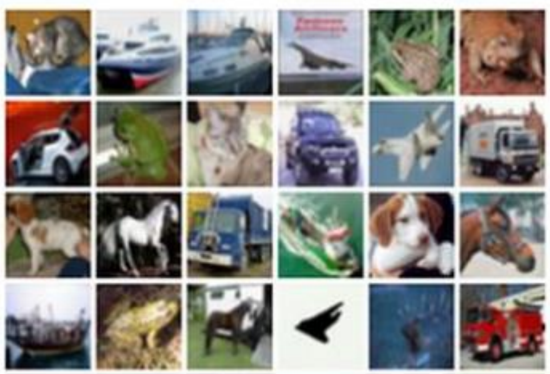
Generative Adversarial Networks

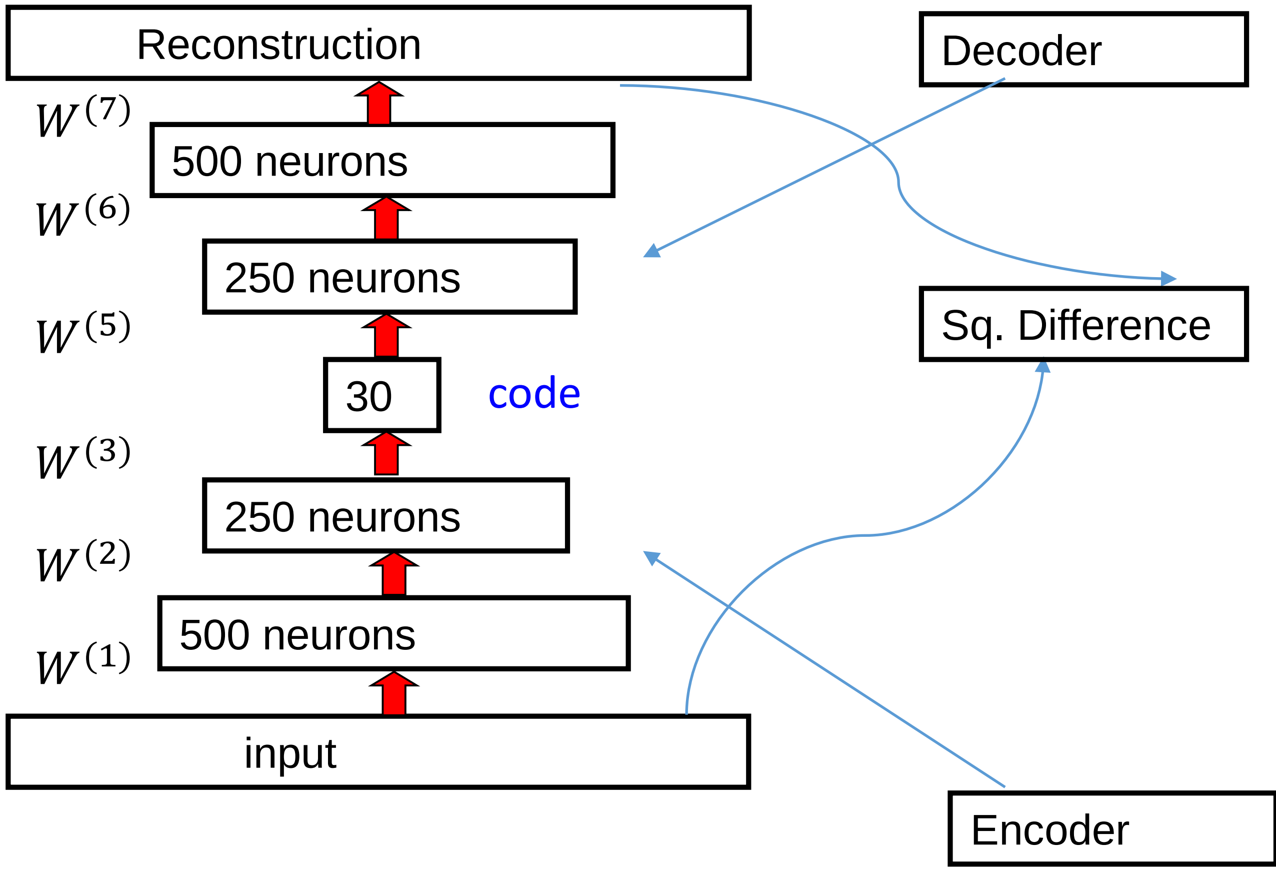
Some slides from Fei-Fei Li, Justing Johnson, Serena Yeung, and Roger Grosse, Lisa Zhang

ECE324, Winter 2020

Michael Guerzhoy

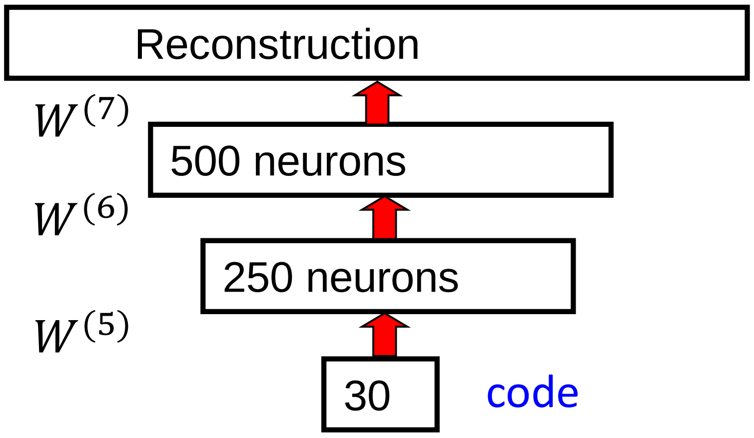
Autoencoders





- Learn to compute a code that can be used to generate a *reconstruction*
- The reconstructions are generally blurry
- Try to minimize the squared difference between the input and the reconstruction

How does the decoder work?



- $W^{(7,i,:)}$ is the i-th template for the image
- The second-to-last layer defines the coefficients for each of the templates
- The code contains information about how to compute those coefficients
  - (For faces) Whose face is it?
  - (For faces) Which way is the person looking?

Example generated images:

- Generated using a variant of autoencoders https://www.youtube.com/watch?v=XNZIN7Jh3Sg

Why are the outputs blurry for vanilla autoencoders?

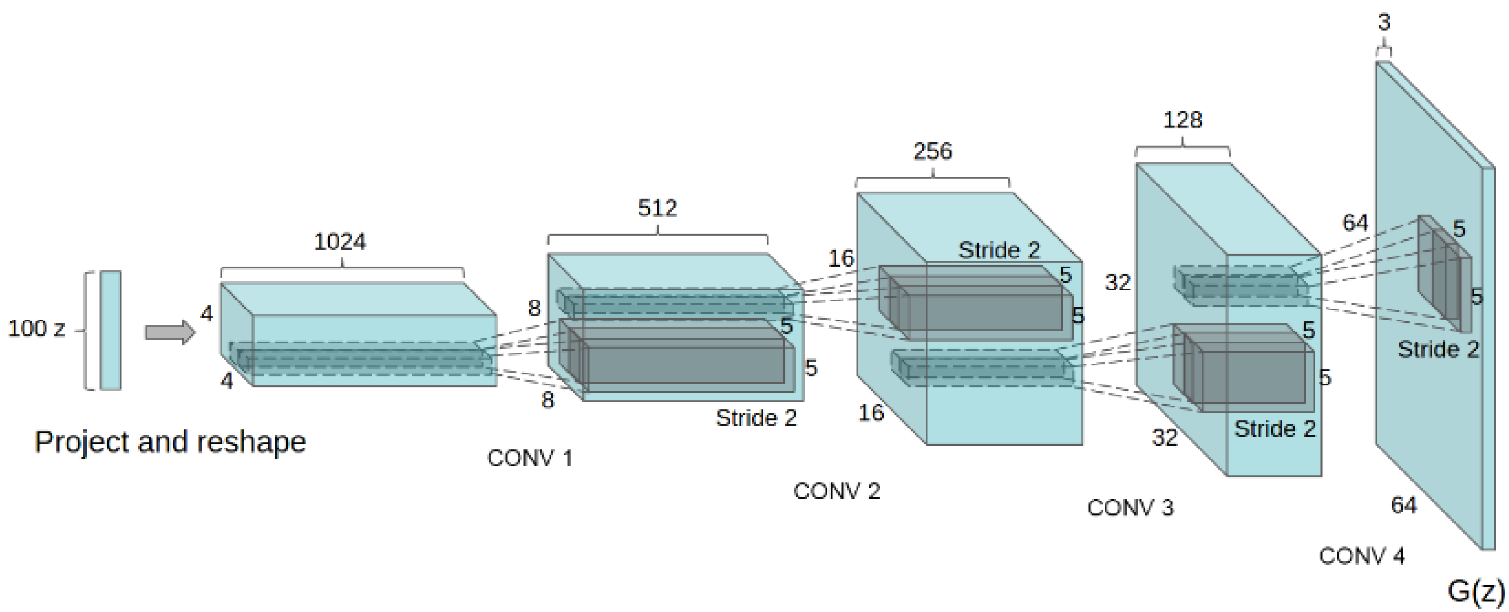
- 700 global templates isn’t bad if we want to reconstruct faces 64x64 in size
  - Don’t even need a deep architecture
- 700 global templates is pretty bad if we want to reconstruct large images
  - Want to get the details in the image right

Local & Hierarchical Templates

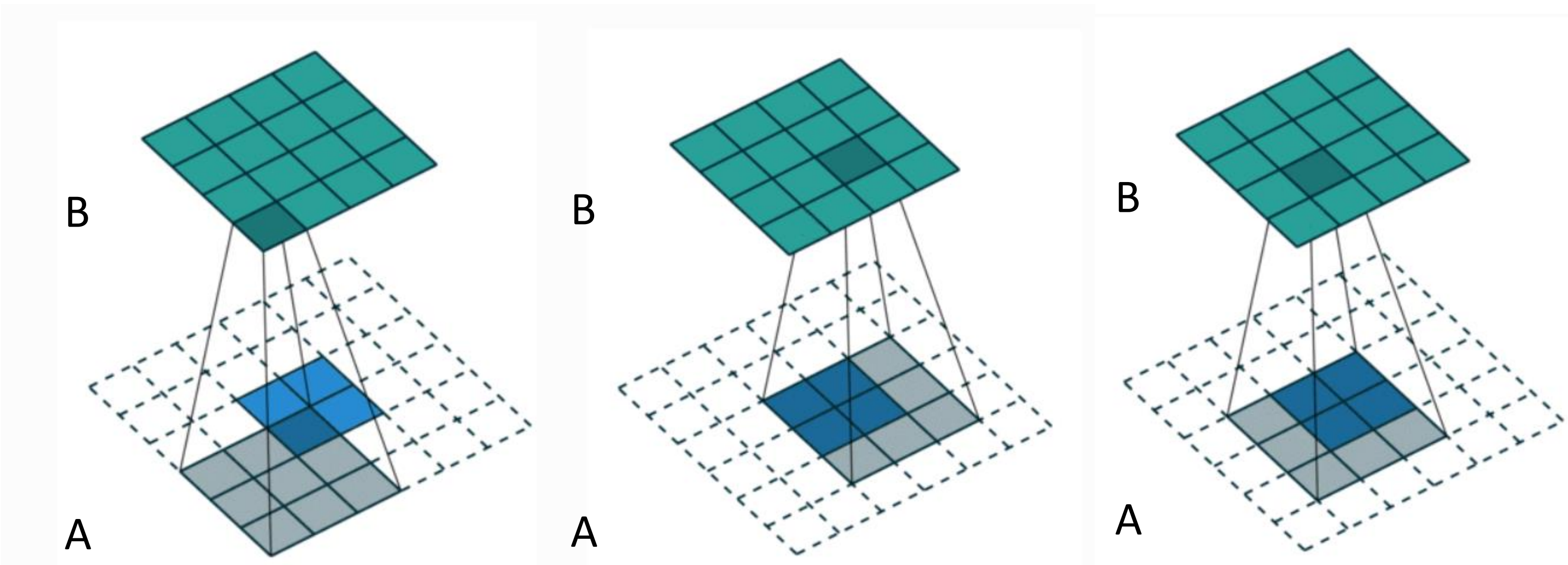
- Want to start with the code and build up the output images
- Want to build up the image from *local* templates
  - Stitch the images together from plausible image patches instead of averaging global templates
  - Want to output an image of an eye at potentially a lot of locations, just store information about what eyes look like once.

➔ Convolutions.

A generator with fractionally-strided convolutions

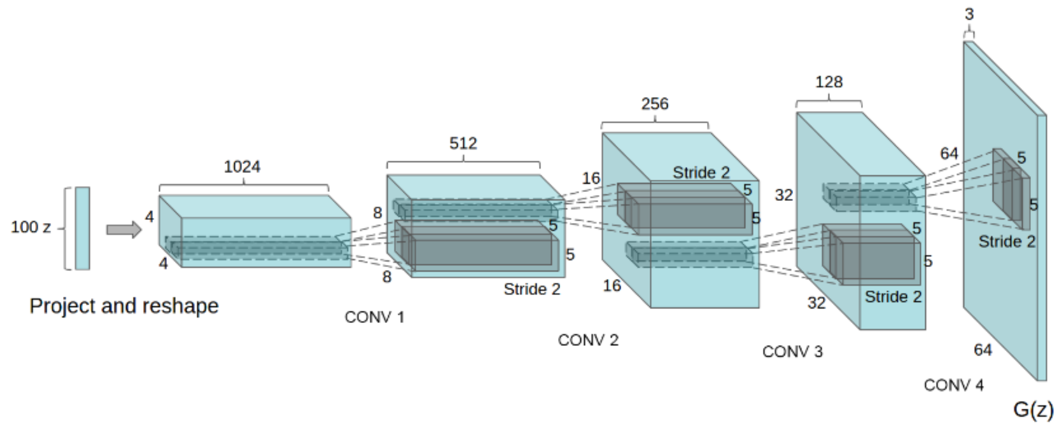


Partially-strided convolutions

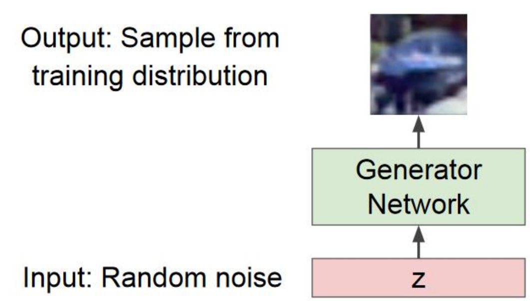


- Can get a 4×4 output from a 2×2 input by zero padding
- A more efficient way of accomplishing the same thing:
  - If a convolution can be computed using $A = CB$ where C is a large matrix (the weights of the *convolution kernel* arranged so that things work out), we can compute $C^T A$ to get a matrix that has the same size as B

A probabilistic generative model



- To generate a random image
  - Sample $z \sim N(0, I)$
    - Each coordinate in z determines the content of the image
  - Run the z though the decoder

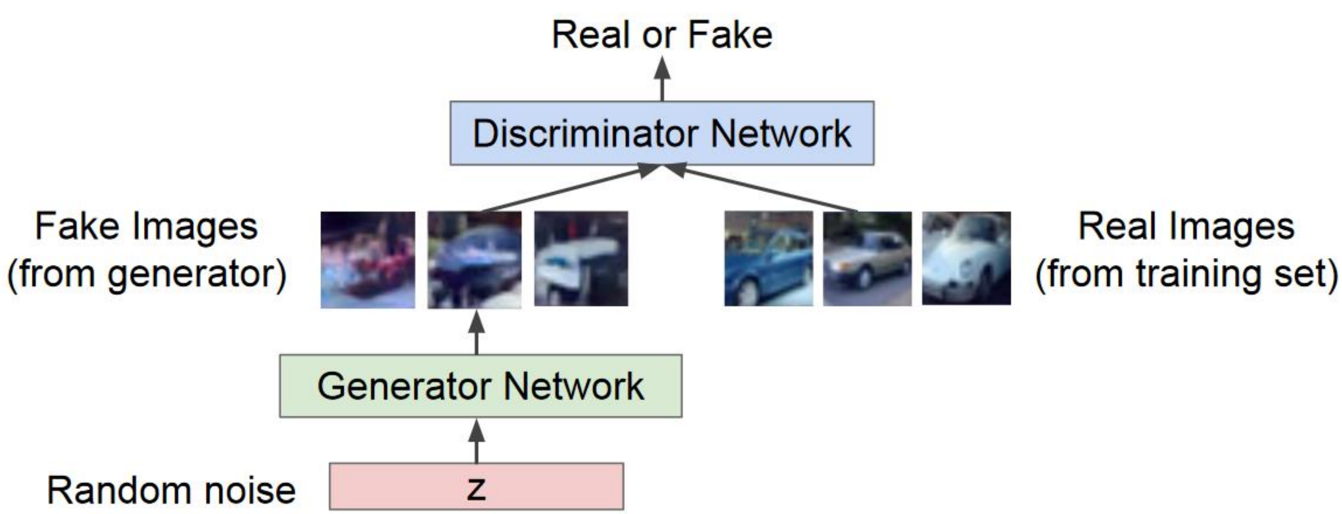


Training deep autoencoders

- Training deep autoencoders is difficult and doesn’t work very well
- Convolutions and down-sampling means exact location information is lost
- An active research area

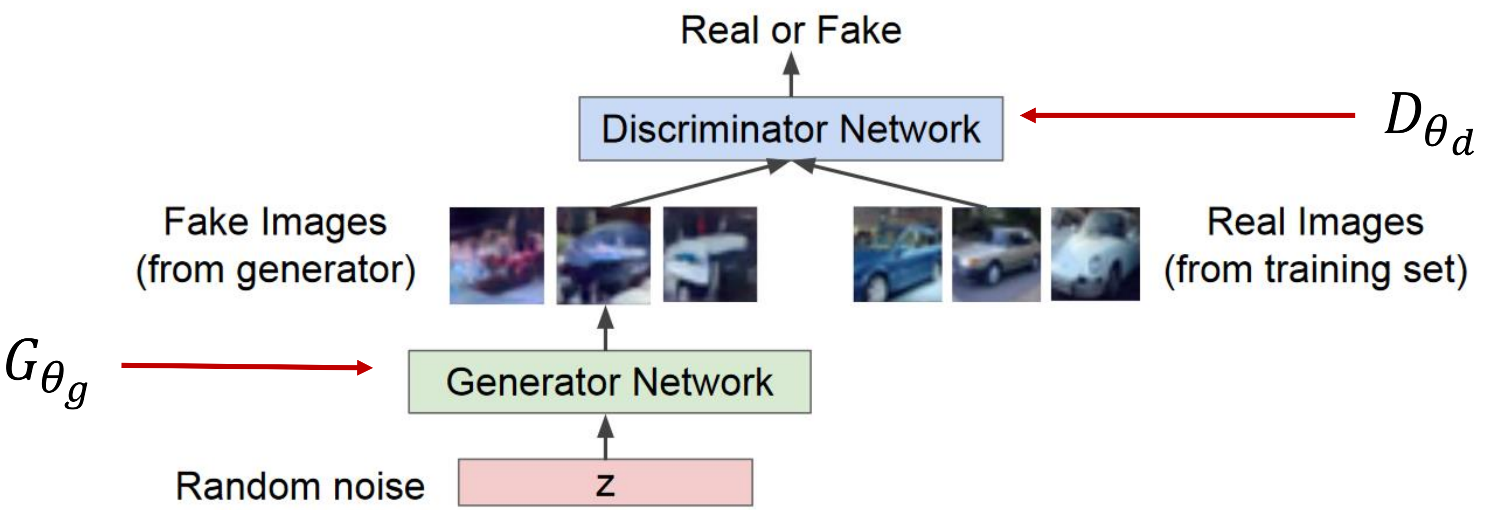
Generative Adversarial Nets (GANs)

- Idea: train two networks
  - **Generator network**: try to fool the discriminator by generating real-looking images
  - **Discriminator network**: try to distinguish between real and fake images



Training GANs: Two-Player Game

- Play a minimax game: given that the discriminator will try to do the best job it can, the generator is set to make the discriminator as wrong as possible.
- The discriminator outputs a probability

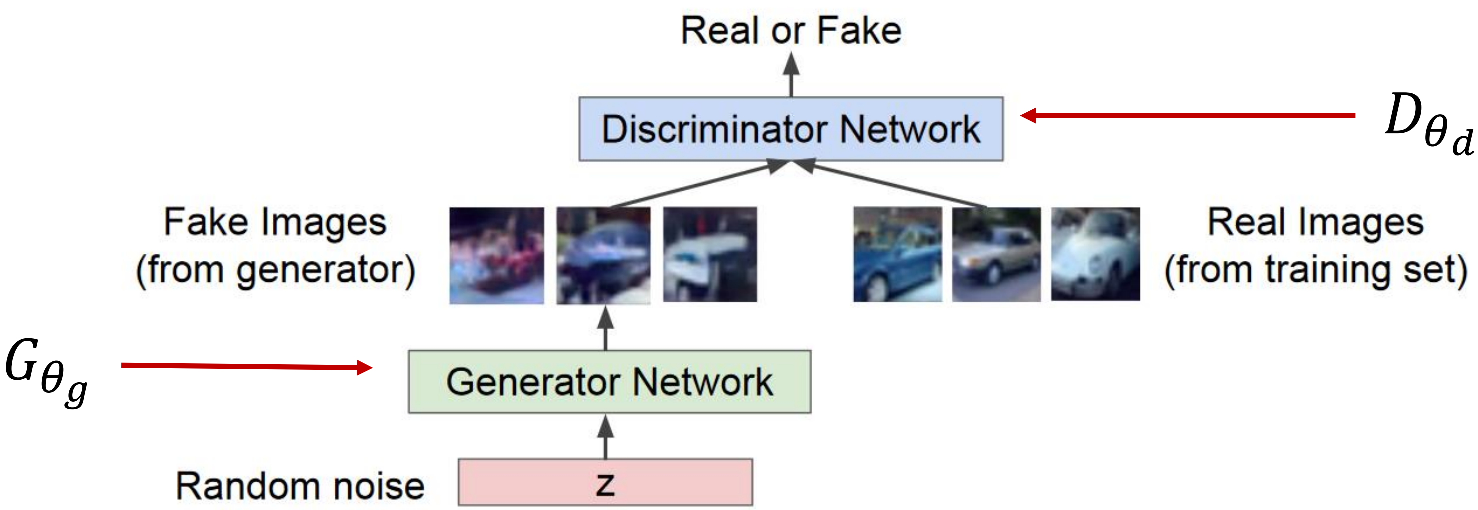


Training GANs: Two-Player Game

$$\min_{\theta_g}\max_{\theta_d}[E_{x\sim p_{data}}\log D_{\theta_d}(x) + E_{z\sim p(z)}\log\left(1 - D_{\theta_d}\left(G_{\theta_g}(z)\right)\right)]$$

x is randomly sampled from the training data. The discriminator wants to output 1

z is randomly sampled, and then a fake image is generated by the generator from the code z. The discriminator wants to output 0



Training GANs: a Two-player game

$$\min_{\theta_g} \max_{\theta_d} [E_{x \sim p_{data}} \log D_{\theta_d}(x) + E_{z \sim p(z)} \log\left(1 - D_{\theta_d}\left(G_{\theta_g}(z)\right)\right)]$$

- Alternate between:
  1. **Gradient ascent** for the discriminator

     $$\max_{\theta_d} [E_{x \sim p_{data}} \log D_{\theta_d}(x) + E_{z \sim p(z)} \log\left(1 - D_{\theta_d}\left(G_{\theta_g}(z)\right)\right)]$$

     Do a better job outputting 1 on real images and 0 on fake images

  2. **Gradient descent** on the generator

     $$\min_{\theta_g} E_{z \sim p(z)} \log\left(1 - D_{\theta_d}\left(G_{\theta_g}(z)\right)\right)$$

     Do a better job making sure the discriminator outputs large numbers on fake images

Modifying the cost function

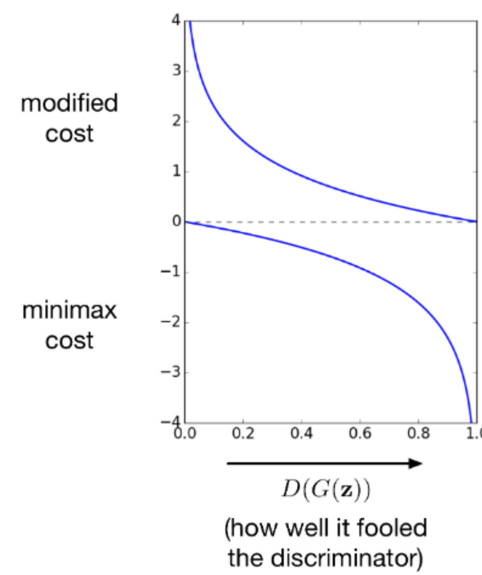
$$\min_{\theta_g} E_{z\sim p(z)} \log\left(1 - D_{\theta_d}\left(G_{\theta_g}(z)\right)\right)$$

- Problem: if the generator is doing a bad job and the discriminator knows it, it's hard to learn from that
  - Modified cost:

$$\mathrm{J} = E_{z\sim p(z)} -\log\left(D_{\theta_d}\left(G_{\theta_g}(z)\right)\right)$$

  - Now, the generator is doing poorly for code z,

$\frac{\partial J}{\partial D_{\theta_d}\left(G_{\theta_g}(z)\right)}$ is large, so that the update to θ_g is large



Training in practice

- Sample real images from the train set to estimate

$$E_{x\sim p_{data}} \log D_{\theta_d}(x) \approx \frac{1}{n}\sum_i \log D_{\theta_d}\left(x^{(i)}\right)$$

- Sample fake images (by first sampling code z and then generating images) to estimate

$$E_{z\sim p(z)} -\log\left(D_{\theta_d}\left(G_{\theta_g}(z)\right)\right) \approx -\frac{1}{n}\sum_j \log\left(D_{\theta_d}\left(G_{\theta_g}\left(z^{(j)}\right)\right)\right)$$

- Can compute the gradients now!

Training GANs

for number of training iterations **do**

for k steps **do**

• Sample minibatch of m noise samples $\{\boldsymbol{z}^{(1)}, \dots, \boldsymbol{z}^{(m)}\}$ from noise prior $p_g(\boldsymbol{z})$.

• Sample minibatch of m examples $\{\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(m)}\}$ from data generating distribution $p_{\text{data}}(\boldsymbol{x})$.

• Update the discriminator by ascending its stochastic gradient:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} \left[\log D_{\theta_d}(x^{(i)}) + \log(1 - D_{\theta_d}(G_{\theta_g}(z^{(i)}))) \right]$$

end for

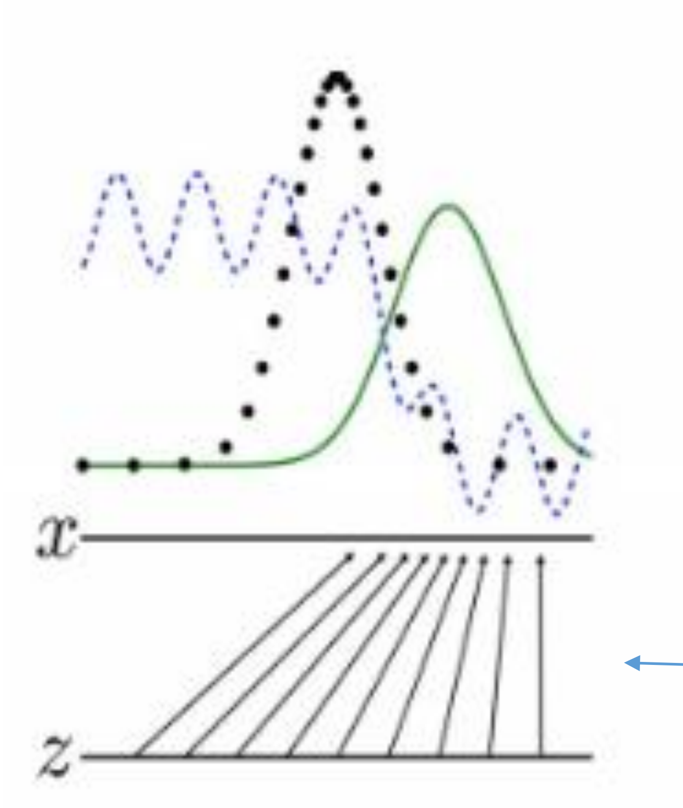
• Sample minibatch of m noise samples $\{\boldsymbol{z}^{(1)}, \dots, \boldsymbol{z}^{(m)}\}$ from noise prior $p_g(\boldsymbol{z})$.

• Update the generator by ascending its stochastic gradient (improved objective):

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} \log(D_{\theta_d}(G_{\theta_g}(z^{(i)})))$$

end for

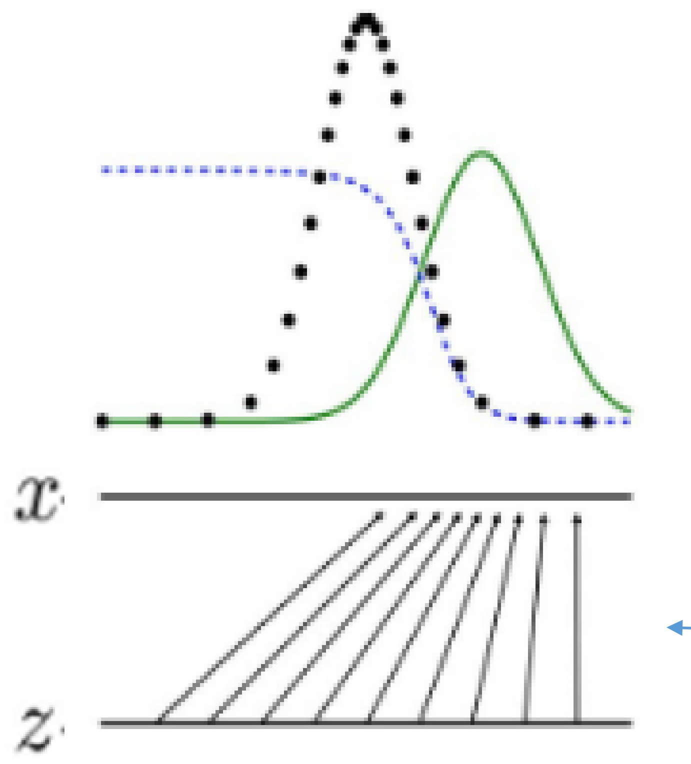
Training a GAN



Generator

Real samples ······ are far away from generated sample distribution —
The discriminator probability ··· is high for real samples and low for fake samples

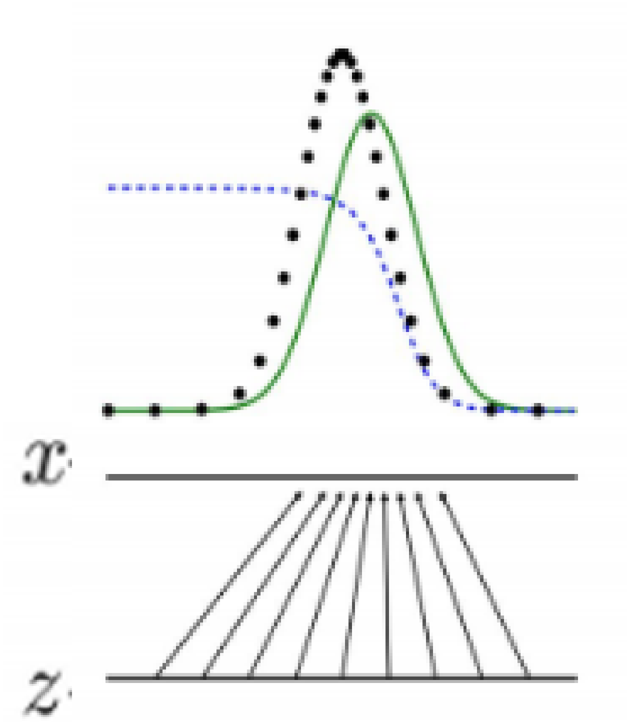
Training a GAN



Generator

Real samples ······ are far away from generated sample distribution —
The discriminator probability ··· is high for real samples and low for fake samples

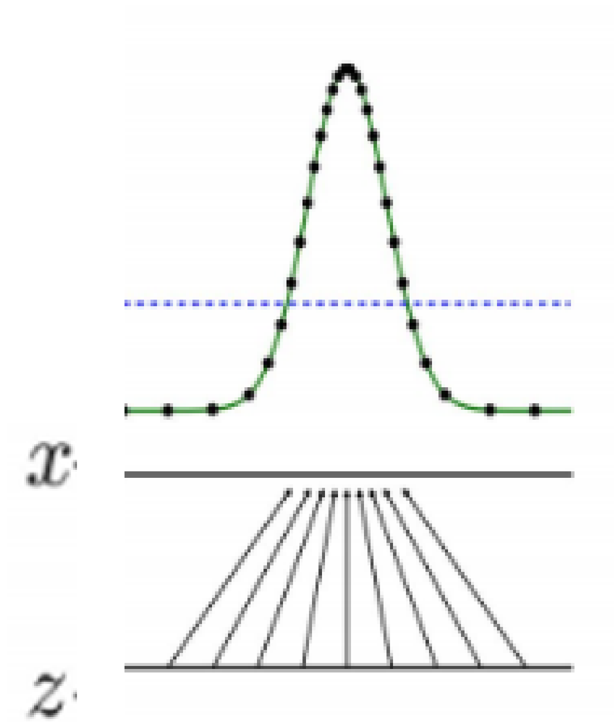
Training a GAN



Generator

Real samples ······ are close to the generated sample distribution —
The discriminator probability ··· is high for most real samples but not all and high for some fake samples

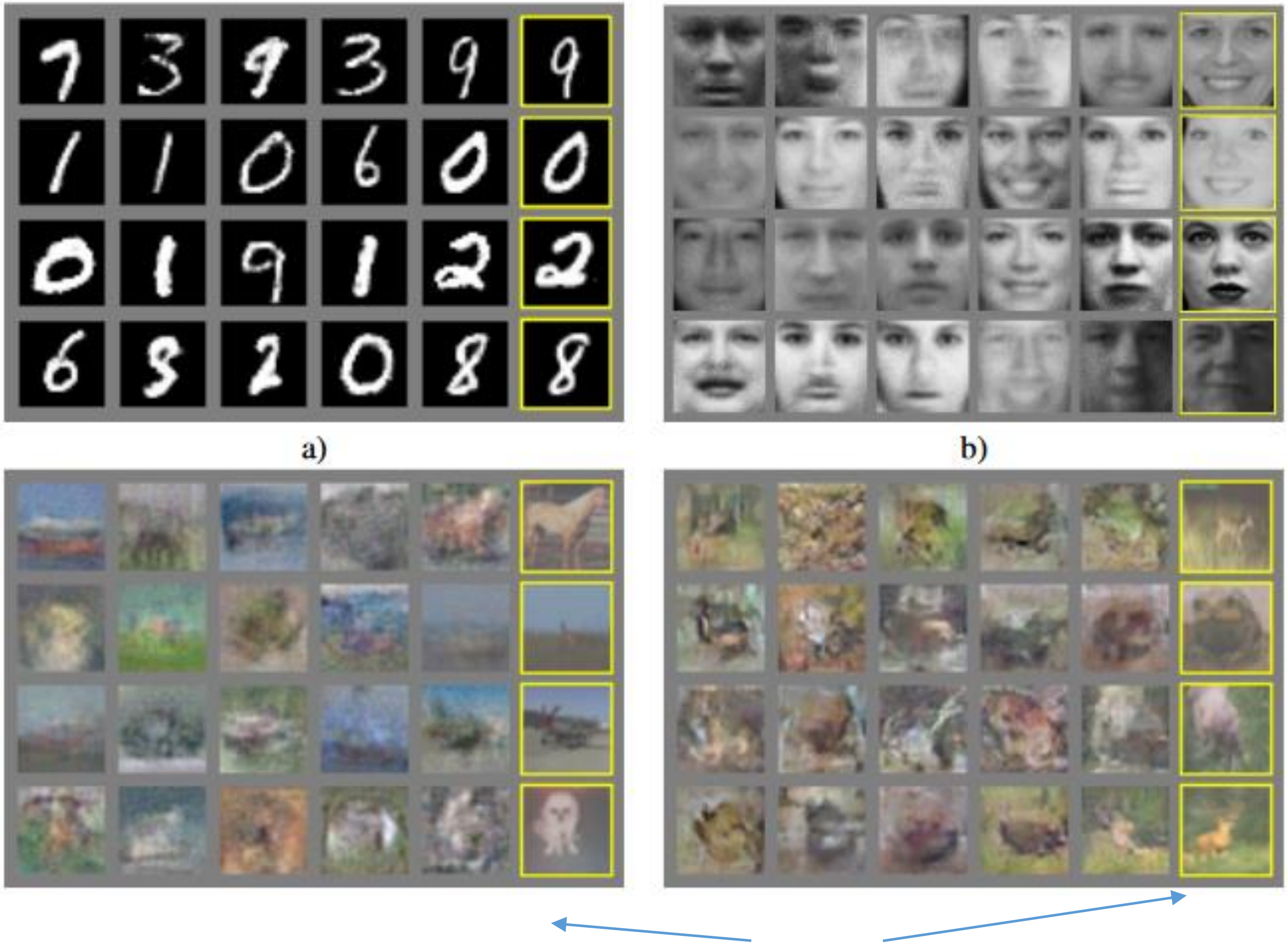
Training a GAN



Generator

Real samples ······ are very close to the generated sample distribution —
The discriminator probability ··· is constant since the discriminator can’t tell real from fake samples

Initial results: Generated Images



Nearest examples from train set

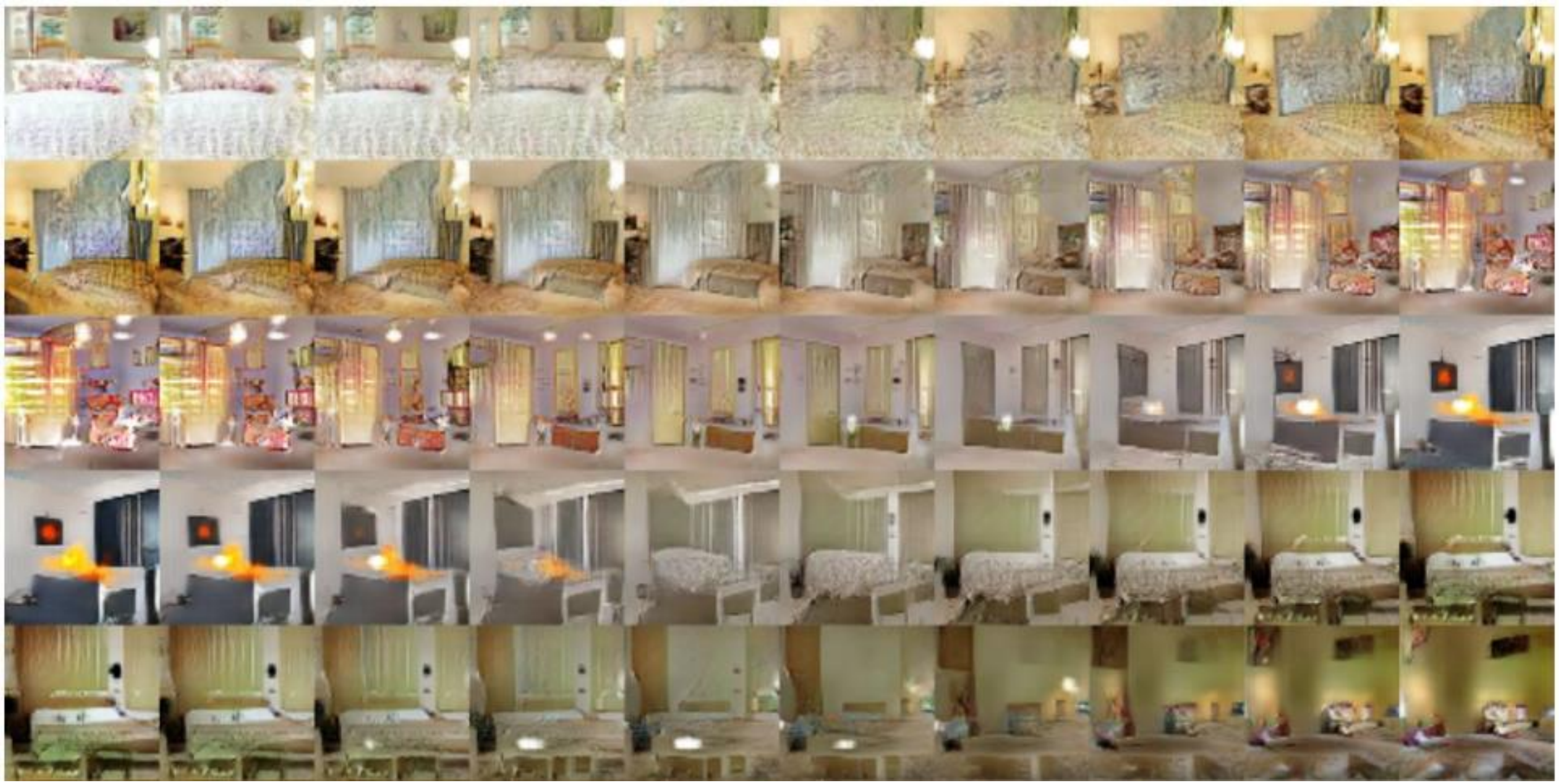
Goodfellow et al. 2014

Convolutional Architecture: Generated images

Radford et al. 2016

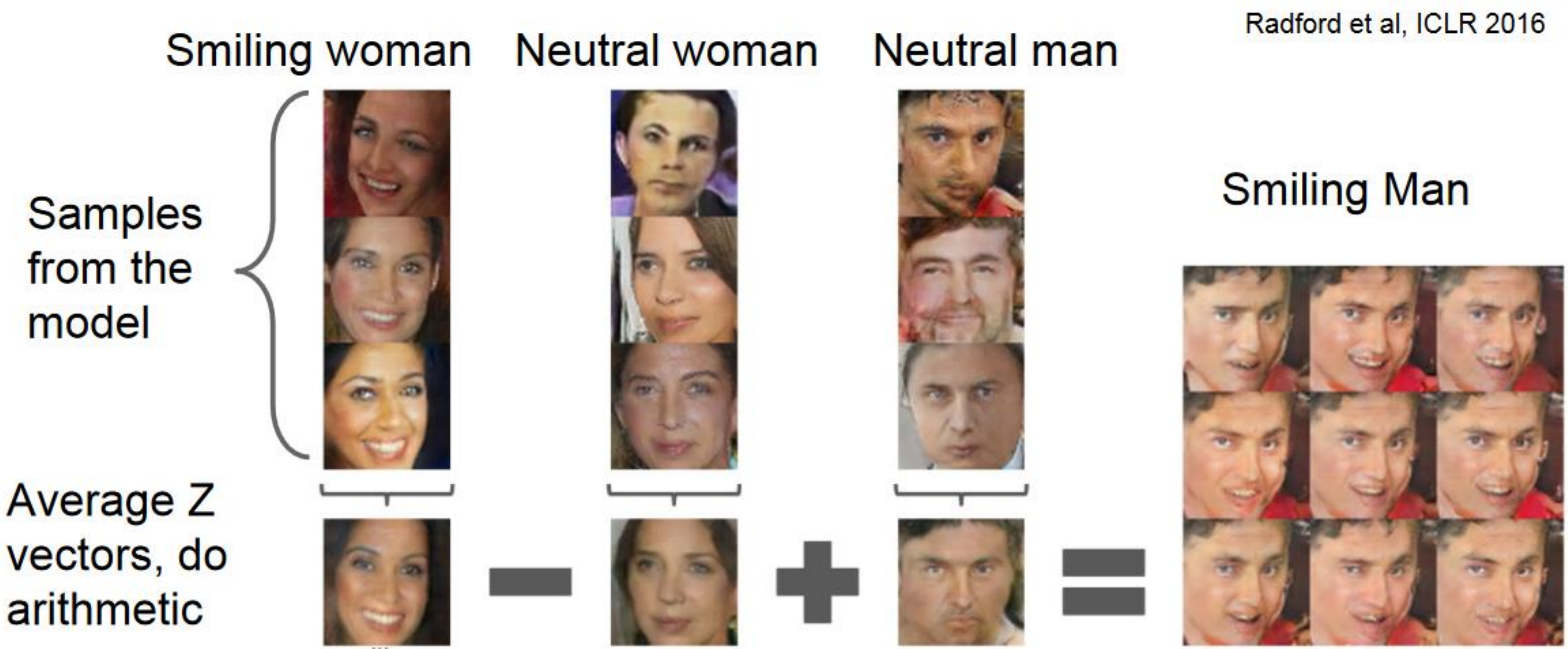
Interpolating between random points in latent (z) space

$z = z_0$ $z = 0.2z_0 + 0.8z_1$ $z = z_1$



Radford et al. 2016

Vector Arithmetic in z space



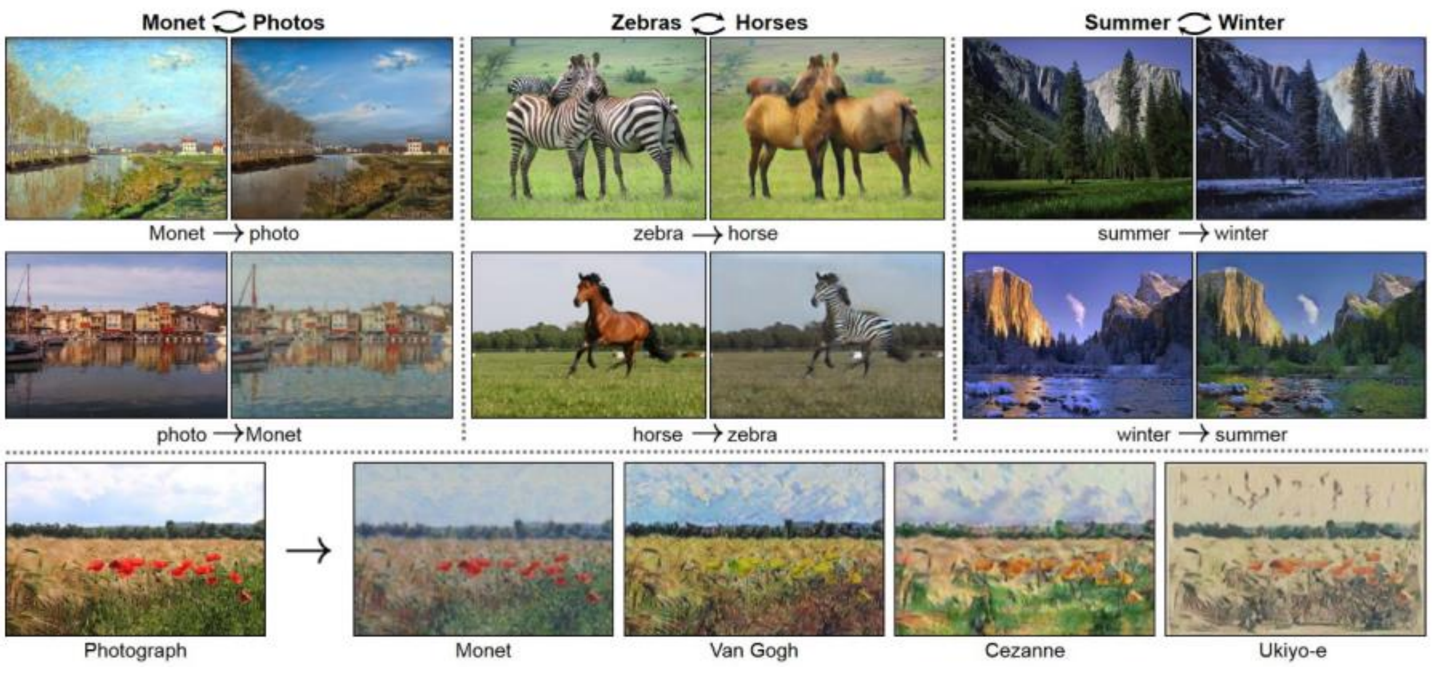
Vector arithmetic in z space

GANs in practice

- Difficult to train (VERY difficult!)
- Difficult to numerically see whether there is progress
  - Plotting the “learning curve” (the minmax objective function) doesn’t help too much
- Difficult to generate globally consistent structure

CycleGAN

- Style transfer problem: change the style of an image while preserving the content

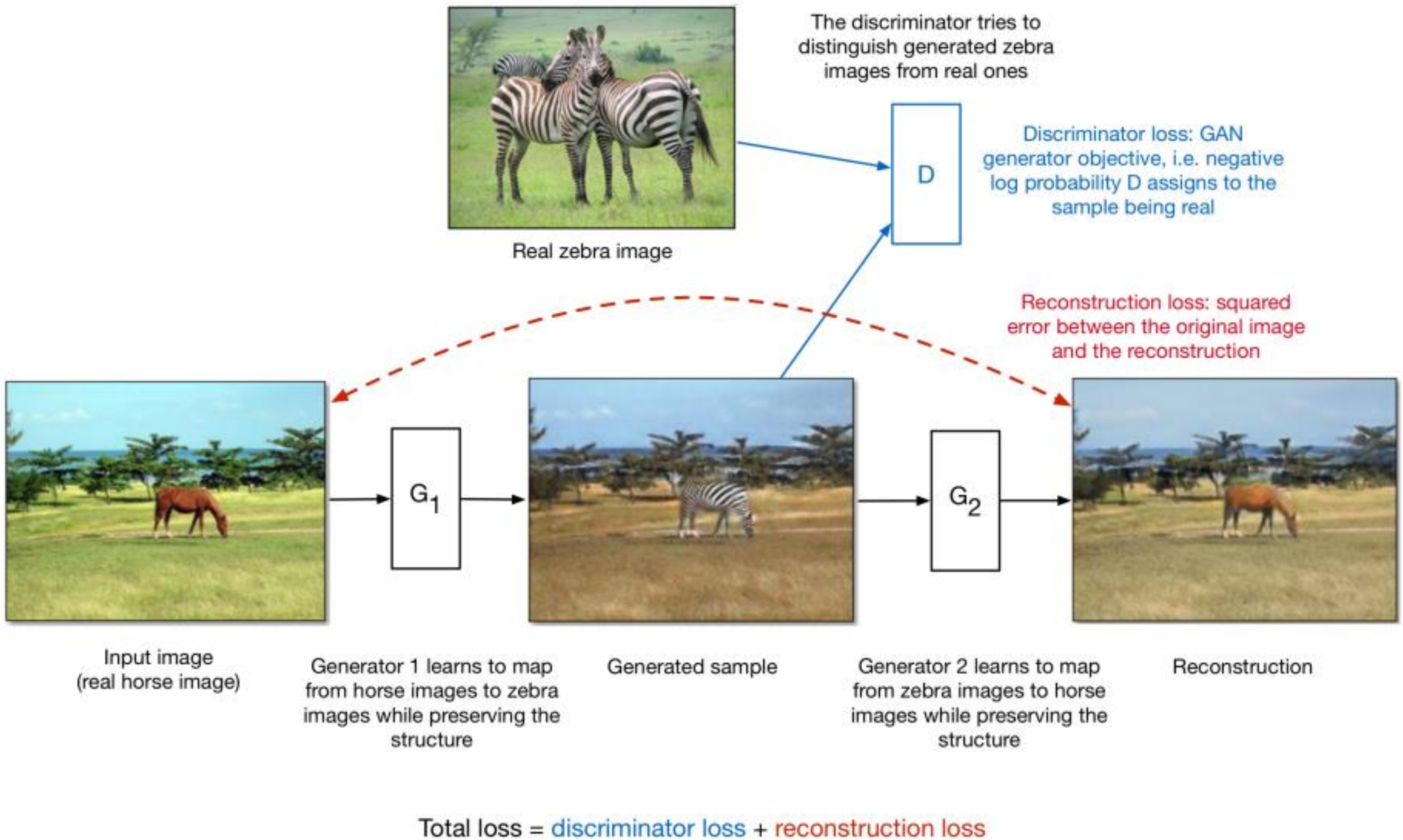


Data: Two unrelated collections of images, one for each style

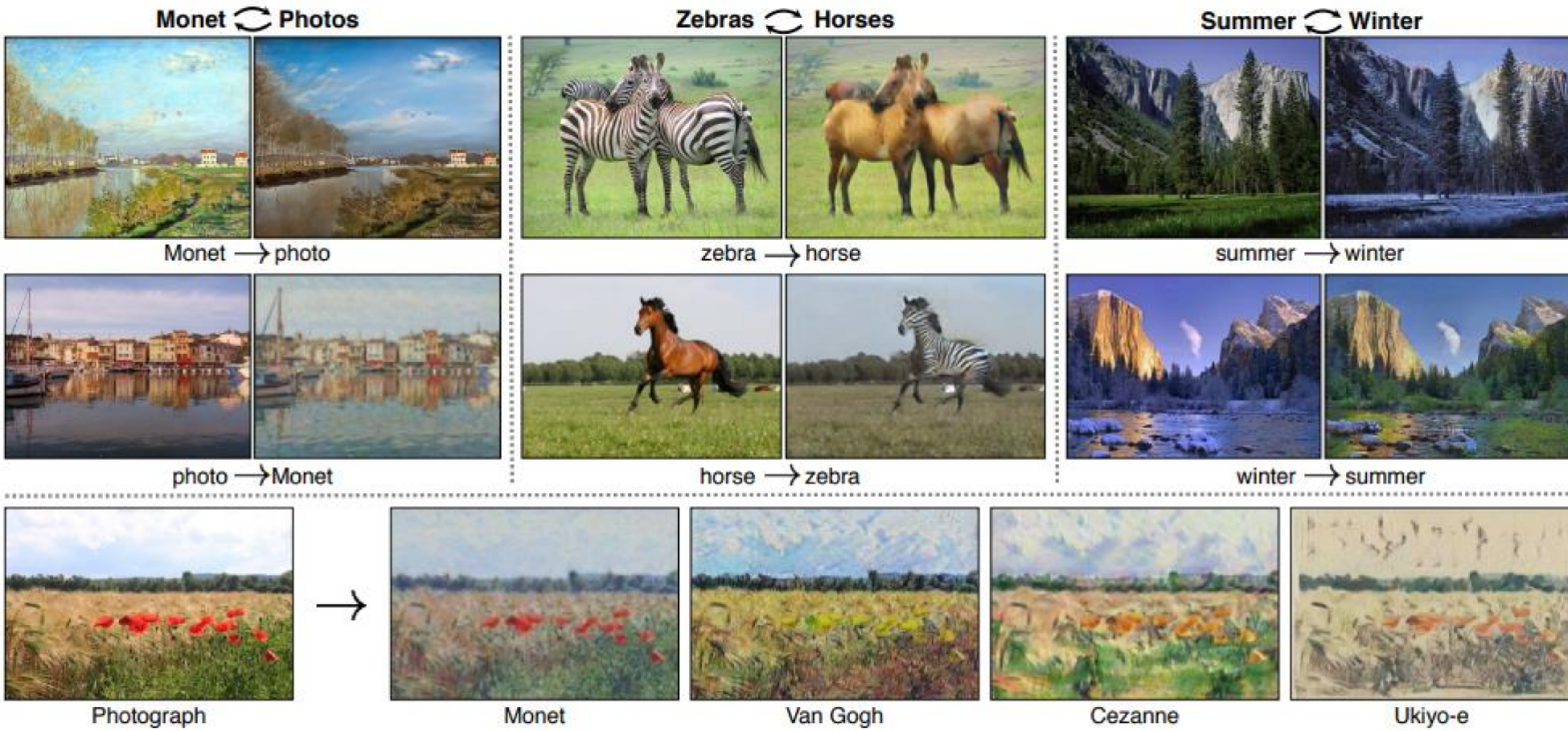
CycleGAN

- If we had paired data (same content in both styles), this would be a supervised learning problem. But this is hard to find.
- The CycleGAN architecture learns to do style transfer from unpaired data.
  - Train two different generator nets to go from style 1 to style 2, and vice versa.
  - Make sure the generated samples of style 2 are indistinguishable from real images by a discriminator net
  - Make sure the generators are cycle-consistent: mapping from style 1 to style 2 and back again should give you almost the original image.

CycleGAN



The discriminator tries to distinguish generated zebra images from real ones
D
Discriminator loss: GAN generator objective, i.e. negative log probability D assigns to the sample being real
Real zebra image
Reconstruction loss: squared error between the original image and the reconstruction
G1
G2
Input image (real horse image)
Generator 1 learns to map from horse images to zebra images while preserving the structure
Generated sample
Generator 2 learns to map from zebra images to horse images while preserving the structure
Reconstruction
Total loss = discriminator loss + reconstruction loss


Monet ⇄ Photos
Zebras ⇄ Horses
Summer ⇄ Winter
Monet → photo
zebra → horse
summer → winter
photo → Monet
horse → zebra
winter → summer
Photograph
Monet
Van Gogh
Cezanne
Ukiyo-e

Training

$$L_{GAN}(G, D_y, X, Y) = E_{y\sim p_{data}(y)}[\log D_y(y)] + \\ E_{x\sim p_{data}(x)}[\log(1 - D_y(G(x))]$$

$$L_{GAN}(F, D_x, Y, X) = E_{x\sim p_{data}(x)}[\log D_x(x)] + \\ E_{y\sim p_{data}(y)}[\log(1 - D_x(F(y))]$$

$$L_{cyc}(G, F) = E_{x\sim p_{data}(x)}\left[\left|F(G(x)) - x\right|_1\right] + \\ E_{y\sim p_{data}(y)}[|G(F(y)) \ - y|_1]$$

$$L(G, F, D_x, D_y) = L_{GAN}(G, D_y, X, Y) + L_{GAN}(F, D_x, Y, X) + L_{cyc}(G, F)$$

$$G^*, F^* = argmin_{G,F} \max_{D_x, D_y} L(G, F, D_x, D_y)$$