrix of feature representations for samples from $p(\mathbf{x}|\theta)$.

Evaluation of likelihood-free models



Limitations

Inception Score

$$IS = \exp(\mathbb{E}_{\mathbf{x}} KL(p(y|\mathbf{x}) || p(y)))$$

- ▶ If generator produces images with a different set of labels from the classifier training set, IS will be low.
- ▶ If generator produces one image per class, the IS will be perfect (there is no measure of intra-class diversity).

Frechet Inception Distance

$$FID = \|\mathbf{m}_\pi - \mathbf{m}_p\|_2^2 + \text{Tr} \left(\boldsymbol{\Sigma}_\pi + \boldsymbol{\Sigma}_p - 2\sqrt{\boldsymbol{\Sigma}_\pi \boldsymbol{\Sigma}_p} \right)$$

- ▶ Needs a large sample size for evaluation.
- ▶ Calculation of FID is slow.
- ▶ Estimates only two sample moments.

Both scores depend on the pretrained classifier $p(y|\mathbf{x})$.

Barratt S., Sharma R. A Note on the Inception Score, 2018

Heusel M. et al. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium, 2017

Outline

1. Evaluation of likelihood-free models

Inception score

Frechet Inception Distance

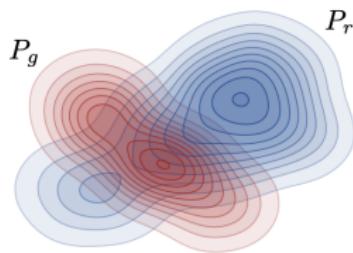
Precision-Recall

2. Evolution of GANs

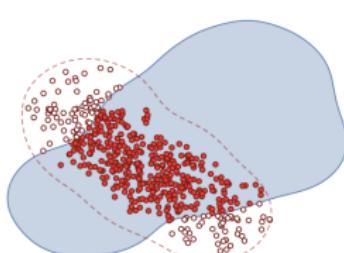
Precision-Recall for Generative Models

What do we want from samples

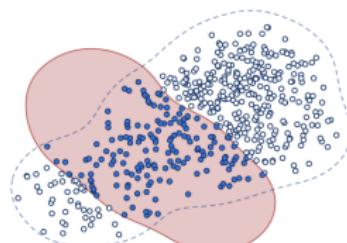
- ▶ **Sharpness:** generated samples should be of high quality.
- ▶ **Diversity:** their variation should match that observed in the training set.



(a) Example distributions



(b) Precision



(c) Recall

- ▶ **Precision** denotes the fraction of generated images that are realistic.
- ▶ **Recall** measures the fraction of the training data manifold covered by the generator.

Precision-Recall for generative models

- ▶ $\mathcal{S}_\pi = \{\mathbf{x}_i\}_{i=1}^n \sim \pi(\mathbf{x})$ – real samples;
- ▶ $\mathcal{S}_p = \{\mathbf{x}_i\}_{i=1}^n \sim p(\mathbf{x}|\theta)$ – generated samples.

Embed samples using pretrained classifier network (as previously):

$$\mathcal{G}_\pi = \{\mathbf{g}_i\}_{i=1}^n, \quad \mathcal{G}_p = \{\mathbf{g}_i\}_{i=1}^n.$$

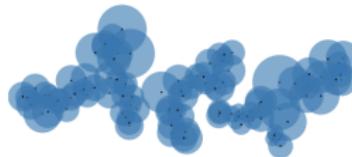
Define binary function:

$$f(\mathbf{g}, \mathcal{G}) = \begin{cases} 1, & \text{if exists } \mathbf{g}' \in \mathcal{G} : \|\mathbf{g} - \mathbf{g}'\|_2 \leq \|\mathbf{g}' - \text{NN}_k(\mathbf{g}', \mathcal{G})\|_2; \\ 0, & \text{otherwise.} \end{cases}$$

$$\text{Precision}(\mathcal{G}_\pi, \mathcal{G}_p) = \frac{1}{n} \sum_{\mathbf{g} \in \mathcal{G}_p} f(\mathbf{g}, \mathcal{G}_\pi); \quad \text{Recall}(\mathcal{G}_\pi, \mathcal{G}_p) = \frac{1}{n} \sum_{\mathbf{g} \in \mathcal{G}_\pi} f(\mathbf{g}, \mathcal{G}_p).$$

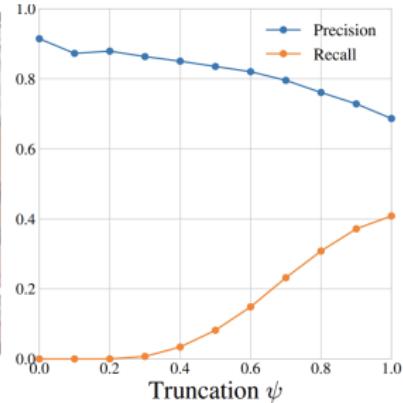
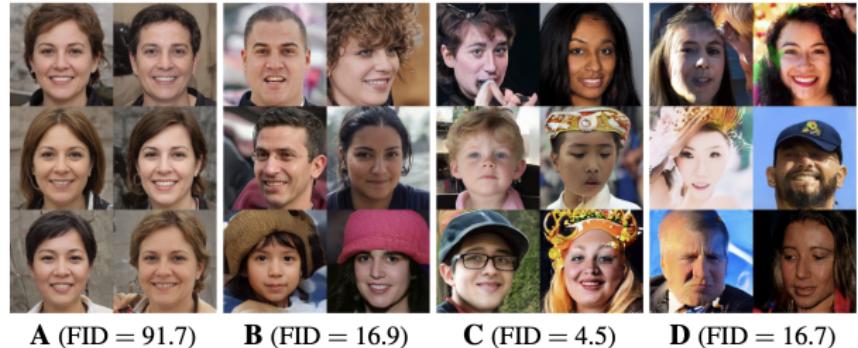
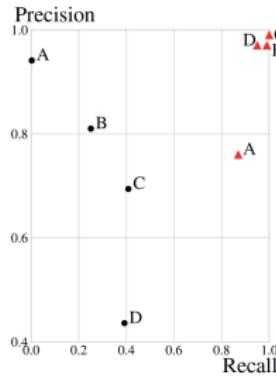


(a) True manifold



(b) Approx. manifold

Precision-Recall for generative models



Truncation trick

BigGAN: truncated normal sampling

$$p(\mathbf{z}|b) = \mathcal{N}(\mathbf{z}|0, \mathbf{I}) / \int_{-\infty}^{\psi} \mathcal{N}(\mathbf{z}|0, \mathbf{I}) d\mathbf{z}$$

Components of $\mathbf{z} \sim \mathcal{N}(0, \mathbf{I})$ which fall outside a predefined range are resampled.

StyleGAN

$$\mathbf{z}' = \hat{\mathbf{z}} + \psi \cdot (\mathbf{z} - \hat{\mathbf{z}}), \quad \hat{\mathbf{z}} = \mathbb{E}_{\mathbf{z}} \mathbf{z}$$

- ▶ Constant ψ is a tradeoff between diversity and fidelity.
- ▶ $\psi = 0.7$ is used for most of the results.

Outline

1. Evaluation of likelihood-free models

Inception score

Frechet Inception Distance

Precision-Recall

2. Evolution of GANs

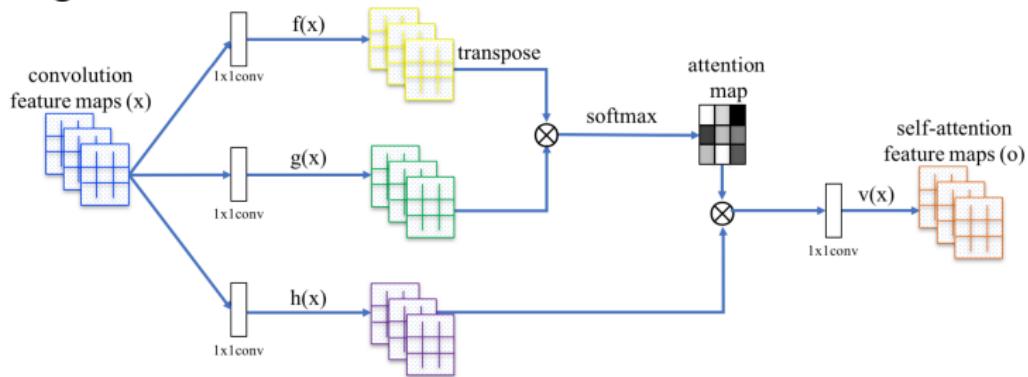
Evolution of GANs



- ▶ **Standard GAN** <https://arxiv.org/abs/1406.2661>
- ▶ **DCGAN** <https://arxiv.org/abs/1511.06434>
- ▶ **CoGAN** <https://arxiv.org/abs/1606.07536>
- ▶ **ProGAN** <https://arxiv.org/abs/1710.10196>
- ▶ **StyleGAN** <https://arxiv.org/abs/1812.04948>

Self-Attention GAN

Convolutional layers process the information in a local neighborhood \Rightarrow inefficient for modeling long-range dependencies in images.

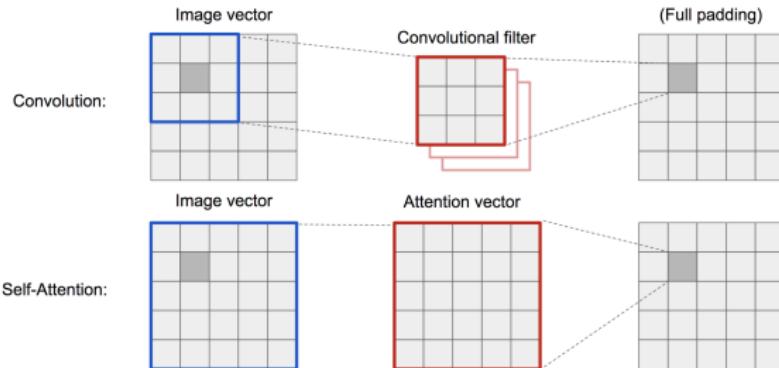


$$\mathbf{f}(\mathbf{x}) = \mathbf{W}_f \mathbf{x}, \quad \mathbf{g}(\mathbf{x}) = \mathbf{W}_g \mathbf{x}, \quad \mathbf{h}(\mathbf{x}) = \mathbf{W}_h \mathbf{x}, \quad \mathbf{v}(\mathbf{x}) = \mathbf{W}_v \mathbf{x}$$

$$s_{ij} = \mathbf{f}(\mathbf{x}_i)^T \mathbf{g}(\mathbf{x}_j), \quad a_{ij} = \frac{\exp s_{ij}}{\sum_{i=1}^N \exp s_{ij}}, \quad \mathbf{o}_j = \mathbf{v} \left(\sum_{i=1}^N a_{ij} \mathbf{h}(\mathbf{x}_i) \right)$$

Self-Attention GAN

Convolution vs Attention



Visualization of attention maps



image credit: <https://lilianweng.github.io/lil-log/2018/06/24/attention-attention.html>
Zhang H. et al. Self-Attention Generative Adversarial Networks, 2018

BigGAN

Batch-size is matter

Batch	Ch.	Param (M)	Shared	Skip- z	Ortho.	Itr $\times 10^3$	FID	IS
256	64	81.5	SA-GAN Baseline			1000	18.65	52.52
512	64	81.5	X	X	X	1000	15.30	58.77(± 1.18)
1024	64	81.5	X	X	X	1000	14.88	63.03(± 1.42)
2048	64	81.5	X	X	X	732	12.39	76.85(± 3.83)
2048	96	173.5	X	X	X	295(± 18)	9.54(± 0.62)	92.98(± 4.27)

Samples (512x512)



Progressive Growing GAN

Problems with HR image generation

- ▶ Disjoint manifolds \Rightarrow gradient problem.
- ▶ Small minibatch \Rightarrow training instability.

Samples (1024x1024)

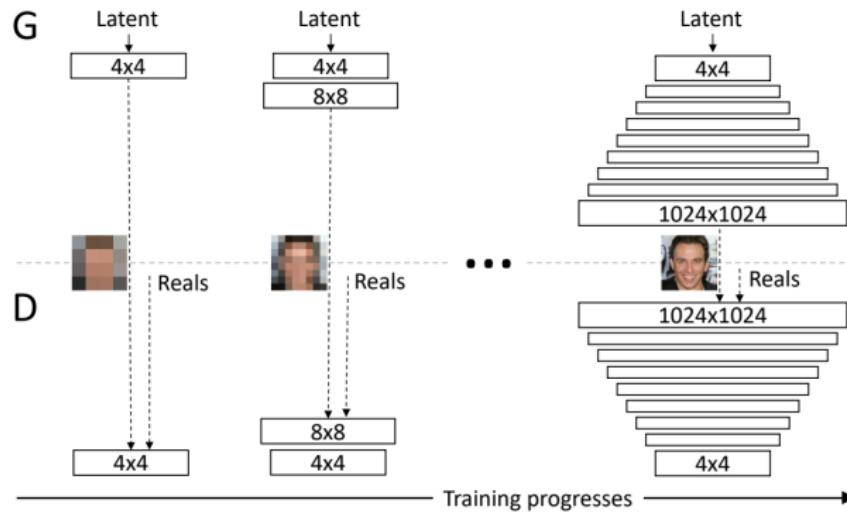


Karras T. et al. *Progressive Growing of GANs for Improved Quality, Stability, and Variation*, 2017

Progressive Growing GAN

Grow both the generator and discriminator progressively, new layers will introduce higher-resolution details as the training progresses.

- ▶ Train GAN which generate 4x4 images (2 convs for G and D).
- ▶ Add upsampling layers to G, downsampling layers to D.
- ▶ Train GAN which generate 8x8 images.
- ▶ etc.



StyleGAN

- ▶ Generating of HR images is hard.
- ▶ Progressive growing greatly simplifies the task.
- ▶ The ability to control specific features of the generated image is very limited.

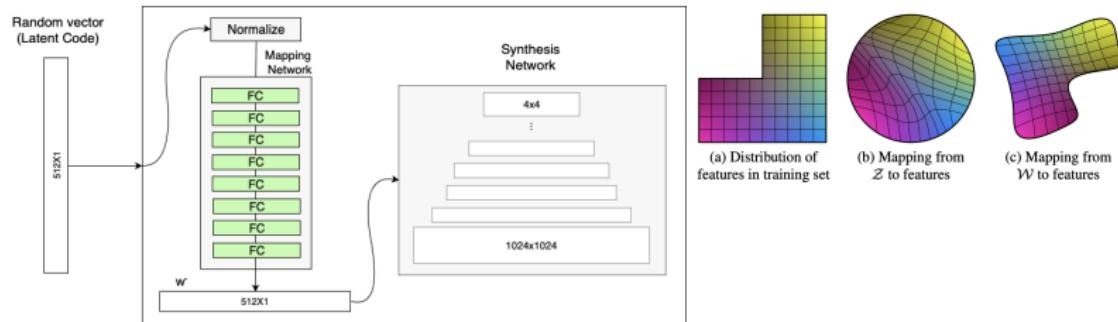
Face image features

- ▶ Coarse (pose, general hair style, face shape). Resolution $4^2 - 8^2$.
- ▶ Middle (finer facial features, hair style, eyes open/closed). Resolution $16^2 - 32^2$.
- ▶ Fine (color scheme (eye, hair and skin) and micro features). Resolution $64^2 - 1024^2$.

StyleGAN

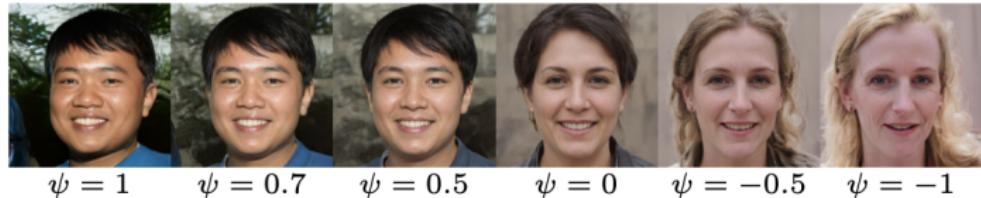
Mapping Network

- ▶ Generator input is likely to be **disentangled**. Each component of input vector \mathbf{z} should be responsible for one generative factor.
- ▶ Mapping network $f : \mathcal{Z} \rightarrow \mathcal{W}$ is used to reduce correlations between components of \mathbf{z} .



StyleGAN

Truncation trick



Samples (1024x1024)



Karras T., Laine S., Aila T. A Style-Based Generator Architecture for Generative Adversarial Networks, 2018

Summary

- ▶ Inception Score and Frechet Inception Distance are the common metrics for GAN evaluation, but both of them have drawbacks.
- ▶ Precision-recall allows to select model that compromises the sample quality and the sample diversity.
- ▶ Self-Attention GAN allows to make huge receptive field and reduce convolutional inductive bias.
- ▶ BigGAN shows that large batch size increase model quality gradually.
- ▶ Progressive growing for GAN learning allows to make training more stable.
- ▶ StyleGAN introduces mapping network to get more disentangled latent representation.