

# Autoencoders and GANs

Materials for this section:

- <https://www.deeplearningbook.org/contents/autoencoders.html>
- <https://www.youtube.com/watch?v=BUNl0To1IVw>
- <https://arxiv.org/pdf/1703.10593.pdf>
- <https://arxiv.org/pdf/1611.07004.pdf>

# Autoencoders

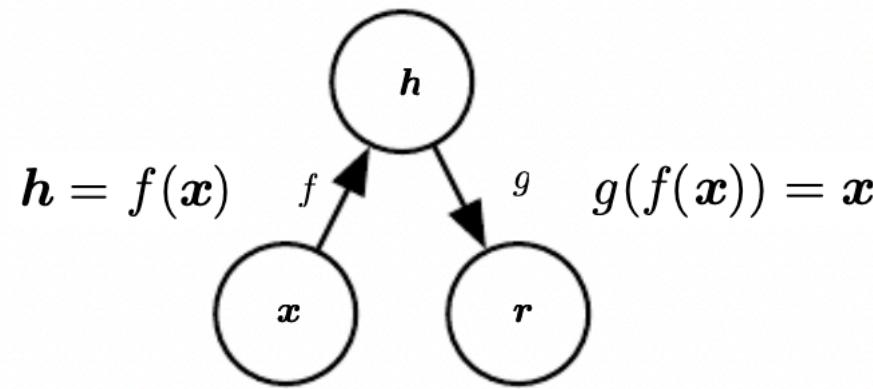
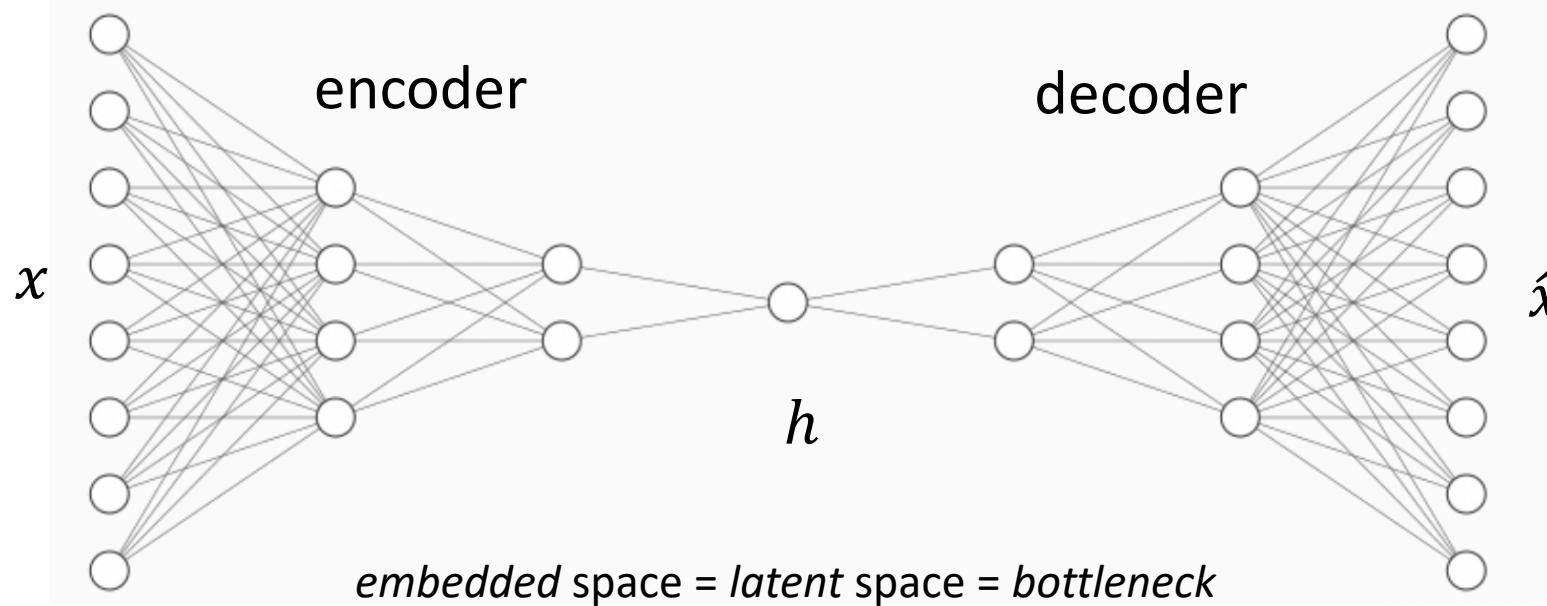


Figure 14.1: The general structure of an autoencoder, mapping an input  $\mathbf{x}$  to an output (called reconstruction)  $\mathbf{r}$  through an internal representation or code  $\mathbf{h}$ . The autoencoder has two components: the encoder  $f$  (mapping  $\mathbf{x}$  to  $\mathbf{h}$ ) and the decoder  $g$  (mapping  $\mathbf{h}$  to  $\mathbf{r}$ ).

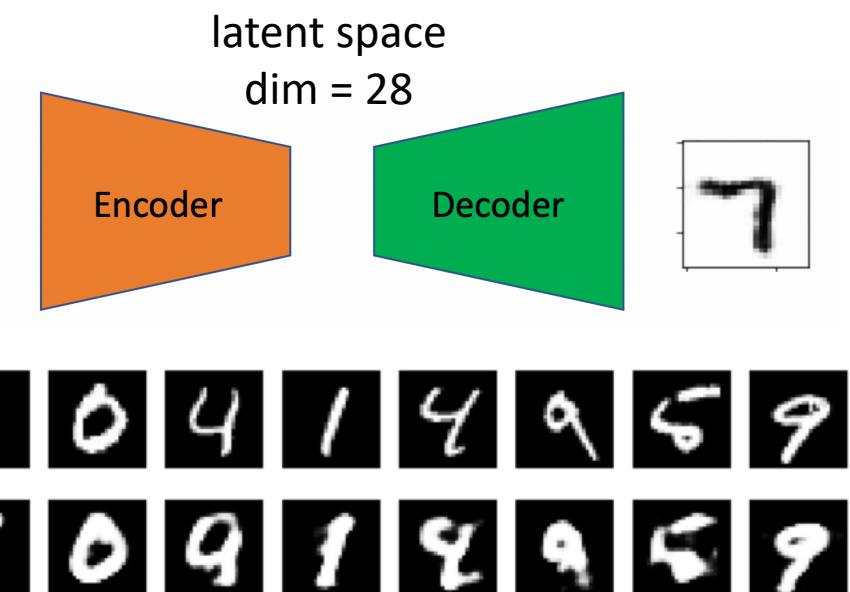
- autoencoders are unsupervised learning models

# Autoencoders



$$\mathcal{L}(x, \hat{x}) = \frac{1}{N} \sum_{i=1}^N \|x_i - \hat{x}_i\|^2$$

input: 28x28



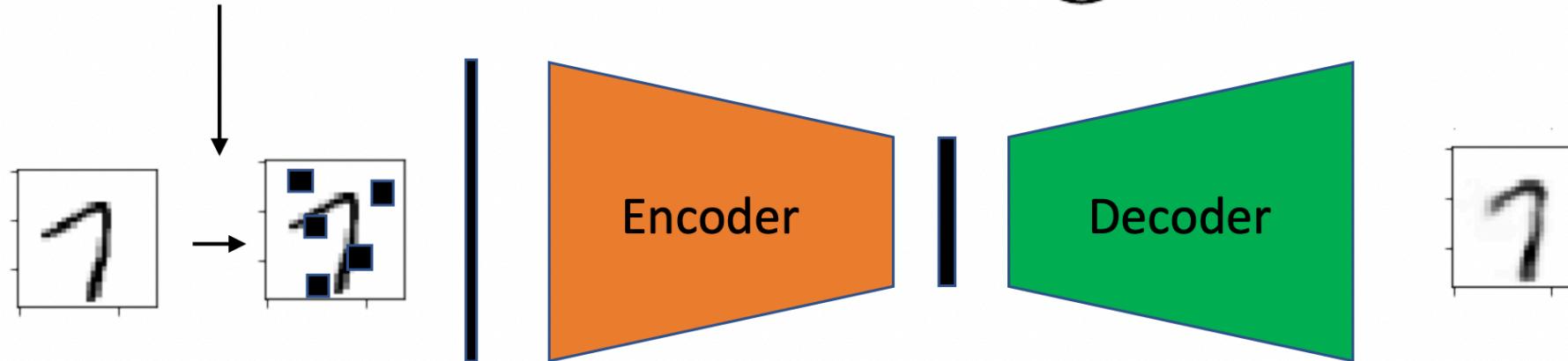
Typically, *latent space*  $h$  have a smaller dimension than input  $x$

- An autoencoder whose code dimension is less than the input dimension is called **undercomplete**

# Autoencoders

## Denoising autoencoders

Add noise to the input to learn how to denoise images



DAE minimizes:

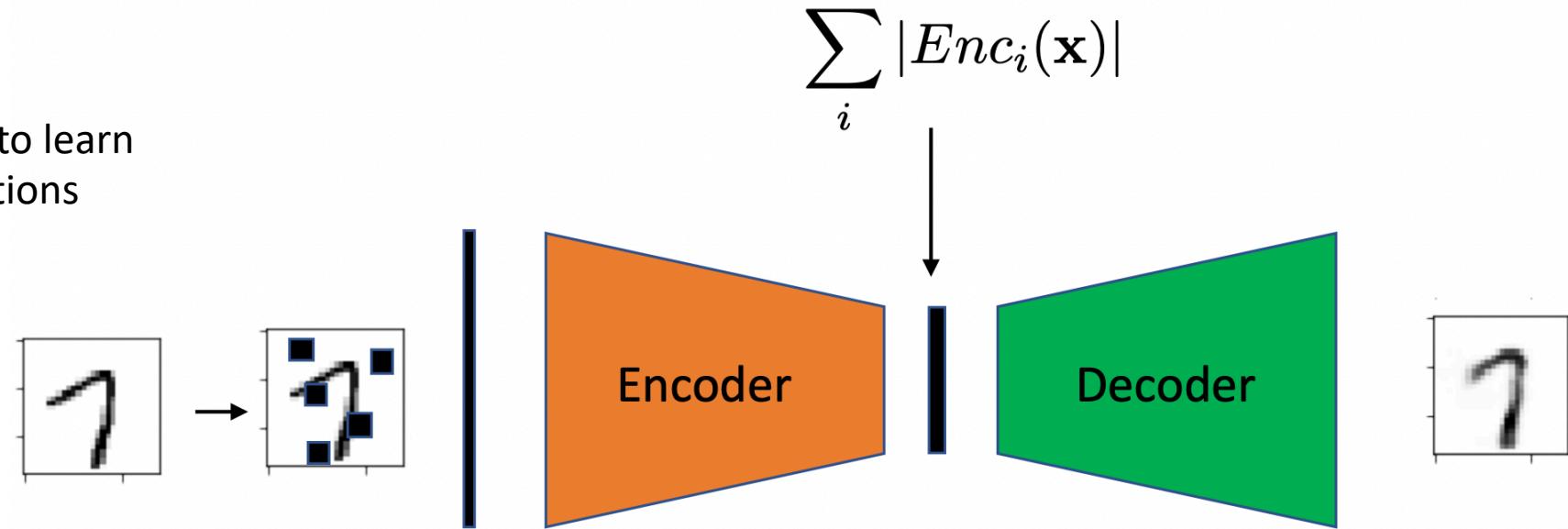
$$L(x, g(f(\tilde{x})))$$

corrupted copy of input

# Autoencoders

## Sparse autoencoders

Add  $L_1$  penalty to the loss to learn **sparse** feature representations

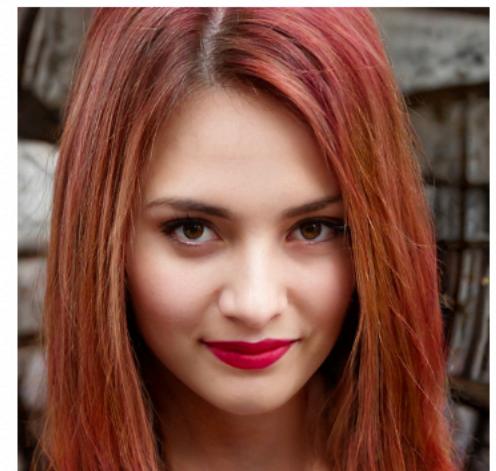
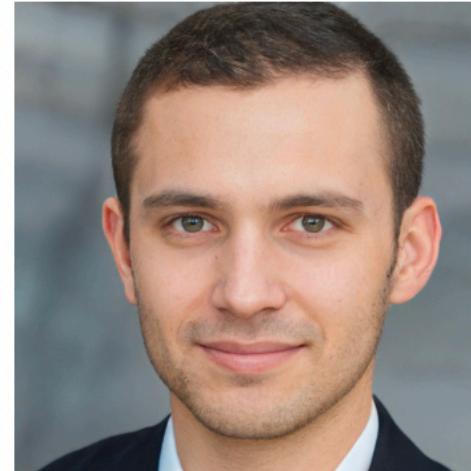


- sparsity of a latent space enforces learning only useful features of an input
- after learning we take trained encoder and use as feature extractor e.g. in classification task

$$\mathcal{L} = \|\mathbf{x} - Dec(Enc(\mathbf{x}))\|_2^2 + \sum_i |Enc_i(\mathbf{x})|$$

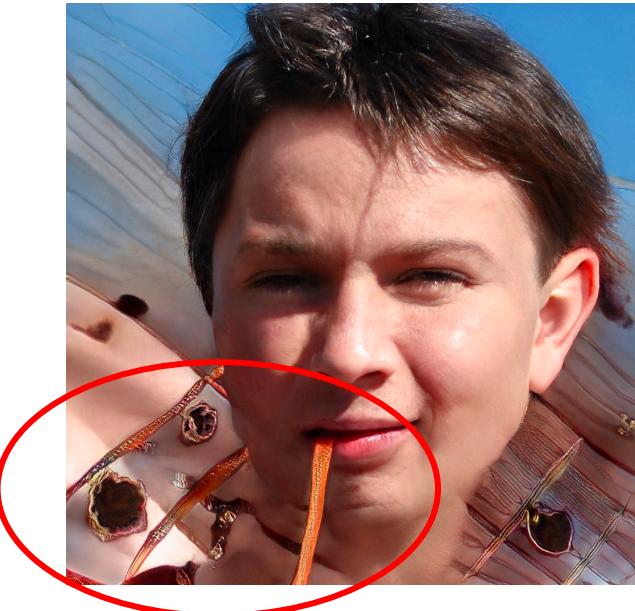
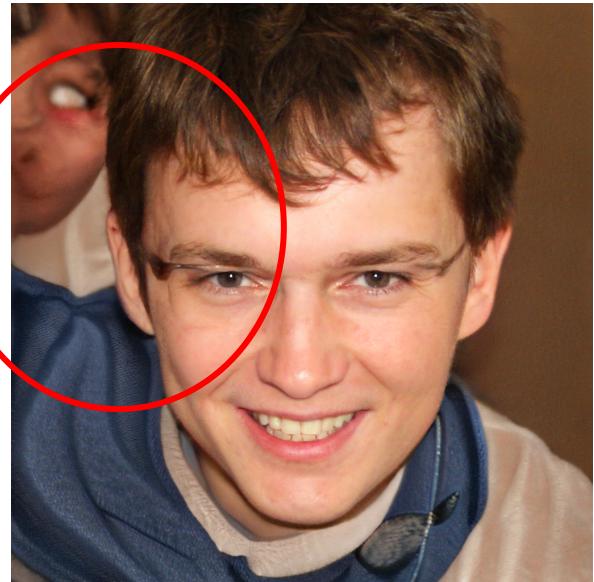
- other application: dimensionality reduction

# Generative models



Which one is unreal?

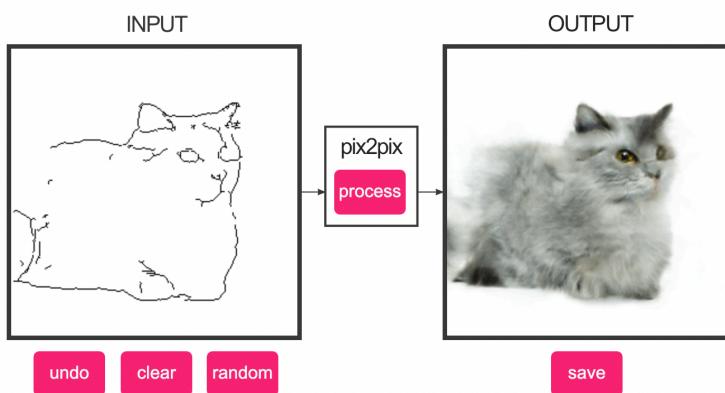
# Generative adversarial networks



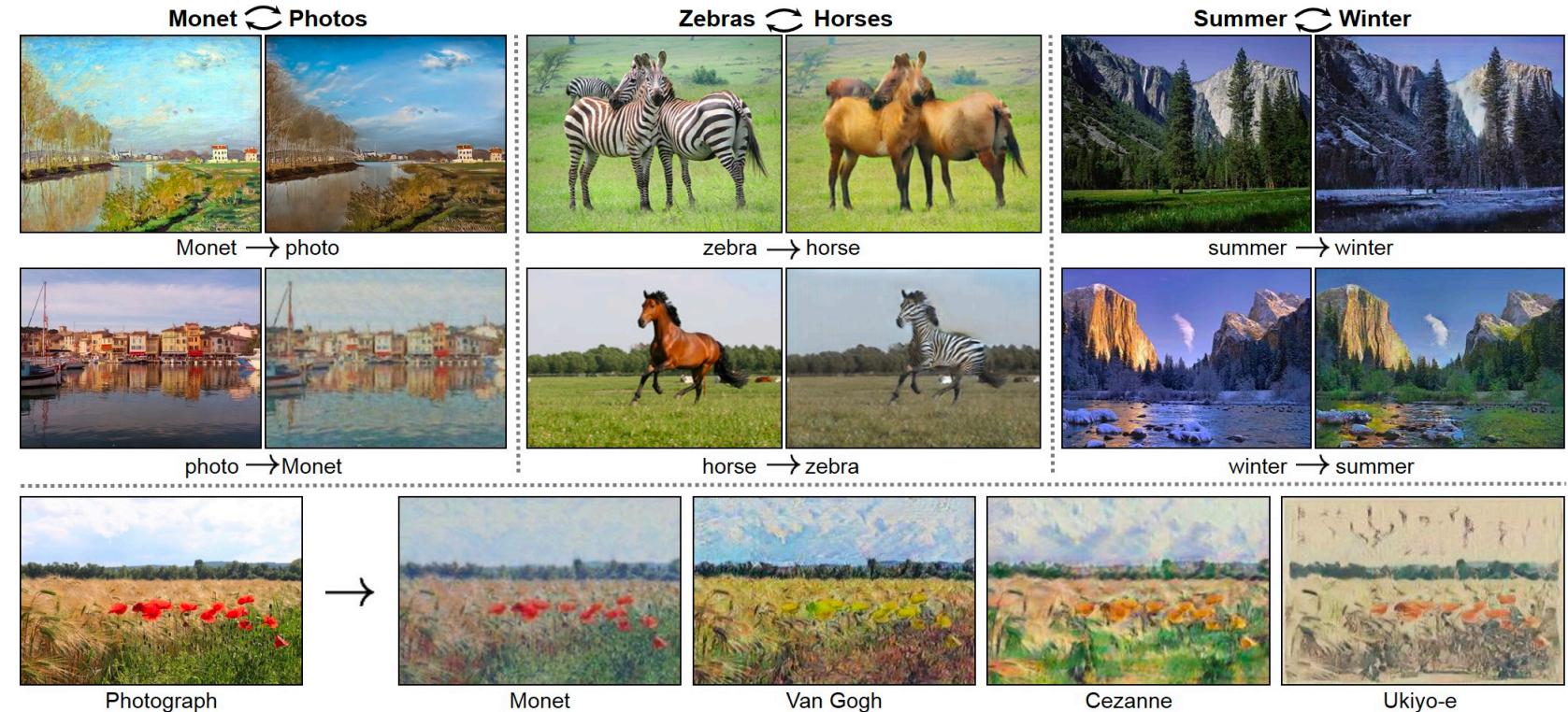
- artifacts

<https://thispersondoesnotexist.com>

# Generative adversarial networks



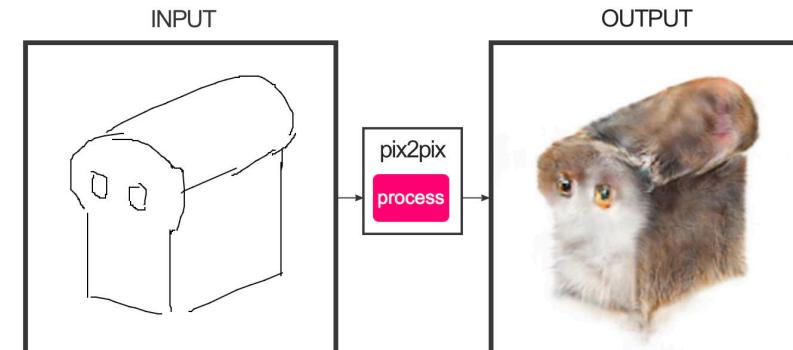
<https://affinelayer.com/pixsrv/>



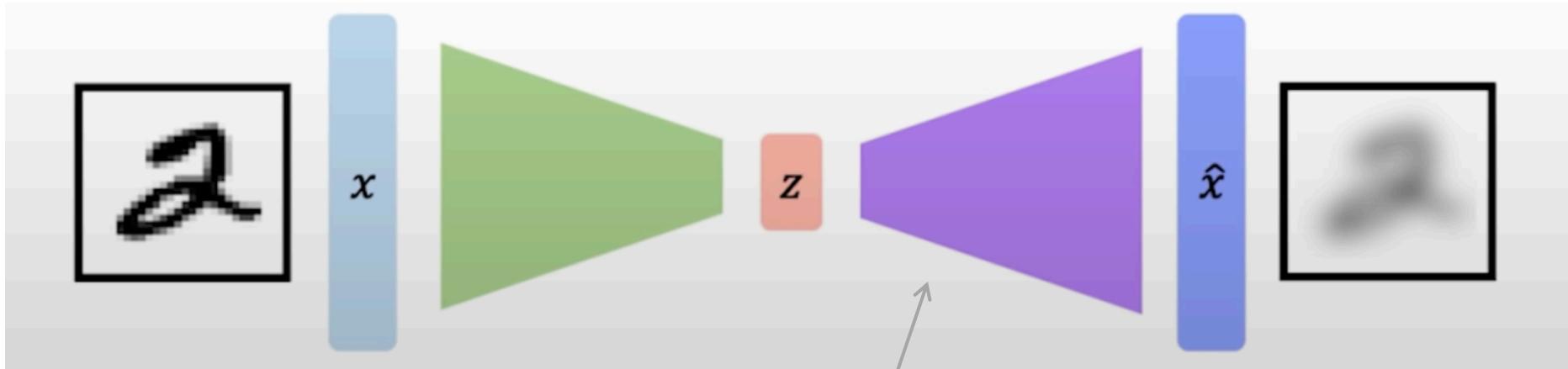
# Generative adversarial networks

GAN = letting two neural networks **compete** with each other to **generate** new data

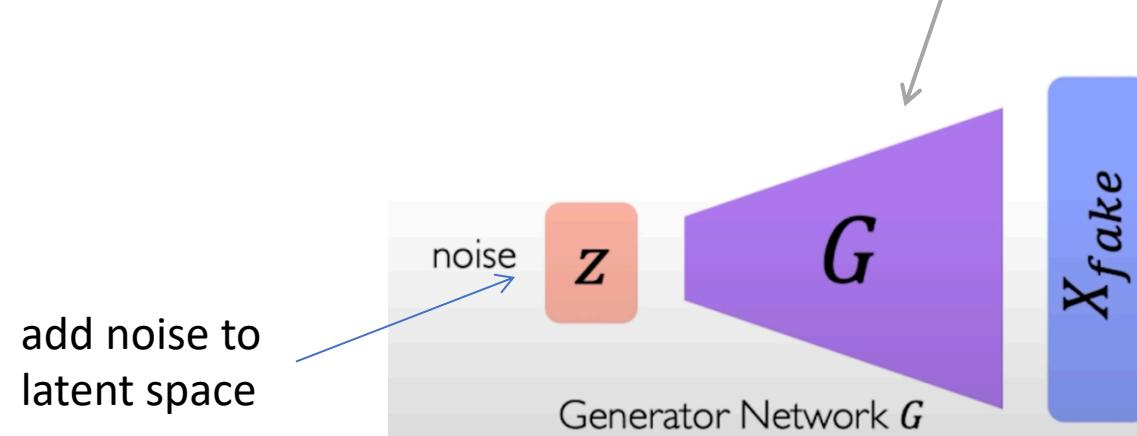
- The original purpose is to generate new data
- Classically for generating new images, but applicable to wide range of domains
- Learns the training set distribution and can generate new images that have never been seen before
- In contrast to e.g., autoregressive models or RNNs (generating one word at a time), GANs generate the whole output all at once



# Generative adversarial networks



well organized latent space from which we can sample -> Variational Autoencoders

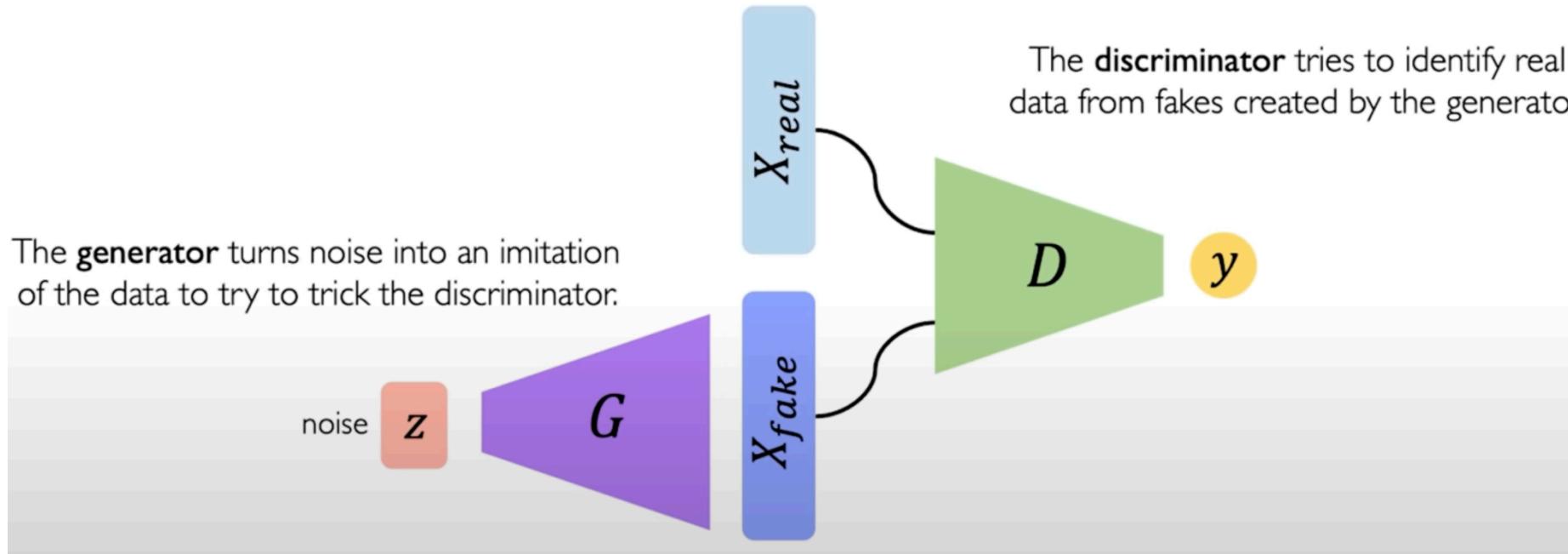


add noise to  
latent space

[youtube.com/watch?v=BUNI0To1IVw](https://youtube.com/watch?v=BUNI0To1IVw)

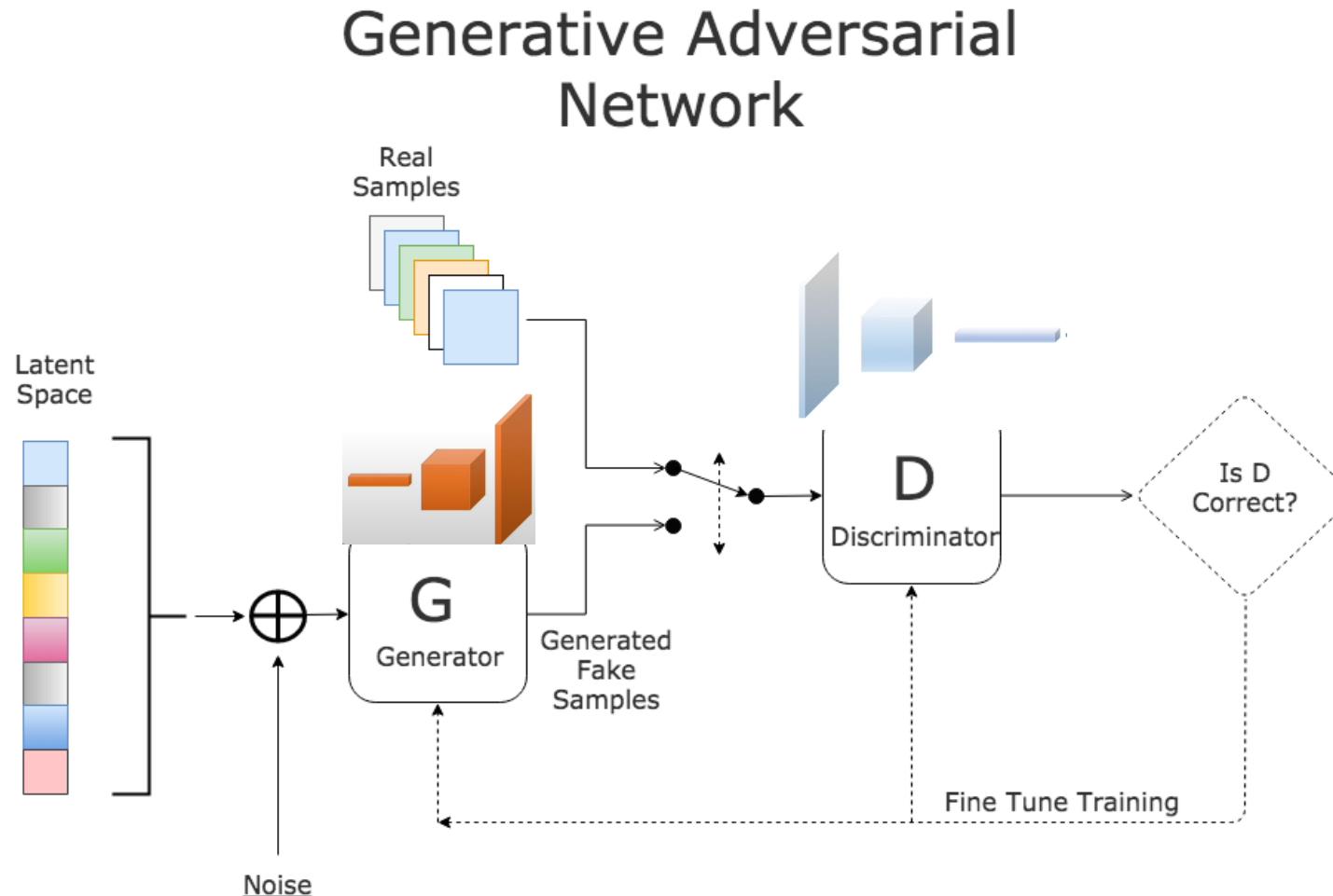
# Generative adversarial networks

The **generator** turns noise into an imitation of the data to try to trick the discriminator.



[youtube.com/watch?v=BUNI0To1IVw](https://youtube.com/watch?v=BUNI0To1IVw)

# Generative adversarial networks



# Generative adversarial networks

## Training GANs

$$\arg \min_G \max_D \mathbb{E}_{\mathbf{z}, \mathbf{x}} [ \log D(G(\mathbf{z})) + \log (1 - D(\mathbf{x})) ]$$

saddle point

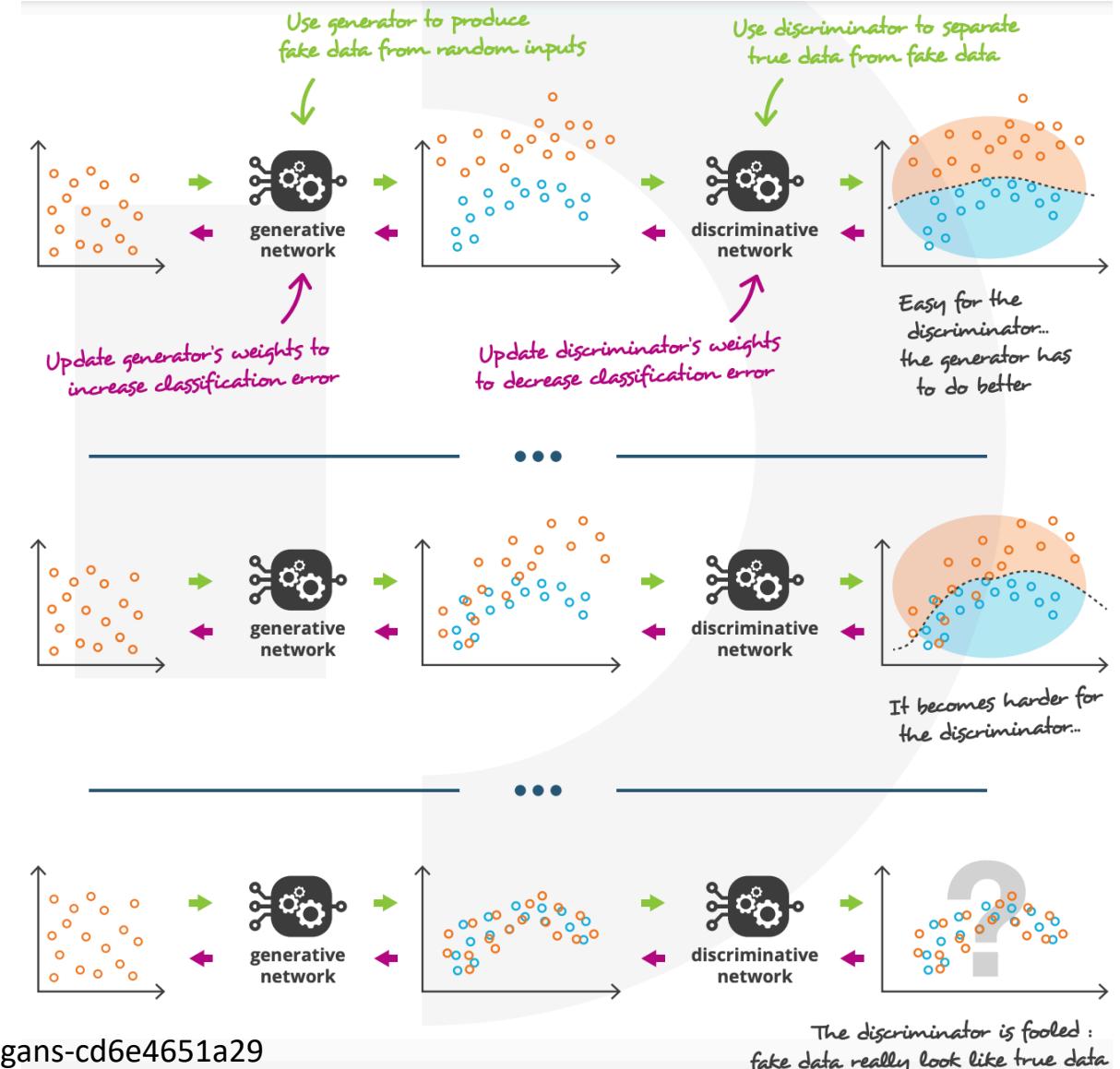
probability  
that D-fake is  
actually fake

D-real is real

Convergence means that the G produces realistic images and the D outputs random predictions (probabilities close to 0.5)

Training problems:

- oscillation between generator and discriminator loss
- mode collapse (generator produces examples of a particular kind only)

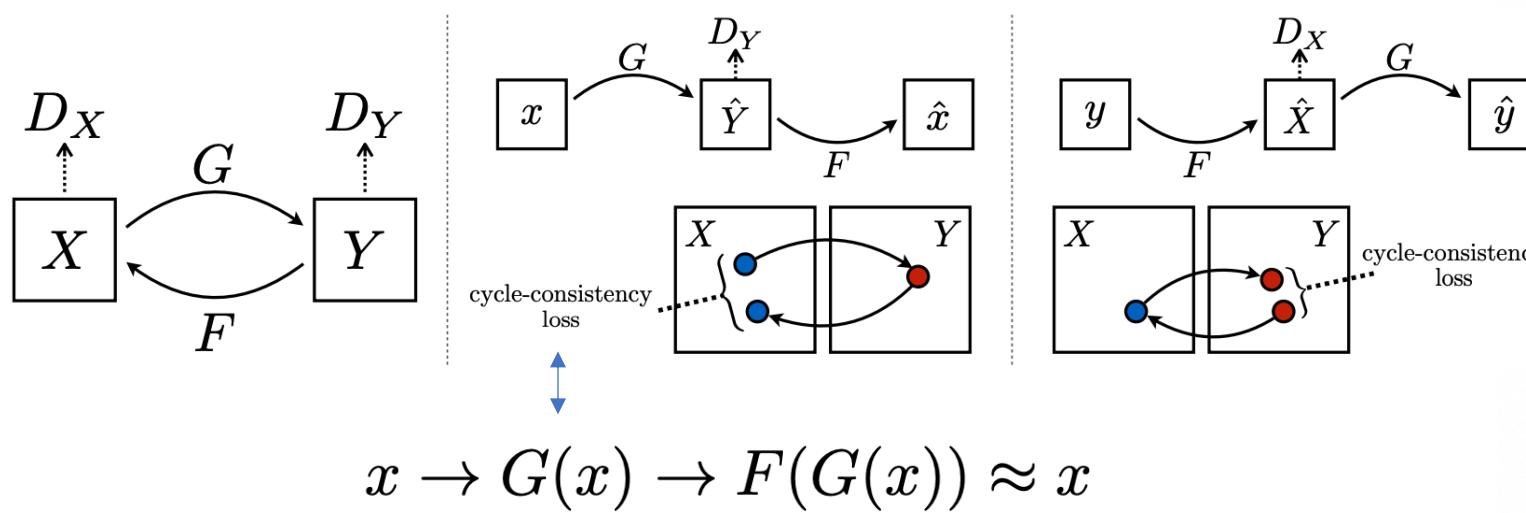
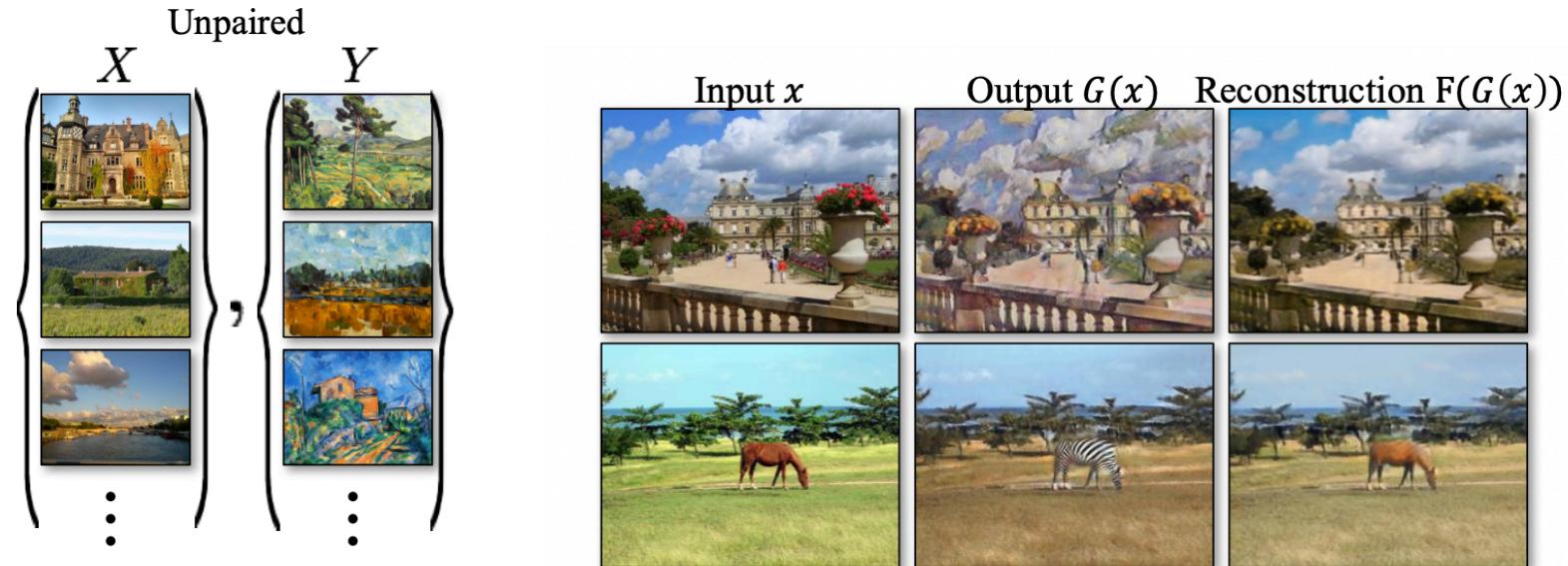


# Cycle GANs

## Cycle GAN

-> unpaired Image-to-Image translation

<https://arxiv.org/pdf/1703.10593.pdf>

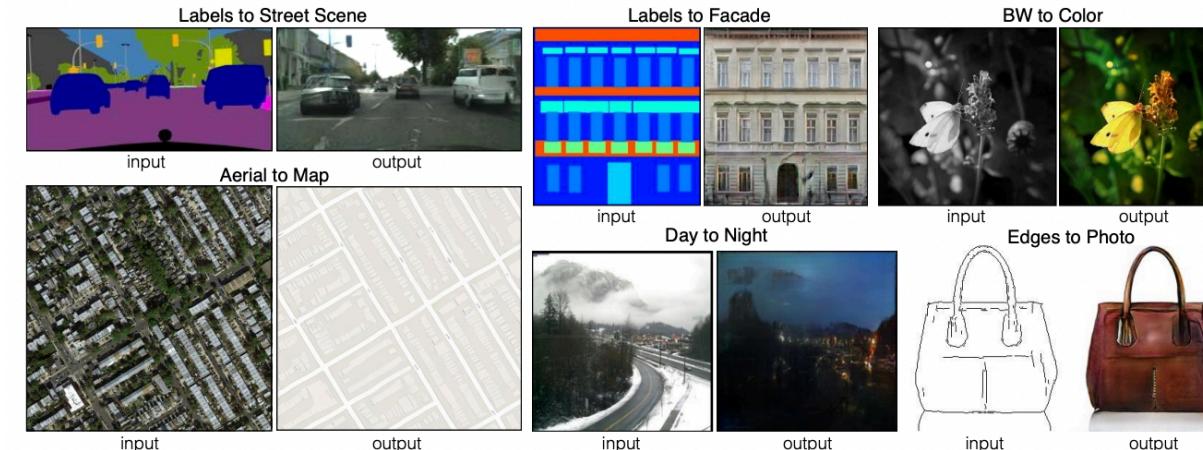
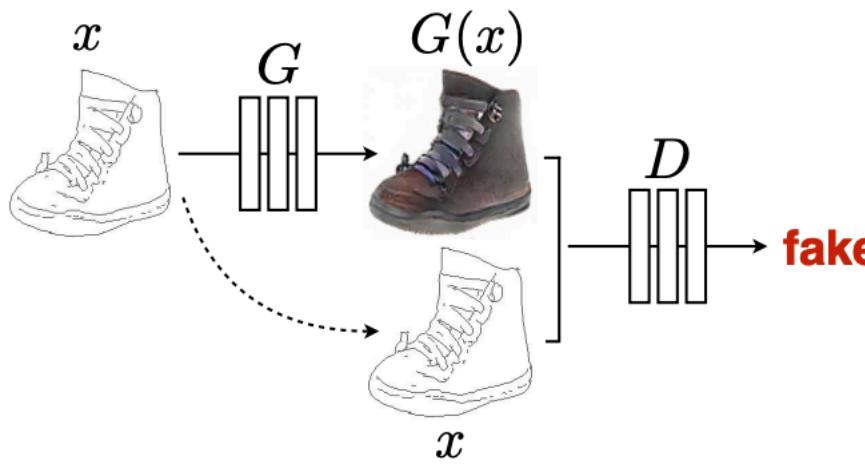


# Conditional GANs

Conditional GAN

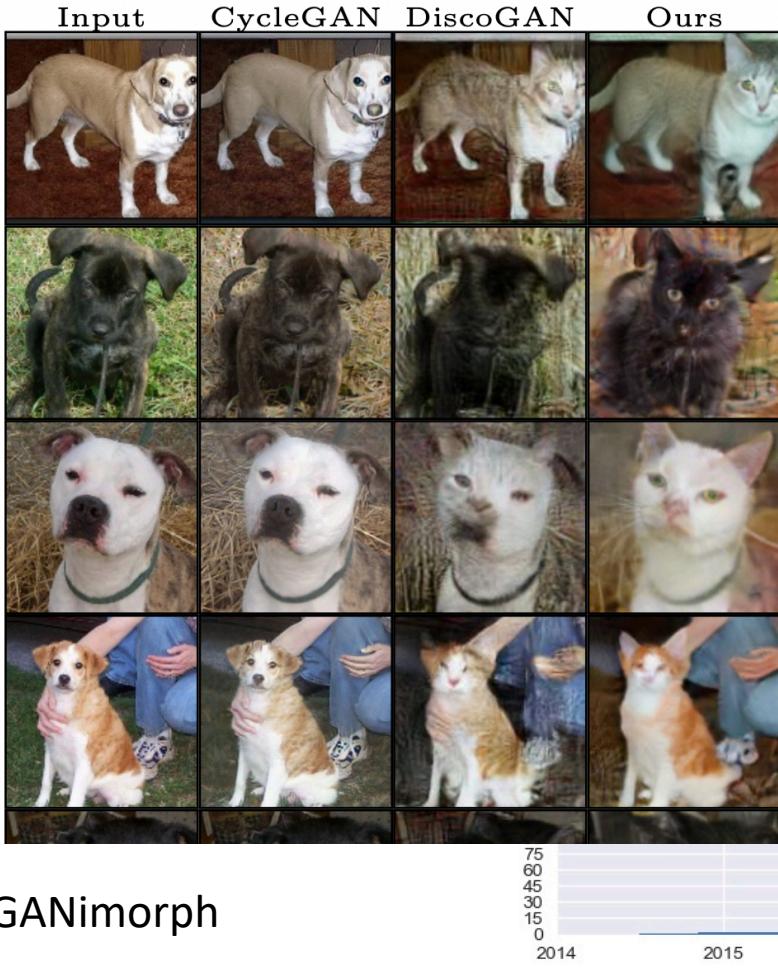
-> e.g. pix2pix model

Unlike an unconditional GAN, in CGANs both the generator and discriminator observe the input edge map:



# Generative adversarial networks zoo

Many more flavors of GANs



> 500 @ 2019

Number of named GAN papers by month

Dog → Cat



[github.com/hindupuravinash/the-gan-zoo](https://github.com/hindupuravinash/the-gan-zoo)

- 3D-ED-GAN - Shape Inpainting using 3D Generative Adversarial Network and Recurrent Convolutional Networks
- 3D-GAN - Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling ([github](#))
- 3D-IWGAN - Improved Adversarial Systems for 3D Object Generation and Reconstruction ([github](#))
- 3D-PhysNet - 3D-PhysNet: Learning the Intuitive Physics of Non-Rigid Object Deformations
- 3D-RecGAN - 3D Object Reconstruction from a Single Depth View with Adversarial Learning ([github](#))
- ABC-GAN - ABC-GAN: Adaptive Blur and Control for improved training stability of Generative Adversarial Networks ([github](#))
- ABC-GAN - GANs for LIFE: Generative Adversarial Networks for Likelihood Free Inference
- AC-GAN - Conditional Image Synthesis With Auxiliary Classifier GANs
- acGAN - Face Aging With Conditional Generative Adversarial Networks
- ACGAN - Coverless Information Hiding Based on Generative adversarial networks
- acGAN - On-line Adaptative Curriculum Learning for GANs
- ACtuAL - ACtuAL: Actor-Critic Under Adversarial Learning
- AdaGAN - AdaGAN: Boosting Generative Models
- Adaptive GAN - Customizing an Adversarial Example Generator with Class-Conditional GANs
- AdvEntuRe - AdvEntuRe: Adversarial Training for Textual Entailment with Knowledge-Guided Examples
- AdvGAN - Generating adversarial examples with adversarial networks
- AE-GAN - AE-GAN: adversarial eliminating with GAN
- AE-OT - Latent Space Optimal Transport for Generative Models
- AEGAN - Learning Inverse Mapping by Autoencoder based Generative Adversarial Nets
- AF-DCGAN - AF-DCGAN: Amplitude Feature Deep Convolutional GAN for Fingerprint Construction in Indoor Localization System
- AffGAN - Amortised MAP Inference for Image Super-resolution
- AIM - Generating Informative and Diverse Conversational Responses via Adversarial Information Maximization
- AL-CGAN - Learning to Generate Images of Outdoor Scenes from Attributes and Semantic Layouts
- ALI - Adversarially Learned Inference ([github](#))
- AlignGAN - AlignGAN: Learning to Align Cross-Domain Images with Conditional Generative Adversarial Networks
- AlphaGAN - AlphaGAN: Generative adversarial networks for natural image matting
- AM-GAN - Activation Maximization Generative Adversarial Nets
- AmbientGAN - AmbientGAN: Generative models from lossy measurements ([github](#))