Learning Generative Models
Generative Adversarial Networks

Sebastian Stober & Jens Johannsmeier

May 7, 2024

OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG
FACULTY OF COMPUTER SCIENCE
AILAB

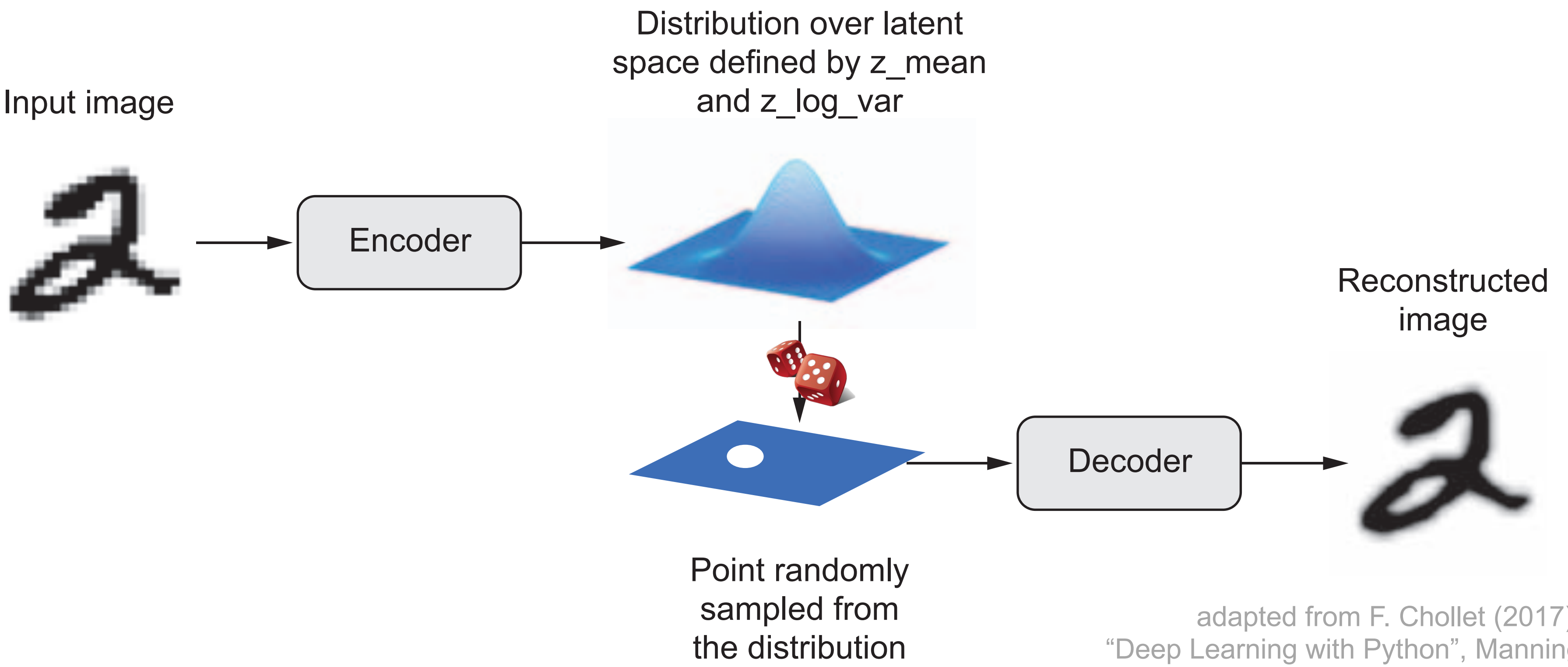
Outline for Today

- VAE Recap

- GAN Discussion
  & VAE Comparison

Recap: VAEs

Variational Auto-Encoder (VAE)



adapted from F. Chollet (2017) "Deep Learning with Python", Manning

encoder output distribution

decoder output distribution

latent prior

$$\mathcal{L}(\theta, \phi, x) = -D_{\mathrm{KL}}\left(q_\phi(z \mid x) \| p_\theta(z)\right) + \mathbb{E}_{q_\phi(z|x)}\left[\log p_\theta(x \mid z)\right]$$

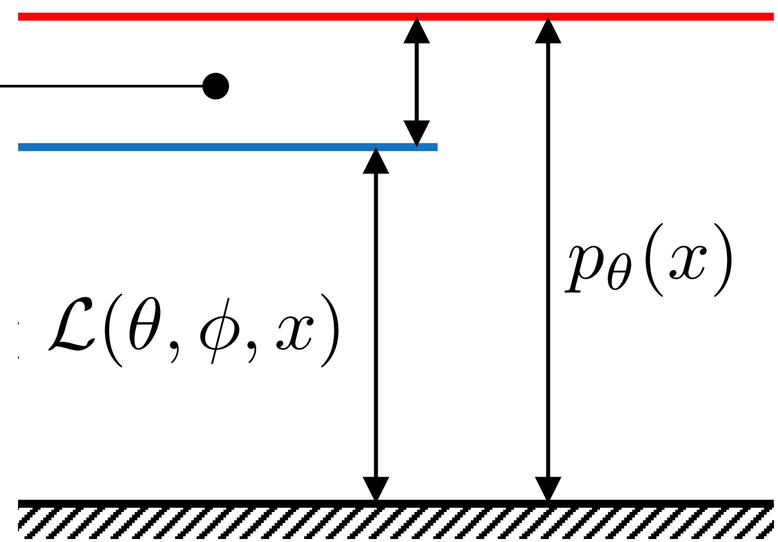
regularization term

reconstruction term

Recap: VAE ELBO

- define a variational lower bound on the data likelihood: $p_\theta(x) \geq \mathcal{L}(\theta, \phi, x)$

error introduced by approximate inference
= KL-Divergence of q and p
- expected to be small for high-capacity q



Recap: VAE ELBO

- define a variational lower bound on the data likelihood: $p_\theta(x) \geq \mathcal{L}(\theta, \phi, x)$

$$\mathcal{L}(\theta, \phi, x) = -\underbrace{D_{\mathrm{KL}}\left(\overbrace{q_\phi(z \mid x)}^{\text{encoder output distribution}} \| \overbrace{p_\theta(z)}^{\text{latent prior}}\right)}_{\text{regularization term}} + \underbrace{\mathbb{E}_{q_\phi(z|x)}\left[\log \overbrace{p_\theta(x \mid z)}^{\text{decoder output distribution}}\right]}_{\text{reconstruction term}}$$

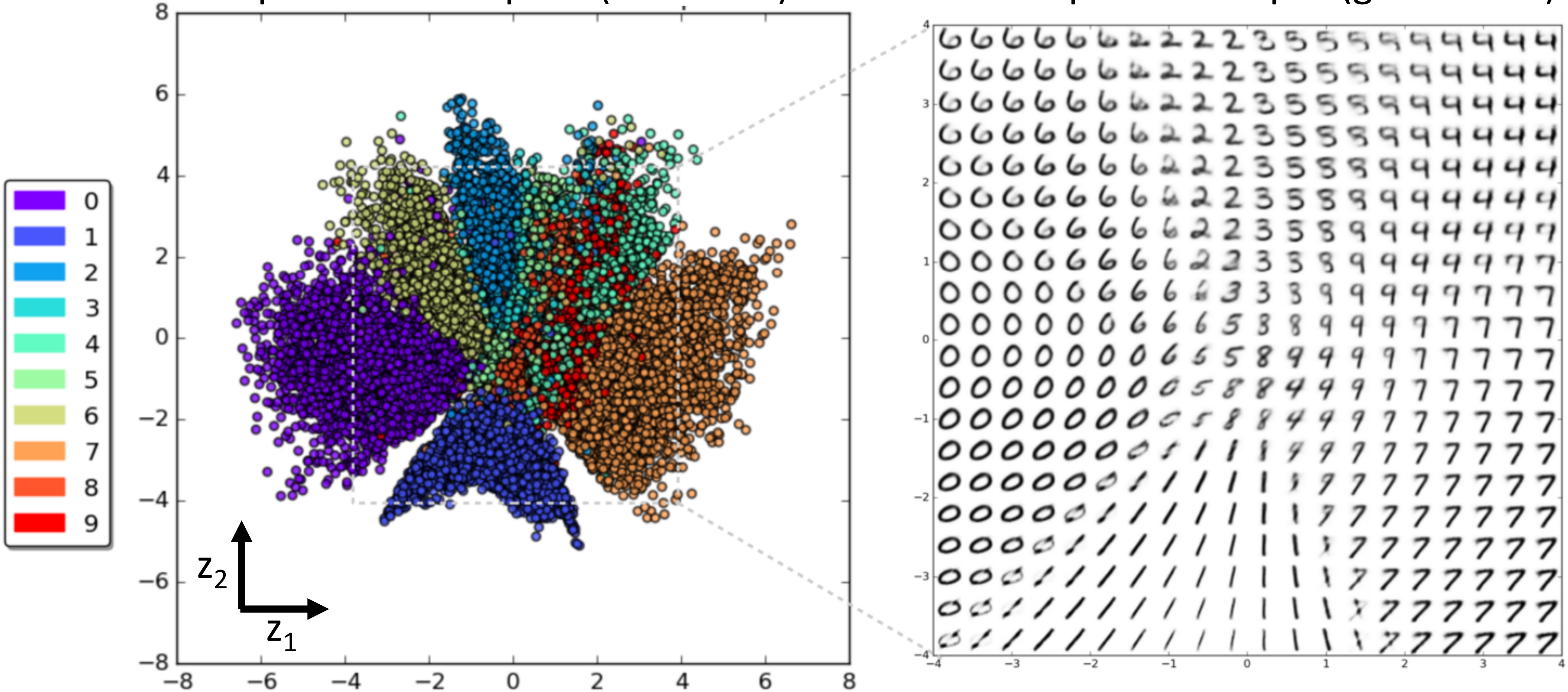
Penalize deviation from prior!

How likely is the output x given the inferred values of the latent variables z?

VAE Introspection

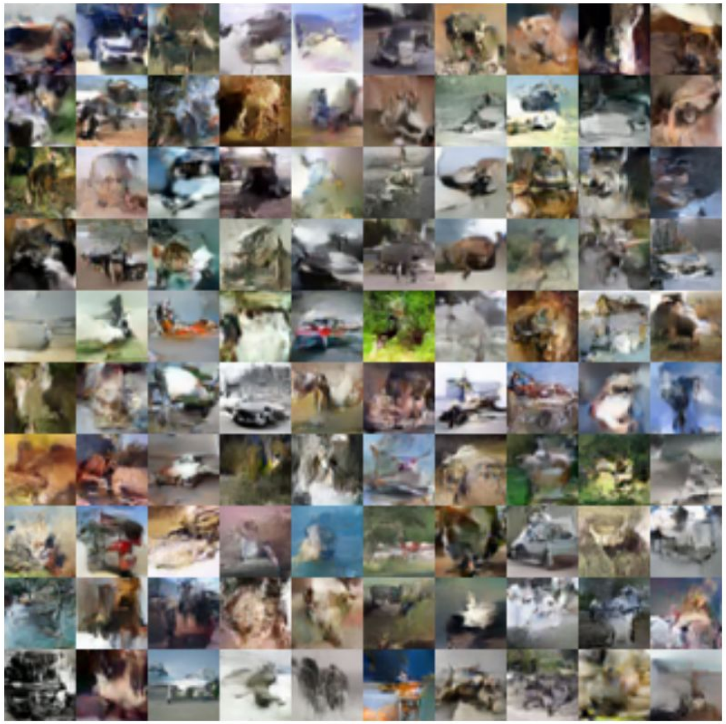
Latent Space Visualization (for MNIST dataset)

input to latent space (inference) latent space to output (generation)



Room for Improvements

- generated images tend to be blurry



Kingma et al. 2016

Larsen et al. 2017

VAEs Summary

- probabilistic spin to traditional auto-encoders
- optimize a (variational) lower bound

pros:

- principled approach to generative models
- inference of $q(z|x)$ can be useful feature representation

cons:

- maximizes only lower bound of likelihood
- samples blurrier and lower quality compared to GANs

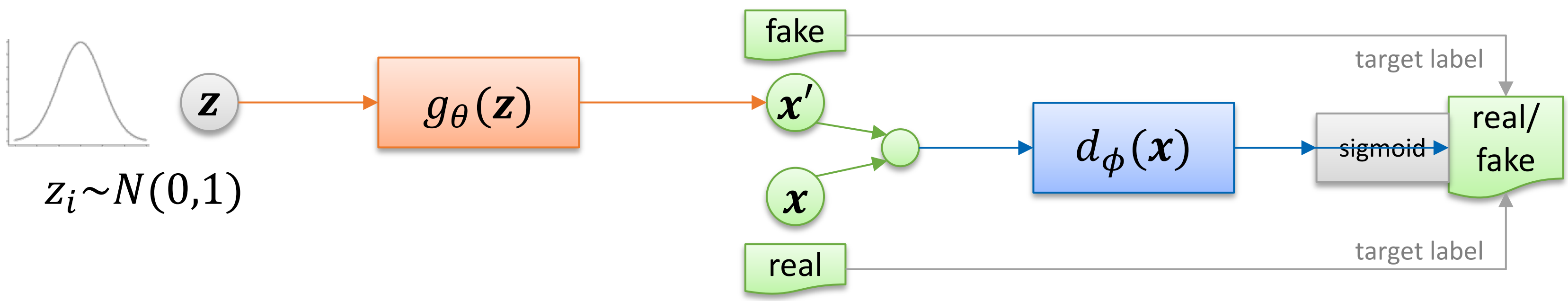
active areas of research:

- more flexible approximations, e.g. richer approximate posterior instead of diagonal Gaussian
- incorporating structure in latent variables

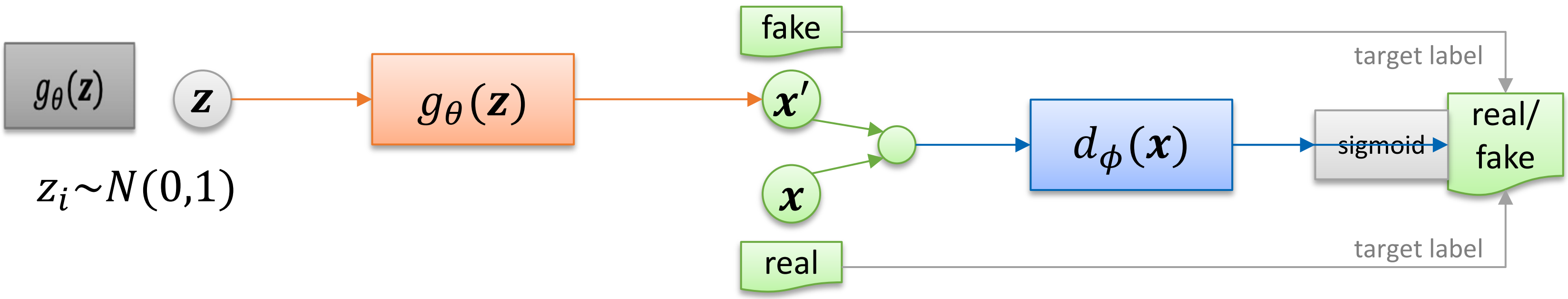
GANs

Discussion

Generative Adversarial Network



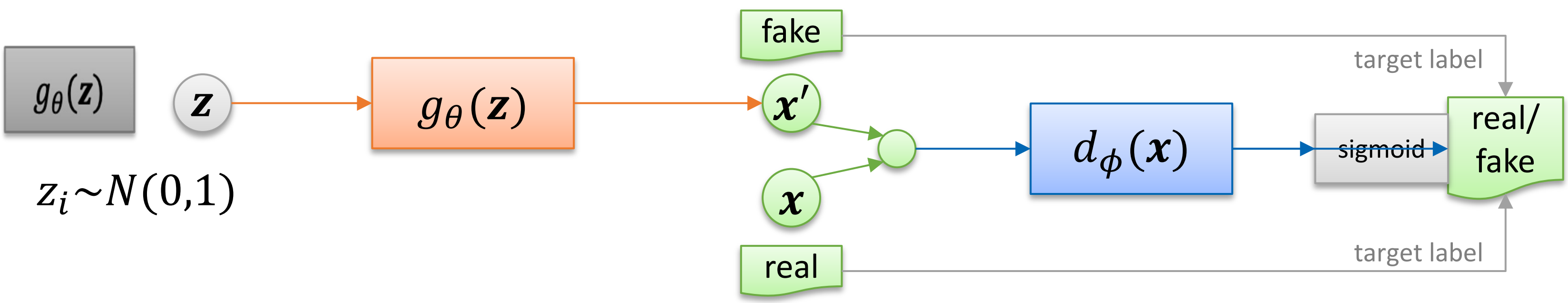
Generative Adversarial Network



objective$_g$ $\min_\theta (1 - d_\phi(g_\theta(z)))$

objective$_d$ $\max_\phi \begin{pmatrix} d_\phi(x) \\ + \left(1 - d_\phi\left(g_\theta(z)\right)\right) \end{pmatrix}$

Generative Adversarial Network



objective$_g$ $\min_\theta (1 - d_\phi(g_\theta(z)))$ objective$_d$ $\max_\phi \begin{pmatrix} d_\phi(x) \\ + \left(1 - d_\phi\left(g_\theta(z)\right)\right) \end{pmatrix}$

update θ to minimize loss$_g$

loss$_g$ $= \log\left(1 - d_\phi\left(g_\theta(z)\right)\right)$

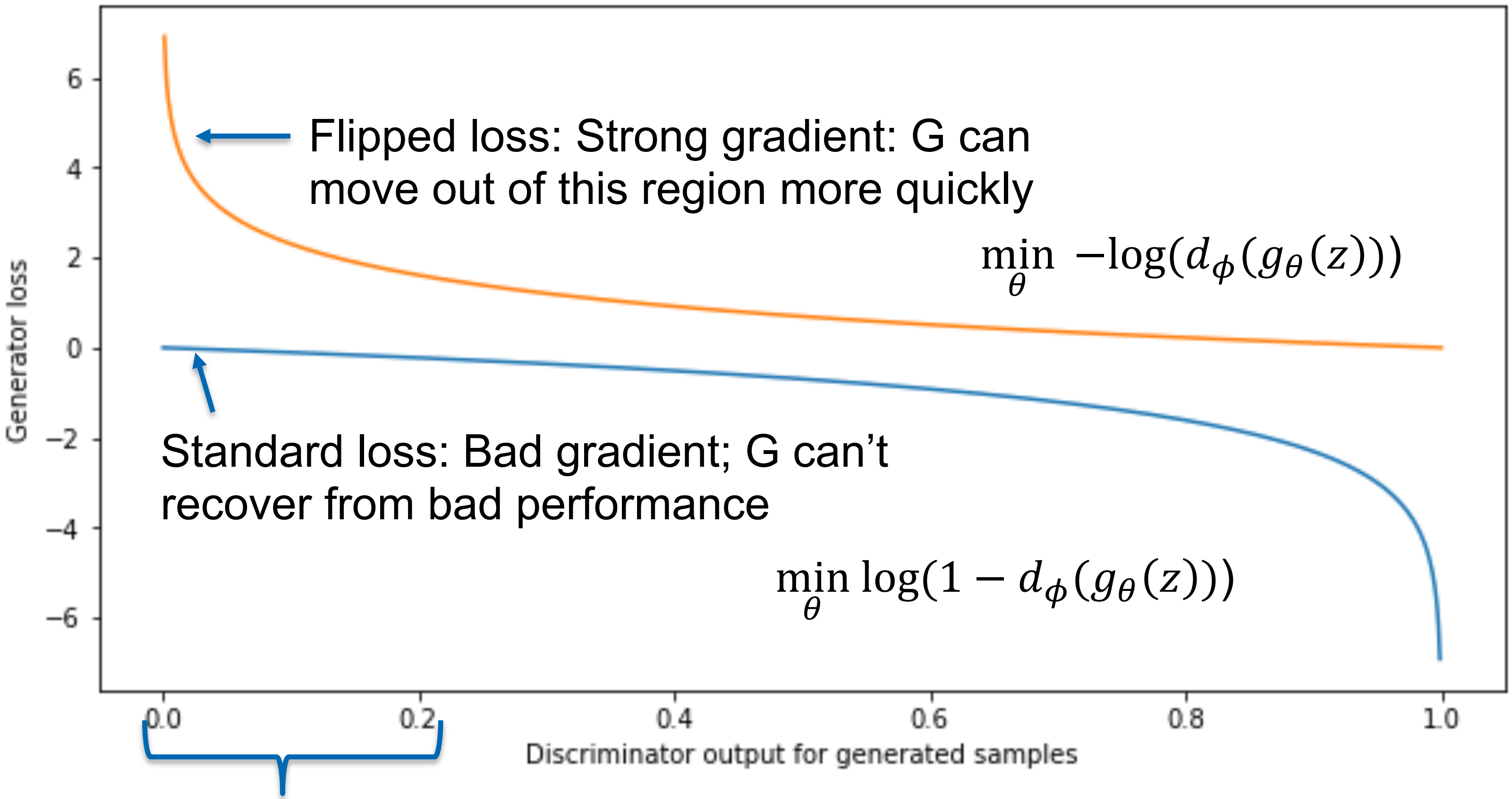
loss$_d$ $= -\log\left(d_\phi(x)\right) - \log\left(1 - d_\phi\left(g_\theta(z)\right)\right)$

update ϕ to minimize loss$_d$

The aim is to converge at $d_\phi(x) = 0.5 = d_\phi(x')$

Flipped Generator loss

In practice, we often minimize $-\log\left(d_\phi(g_\theta(z))\right)$ in the generator.

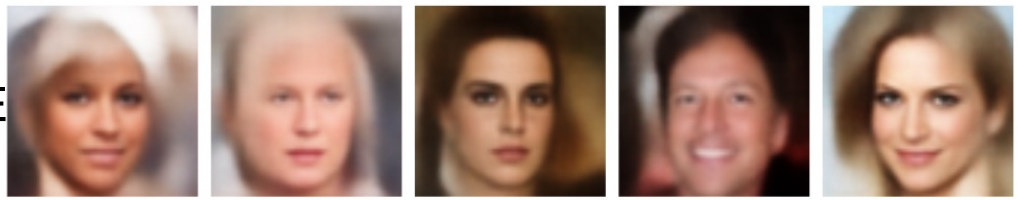


Region of bad generator performance

VAE vs. GAN

samples of VAE/GAN trained on CelebA dataset

vanilla VAE



- rather blurry
- more natural looking

vanilla GAN

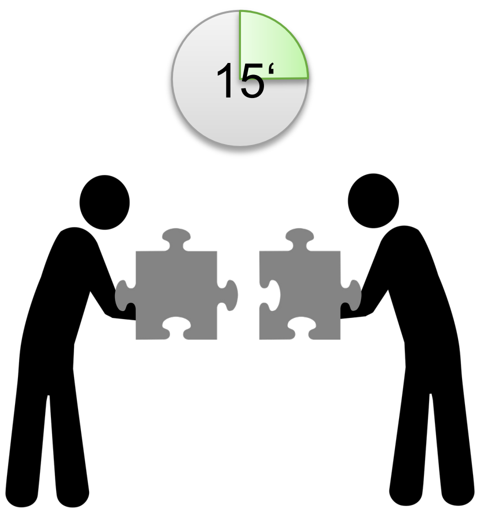
- crisper
- distortions/artifacts

[Larsen et al., 2016]

→ there are more advanced VAEs and GANs as well as VAE/GAN hybrids

VAE vs. GAN

	VAE	GAN
training	• How easy and stable is training? • Is the training related to maximum likelihood?	
latent space	• Is the latent space interpretable? • Is the latent space used everywhere? • Can you perform inference?	
results	• What are typical characteristics of the generated samples? • Which are typical behaviors where you get bad samples?	



VAE vs. GAN

VAE		GAN
rather stable trained with lower bound on likelihood	**training**	very unstable not trained based on likelihood
more interpretable possibly ignored parts of the latent space	**latent space**	not interpretable the whole latent space generates samples
more blurry depend on choice of $p(z)$ and $q(z\|x)$ posterior collapse: learns to ignore latent space and generate ”average” examples	**results**	sharper might strongly vary with each run mode collapsing, unbalanced generator/discriminator

a lot of difficulties
→ there are more advanced VAEs and GANs as well as VAE/GAN hybrids

Inference in GANs

How is inference performed in GANs?

How could you perform inference in GANs?

Outlook: GAN Variants

- In theory, GANs optimize the *Jensen-Shannon-Divergence*
- Other divergences/distances lead to other losses, e.g. Least Squares GAN, Wasserstein GAN…
- In practice, most GANs do not actually optimize a divergence/distance

Outlook: GAN Variants

- Feature matching: train D as normal; G optimizes

$$||\mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}}\mathbf{f}(\boldsymbol{x}) - \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}\mathbf{f}(G(\boldsymbol{z}))||_2^2$$

…on one or more hidden layers f (of D)

- Spectral Normalization: Limit how fast the output of D can change in response to changes in input
- Progressive Growing: Train on low-resolution data first, then successively increase the resolution
- In practice, most functioning GANs use multiple loss functions and an array of “little tricks”