Deep Learning

Alberto Ezpondaburu

Analytical Index

Analytical Index

5

Deep Generative Models

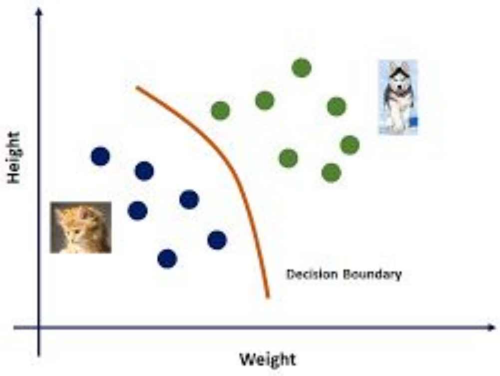
Introduction

5.1

Generative vs Discriminative

Generative vs Discriminative models

Discriminative

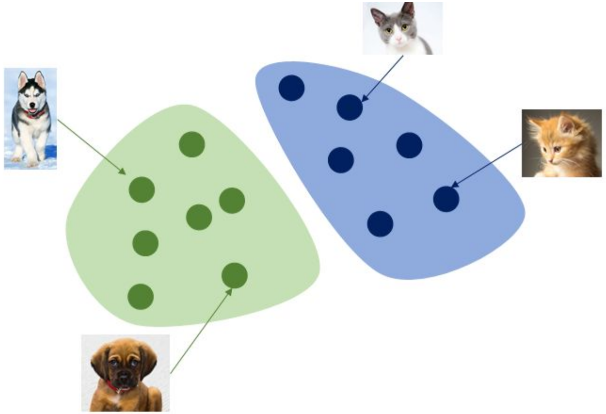


Model of the conditional probability of the target Y, given an observation x

Features $X \rightarrow Y$ Target

$$P(Y|X)$$

Generative



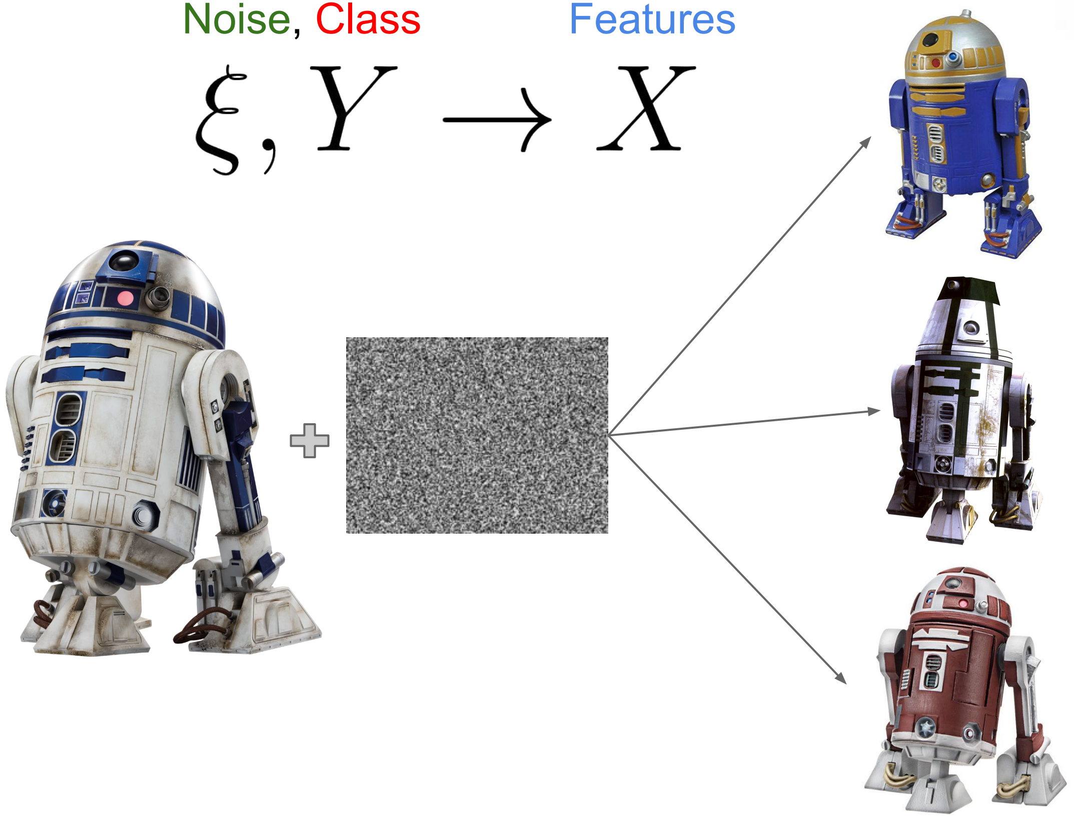
Target $Y \rightarrow X$ Features

$$P(X|Y)$$

Generative

Noise, Class Features

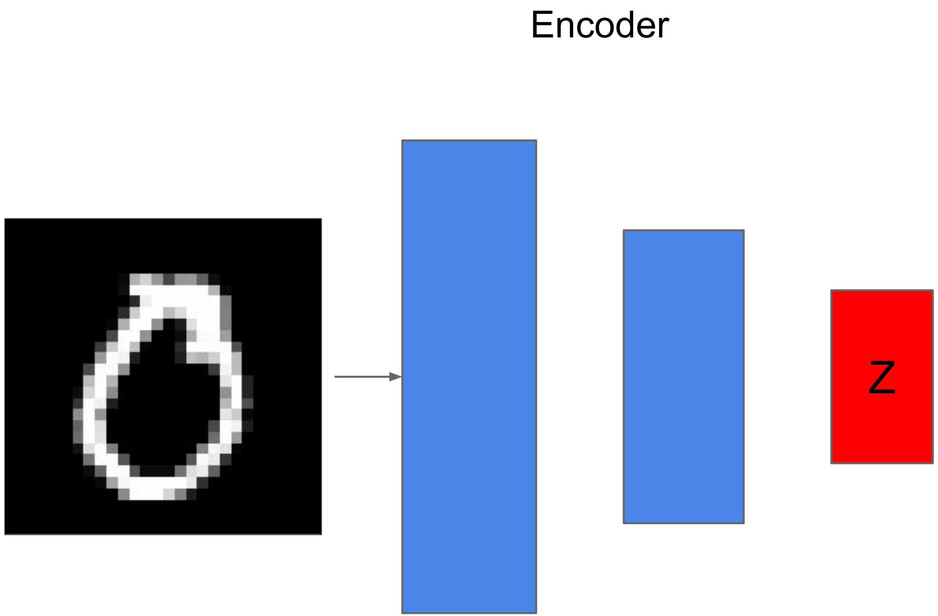
$$\xi, Y \rightarrow X$$



5.2

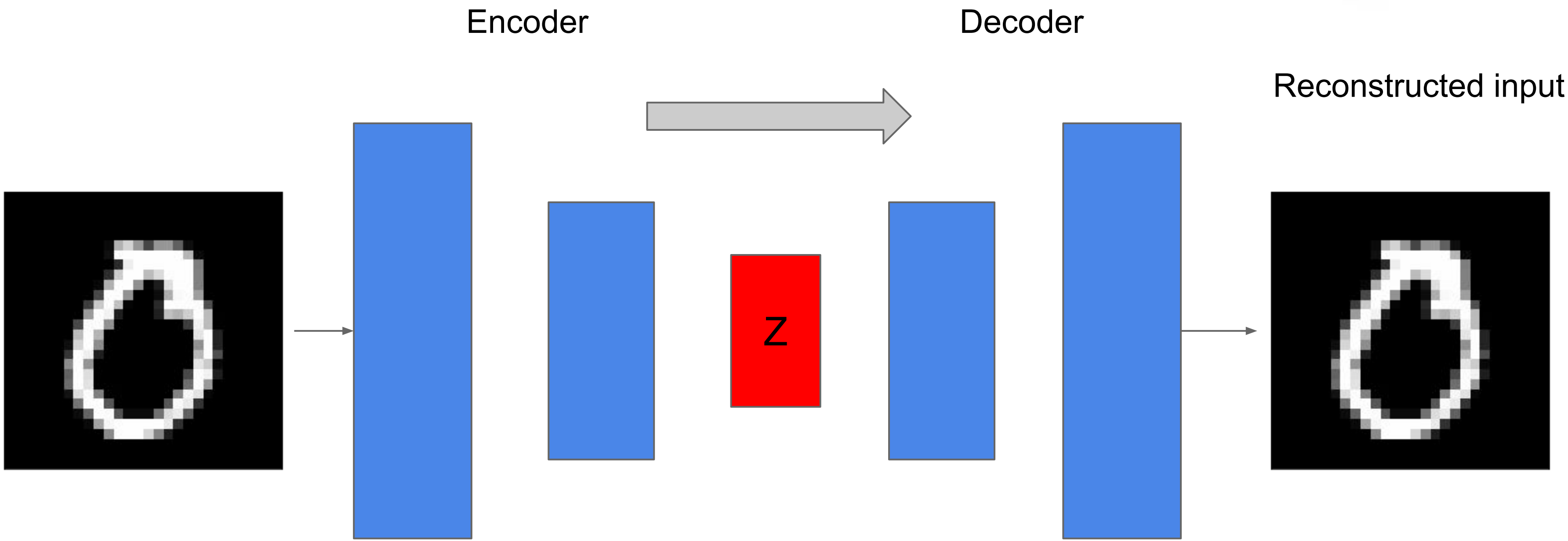
Autoencoders

Autoencoder Intuition



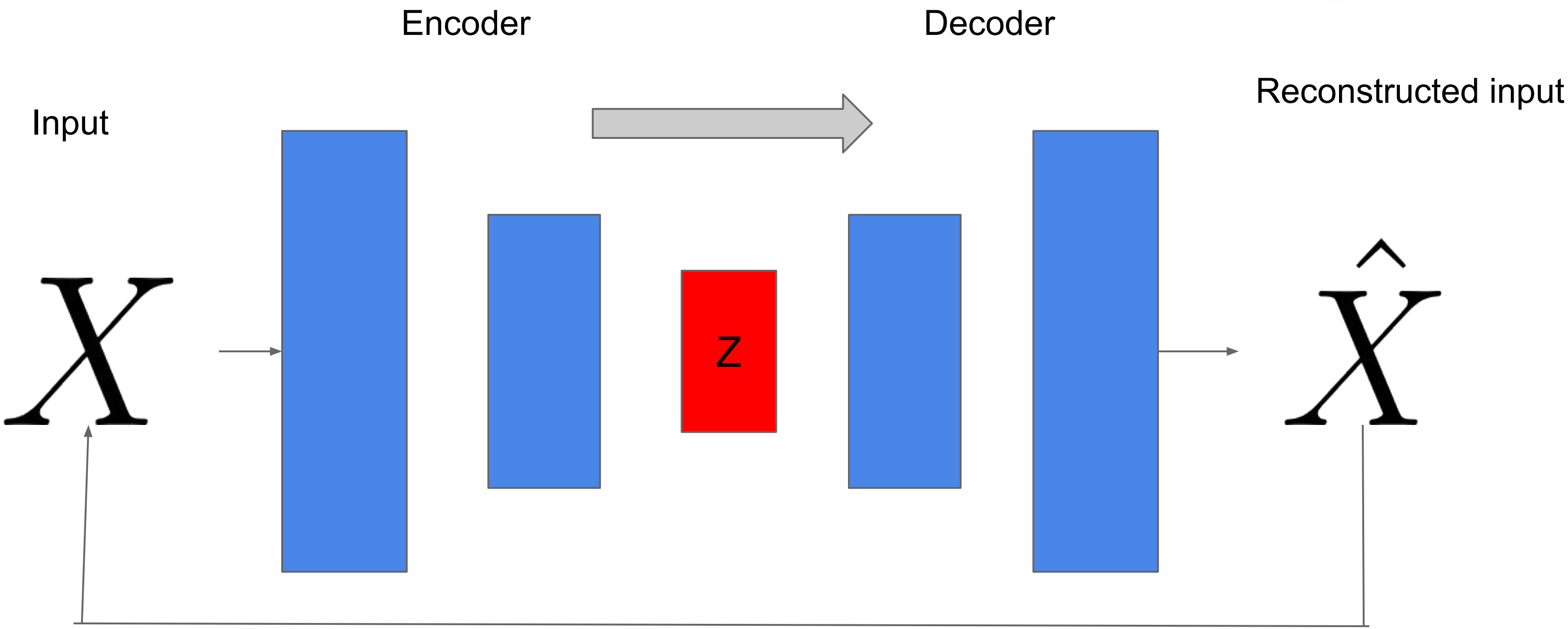
- We want to obtain a lower dimensional representation from the input data.
- Useful for dimensionality reduction and visualization.
- The encoder maps input data into a lower dimensional latent space (z)

Autoencoder Intuition



Train the model to reconstruct the original input data.
The decoder learns to reconstruct the data with the latent space.

Autoencoder: Reconstruction Loss



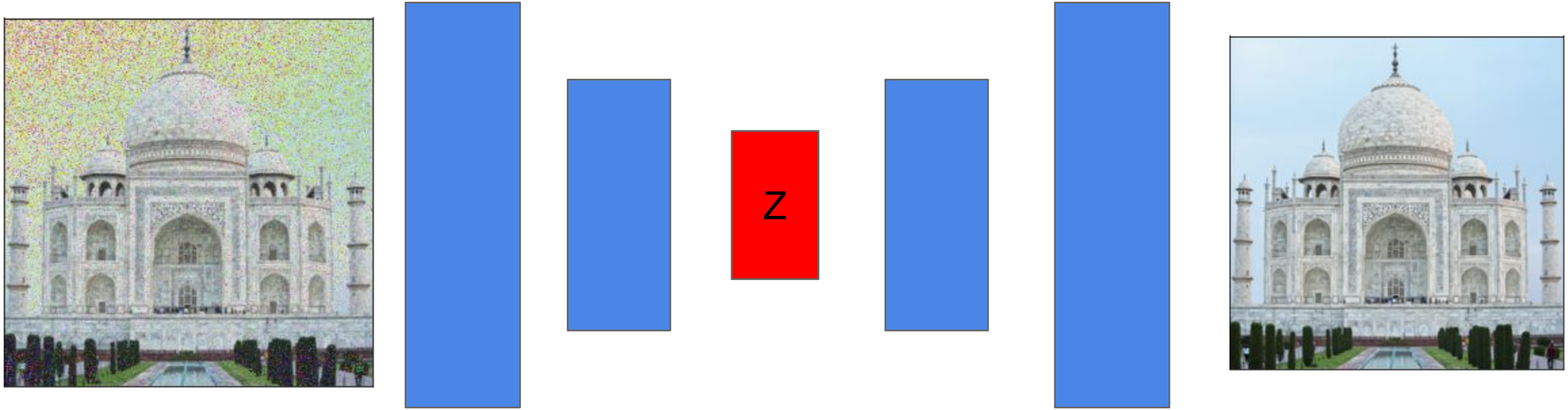
$$\mathcal{L}(X, \hat{X}, W) = ||X - \hat{X}||^2$$

Autoencoder Applications: Compression

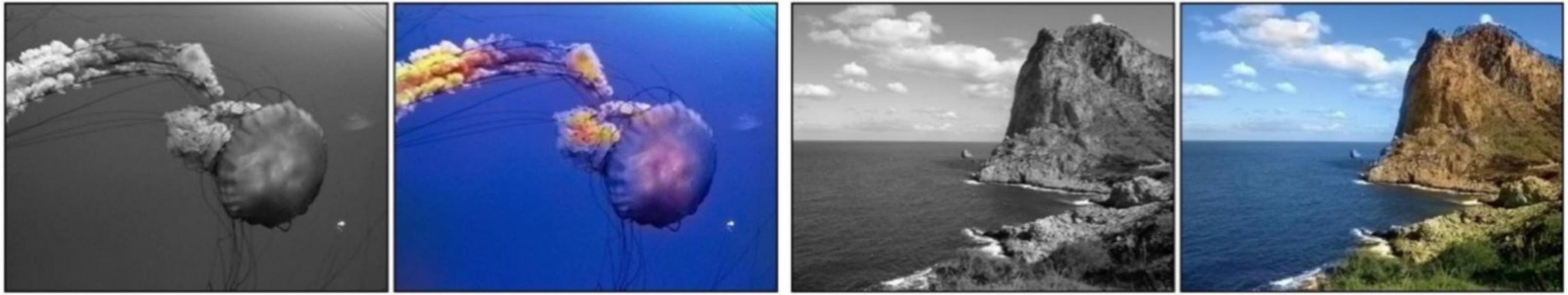
Autoencoder Applications: Denoising

Noisy Image

Denoised Image



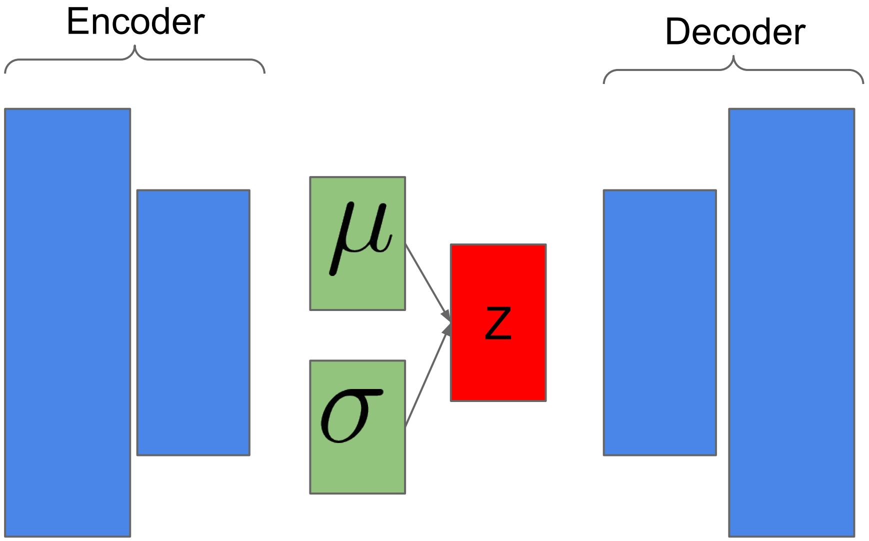
Autoencoder Applications: Colorizing Images



5.3

Variational Autoencoders

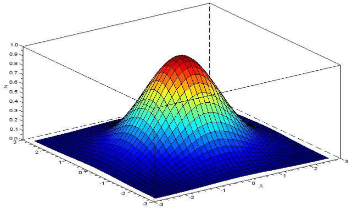
VAEs



Encoder computes: $q_\phi(z|x)$

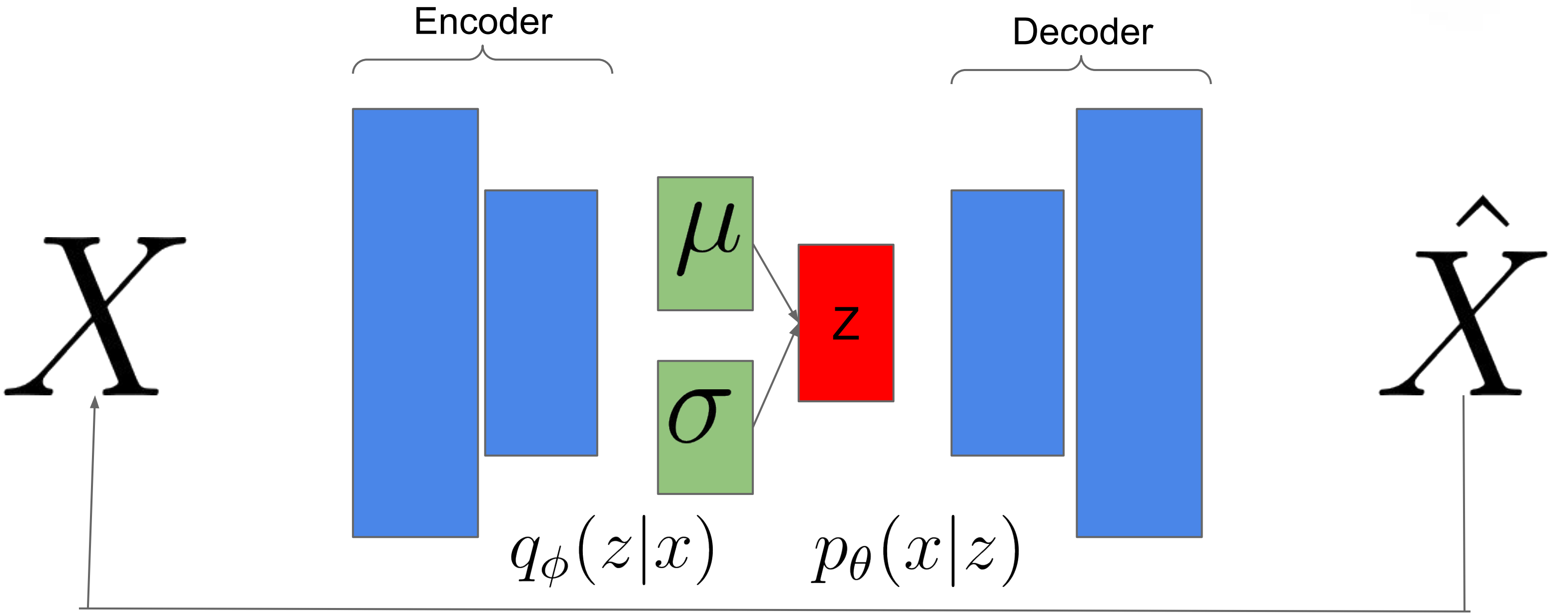
Decoder computes: $p_\theta(x|z)$

$x \xrightarrow{q_\phi(z|x)} z$



$z \xrightarrow{p_\theta(x|z)} \tilde{x}$

VAEs: Loss



$$\mathcal{L} = \underbrace{||X - \hat{X}||^2}_{\text{Reconstruction loss}} + D(\underbrace{q_\phi(z|x)||p(z)}_{\text{Regularization}}))$$

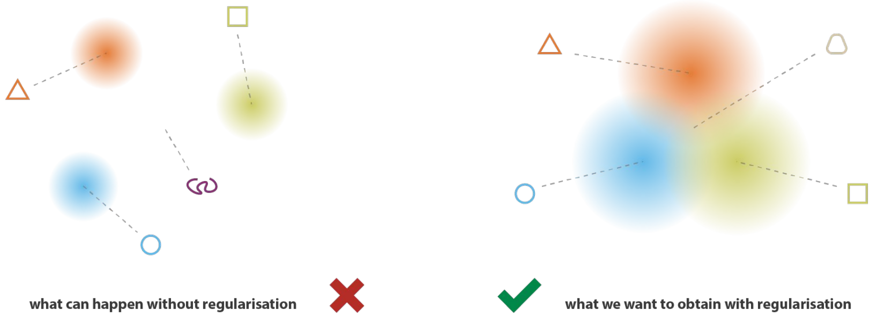
VAEs: Regularization

$$D(q_\phi(z|x)||p(z))$$

Inferred Distribution

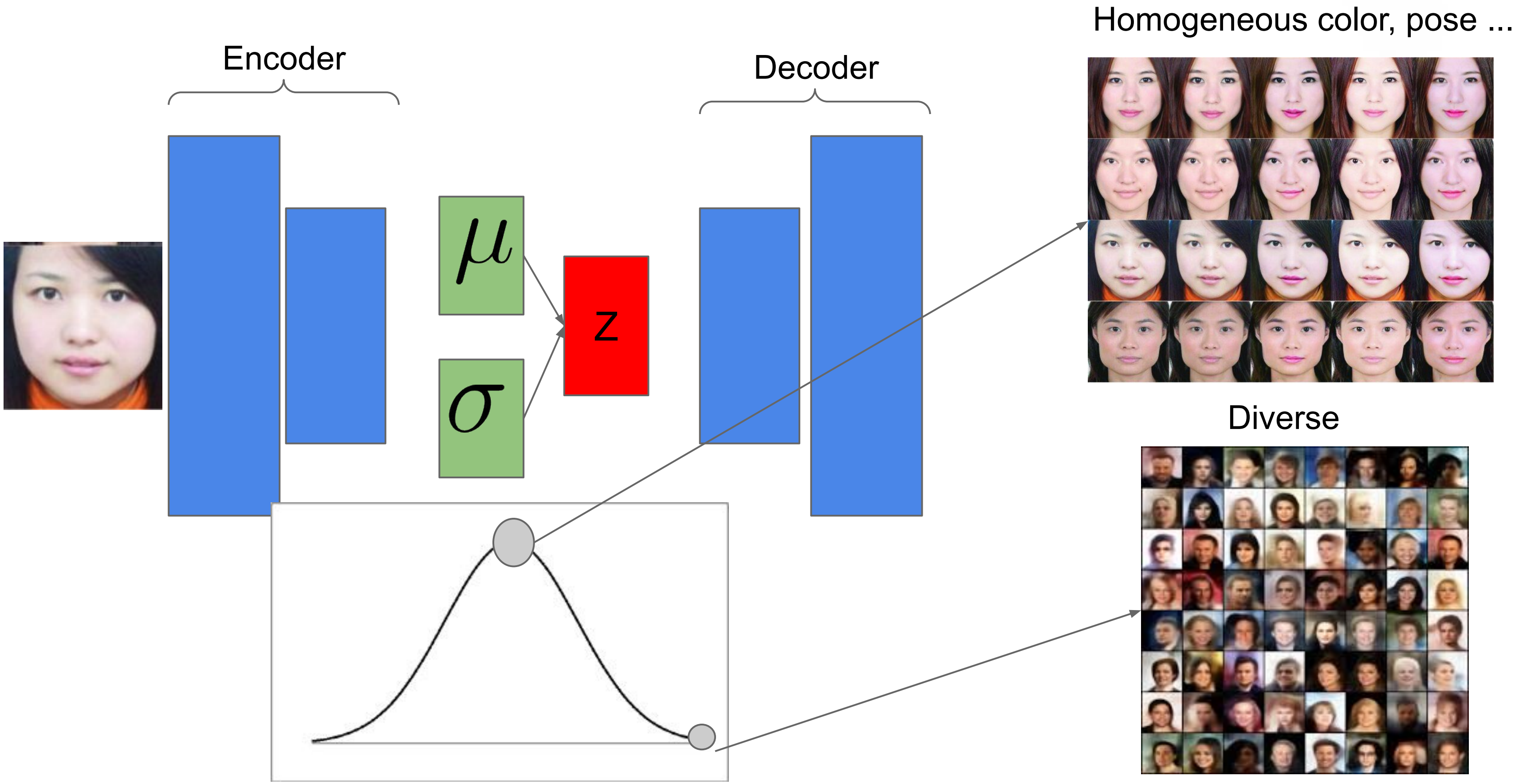
$$\mathcal{N}(0,\ 1)$$

Continuity (two close points in the latent space should not give two completely different contents once decoded)
Completeness (for a chosen distribution, a point sampled from the latent space should give “meaningful” content once decoded).



https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73

VAEs

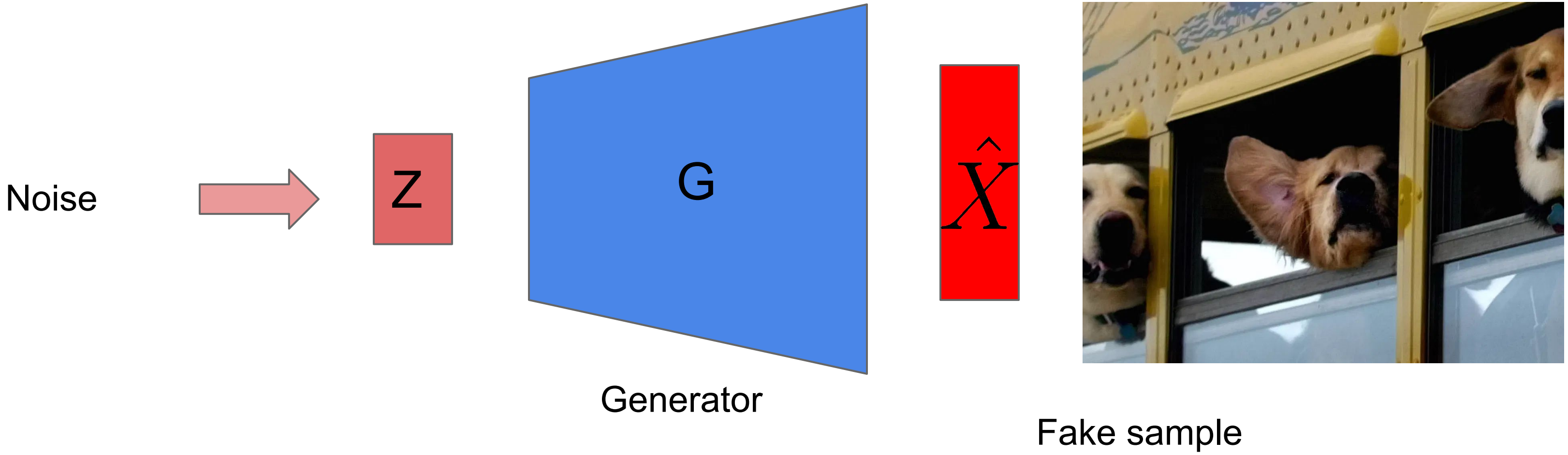


5.4

GANs: Generative Adversarial Networks

GANs: Why do we need to learn the distribution?

Instead of learning the density of the data, why don't we train a model that samples it directly?



GANs: Generator & Discriminator Intuition

Generative ADVERSARIAL Network: We have two competing neural networks.

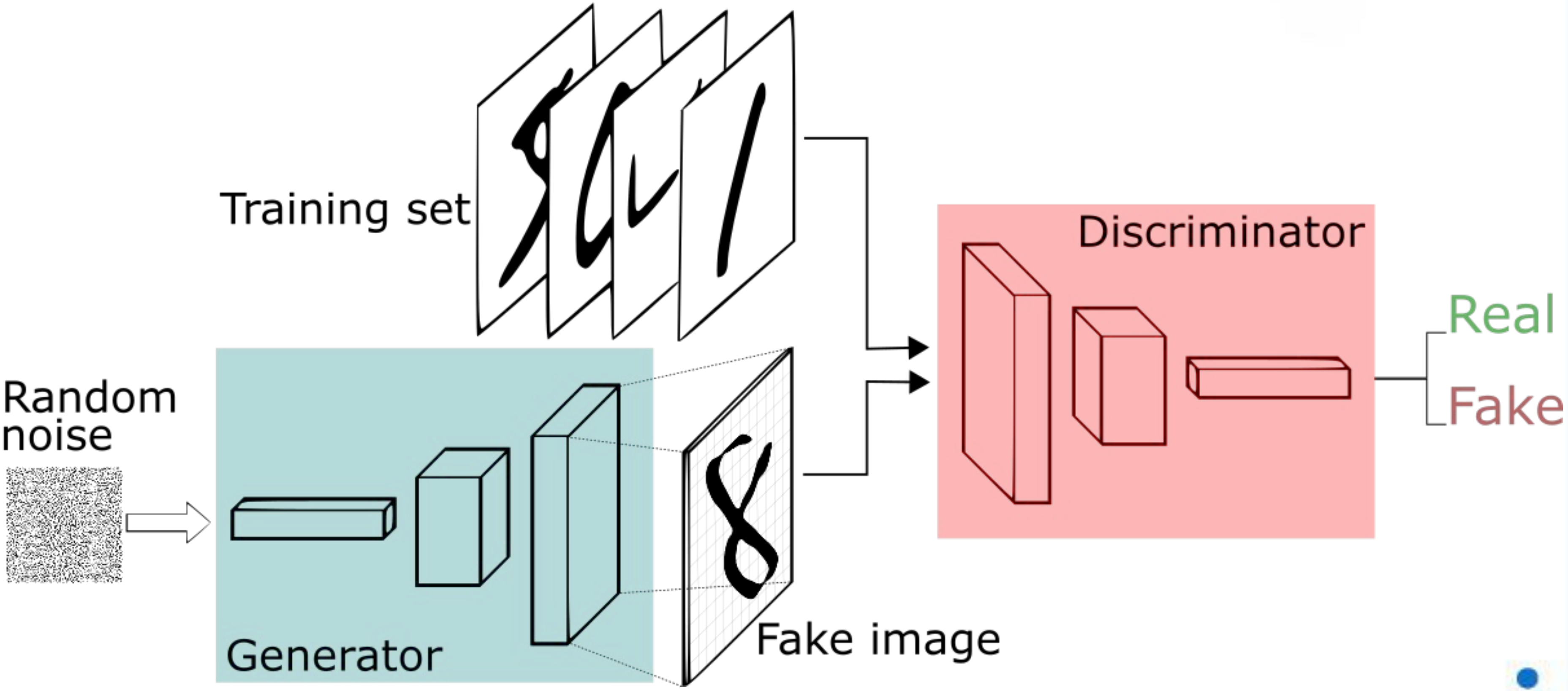
<u>Generator</u>

Learn to sample **fake** data, that **looks real**.

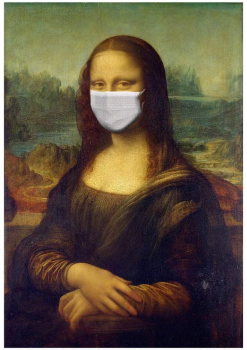
<u>Discriminator</u>

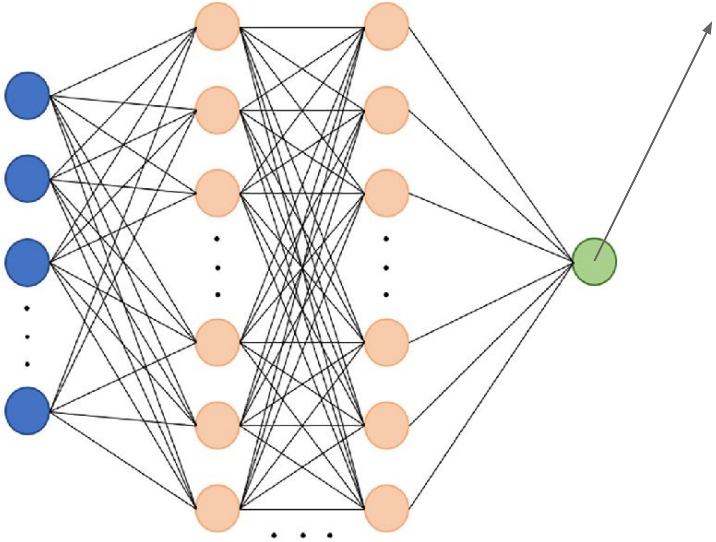
Learns to **distinguish** real from fake data.

GANs: Generator & Discriminator



GANs: Discriminator

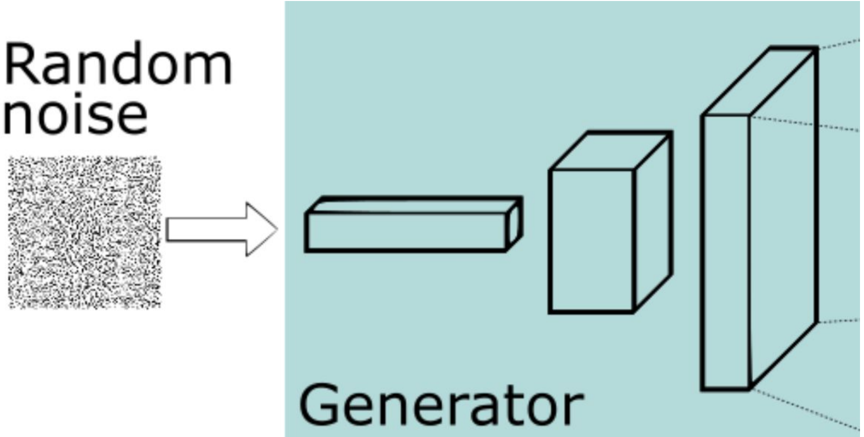


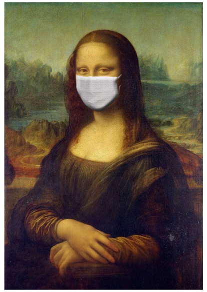


Fake or Real ??

- Is a binary classifier.
- Learns the probability of class Y (**fake** or **real**) given input X.
- Probabilities are the feedback for the generator.

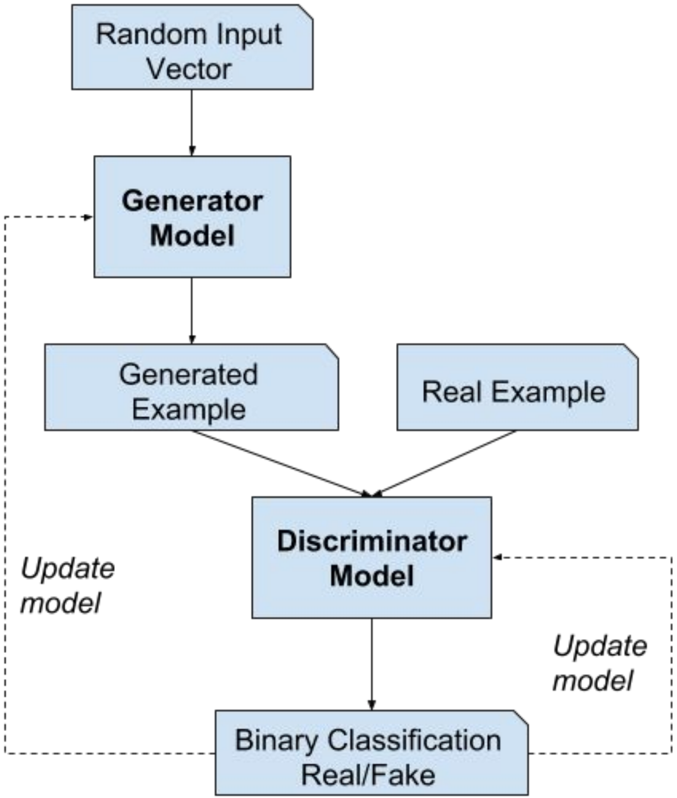
GANs: Generator

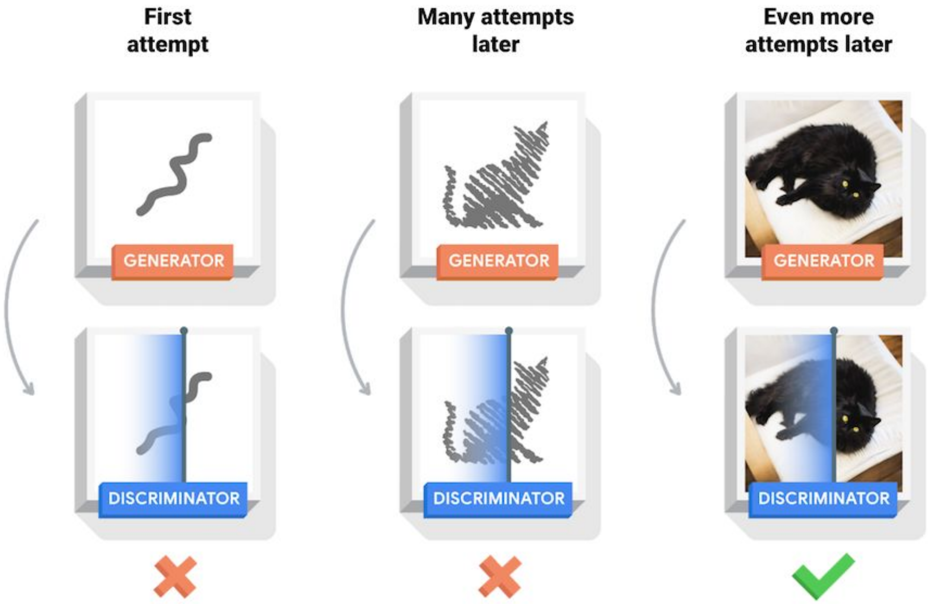




- Produces fake data.
- Learns the approximate distribution of inputs data.
- Takes noise as input

GANs: Training





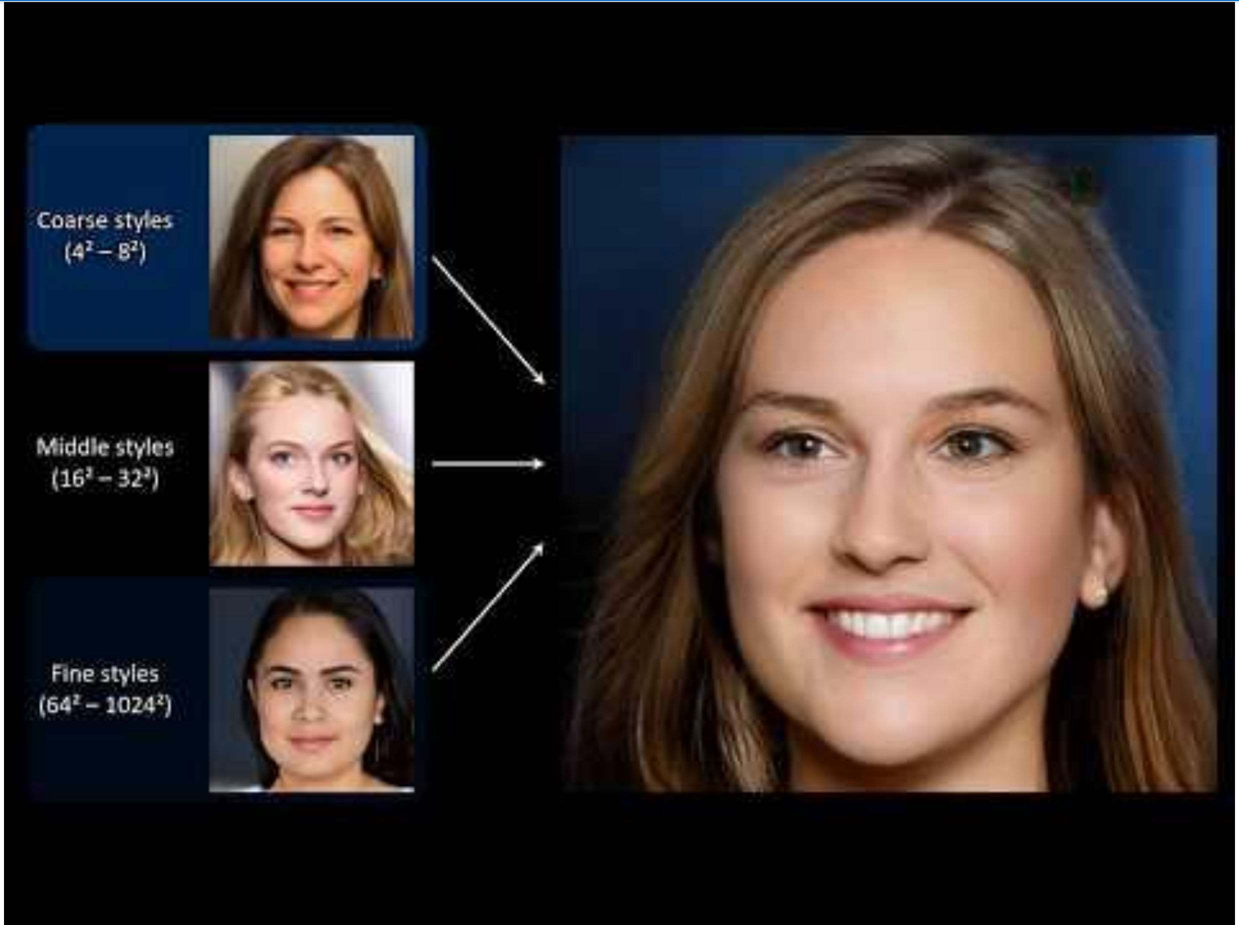
Generator and discriminator, are trained together. The generator generates a batch of samples, and combined with real data are provided to the discriminator and classified as real or fake.

GANs: StyleGAN

StyleGAN is a novel generative adversarial network (GAN) introduced by Nvidia researchers in December 2018

GANs: StyleGAN

GANs: StyleGAN

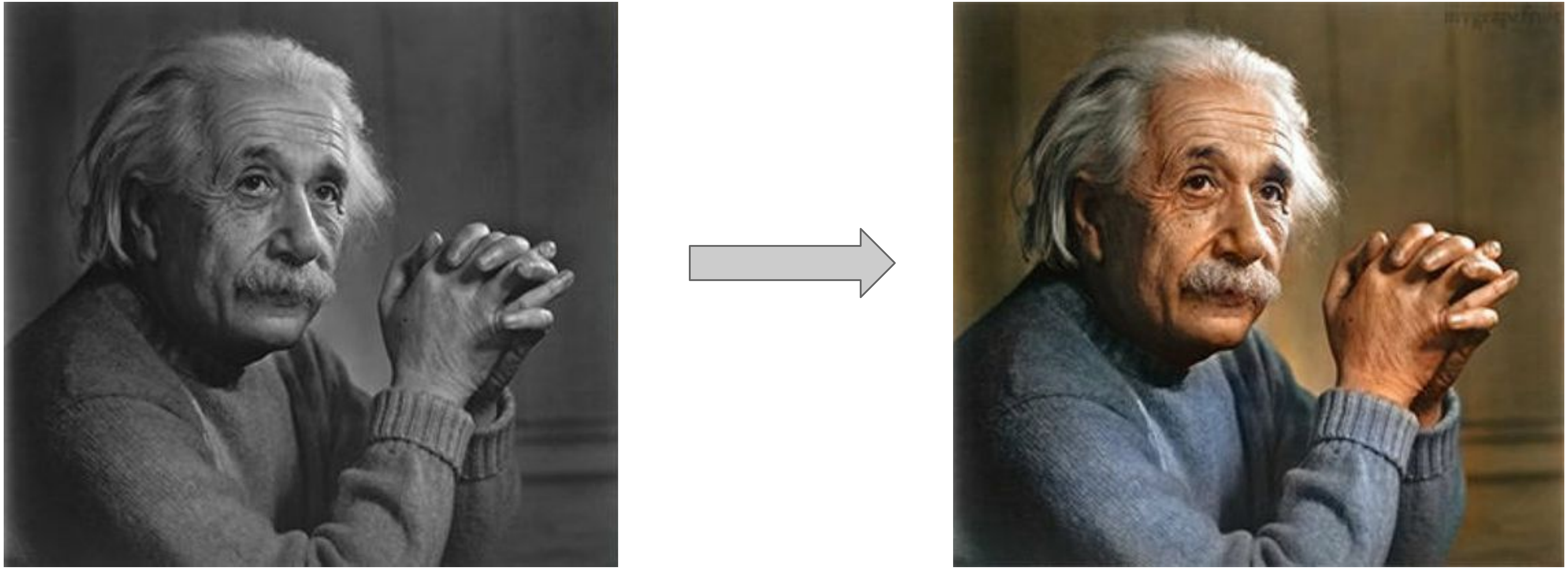


GANs: Create Anime Characters

Towards the Automatic Anime Characters Creation with Generative Adversarial Networks: https://arxiv.org/abs/1708.05509

GANs: Image to Image Translation

Image-to-image translation task of taking images from one domain and transforming them so they have the style (or characteristics) of images from another domain.



GANs: Pix2Pix Image to Image Translation

Paired translation: Train with pairs of images of the two domains

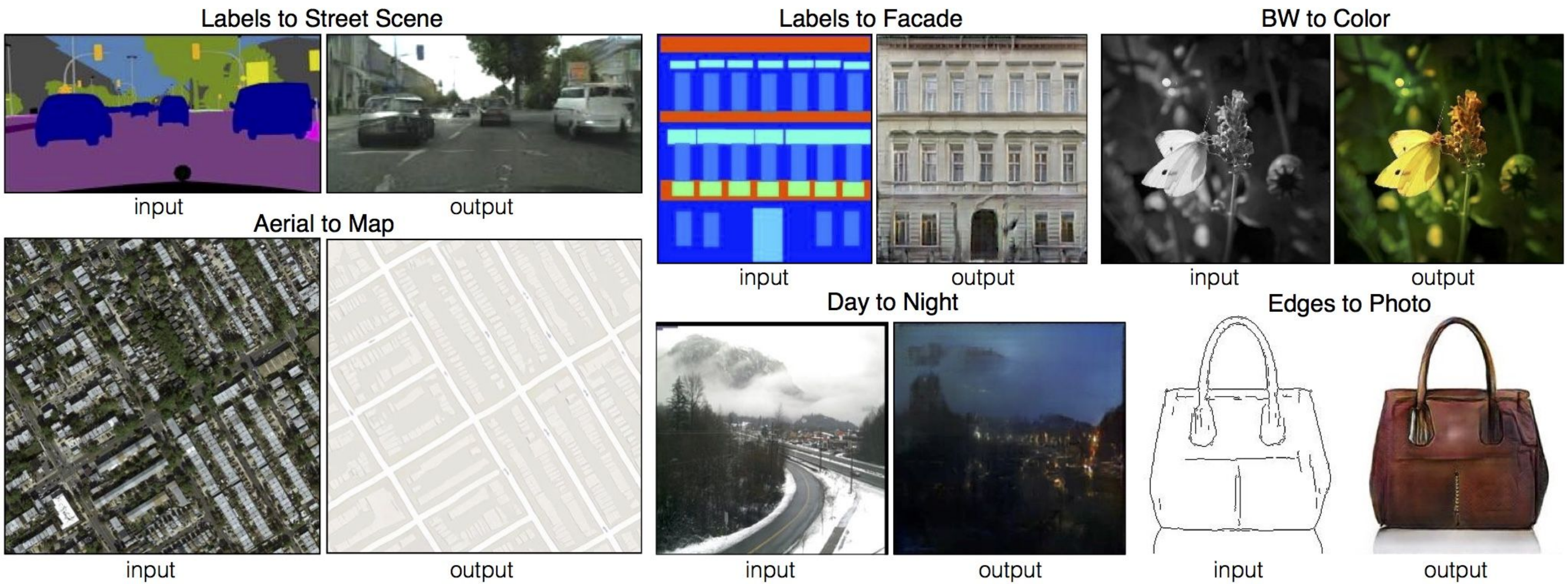
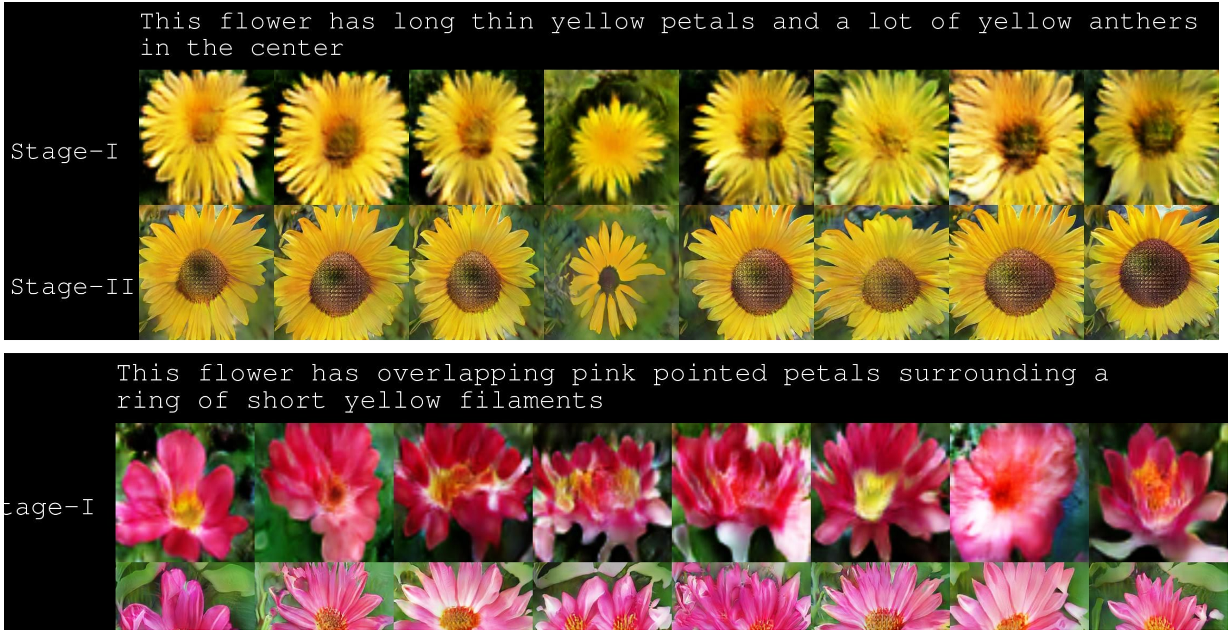


Image-to-Image Translation with Conditional Adversarial Networks

GANs: Face Aging

GANs: Text to Image StackGAN

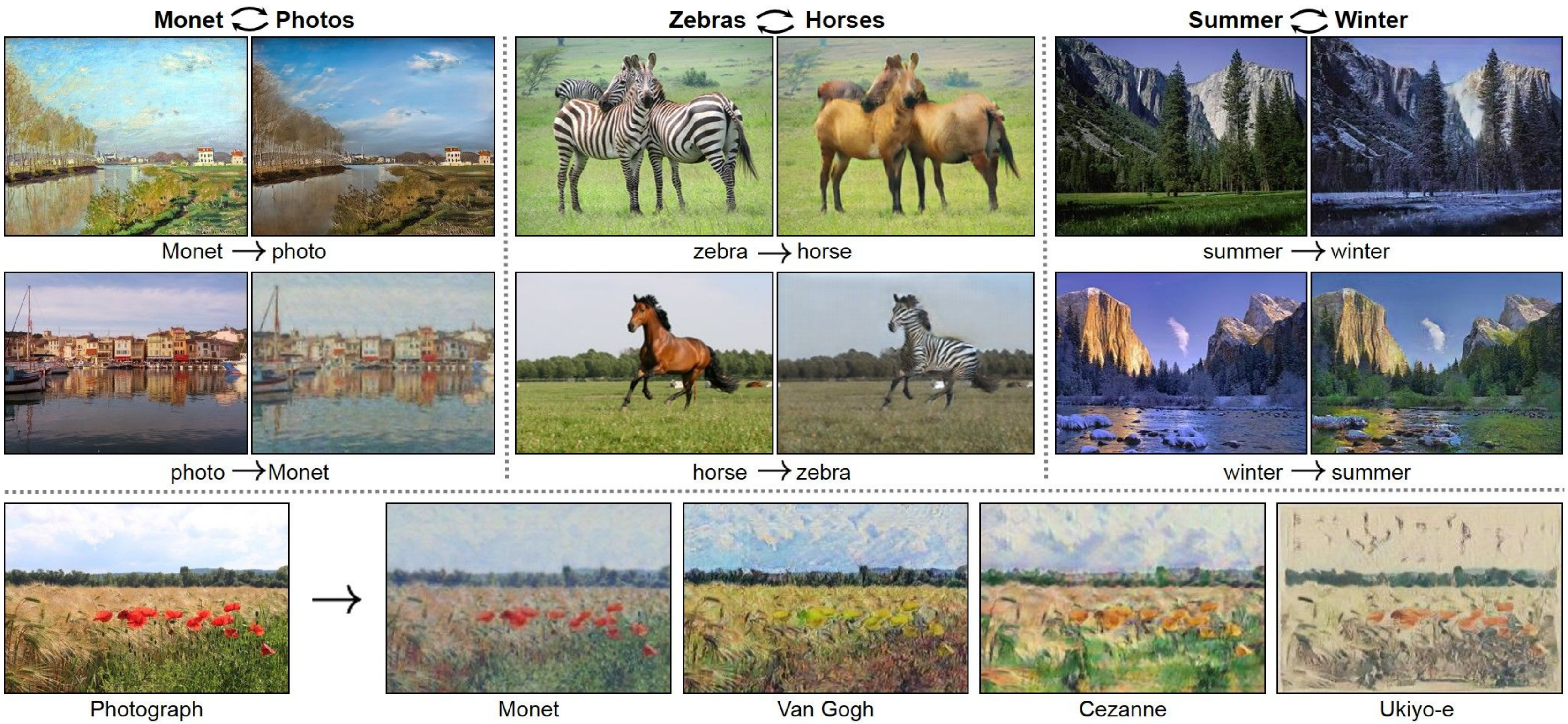


https://github.com/hanzhanggit/StackGAN

GANs: CycleGAN, Image to Image Translation

Unpaired image-to-image: Capture the characteristics of one image domain and figure out how these characteristics could be translated into another image domain.

GANs: CycleGAN, Image to Image Translation



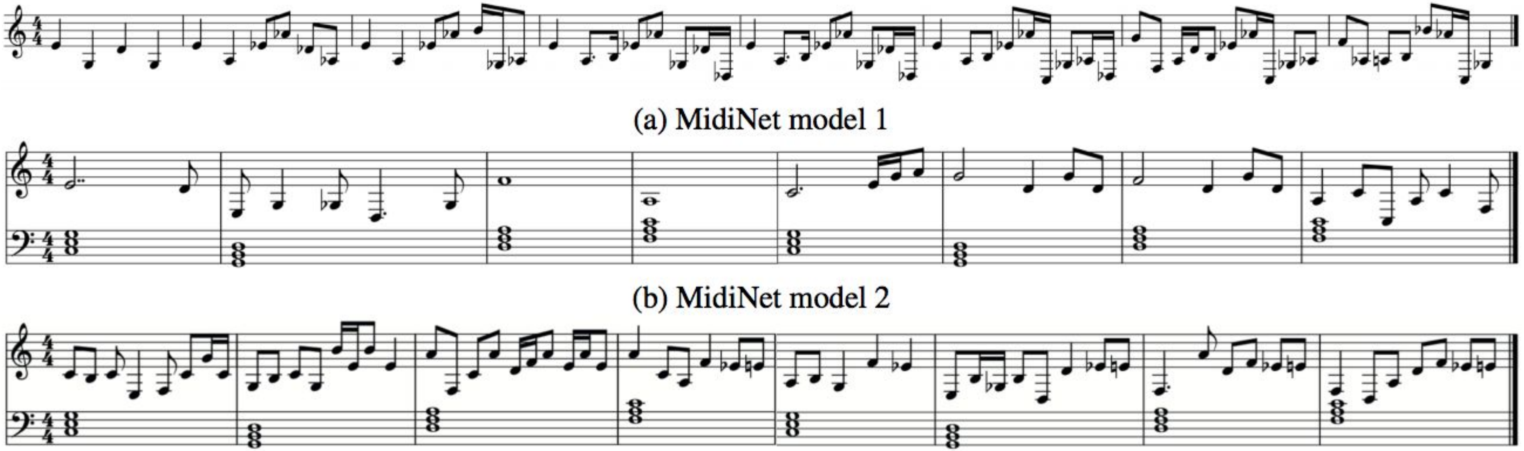
GANs: Multimodal MUNIT

(a) edges ↔ shoes (b) edges ↔ handbags

https://github.com/NVlabs/MUNIT

GANs: GauGAN

GANs: MidiNet



(a) MidiNet model 1

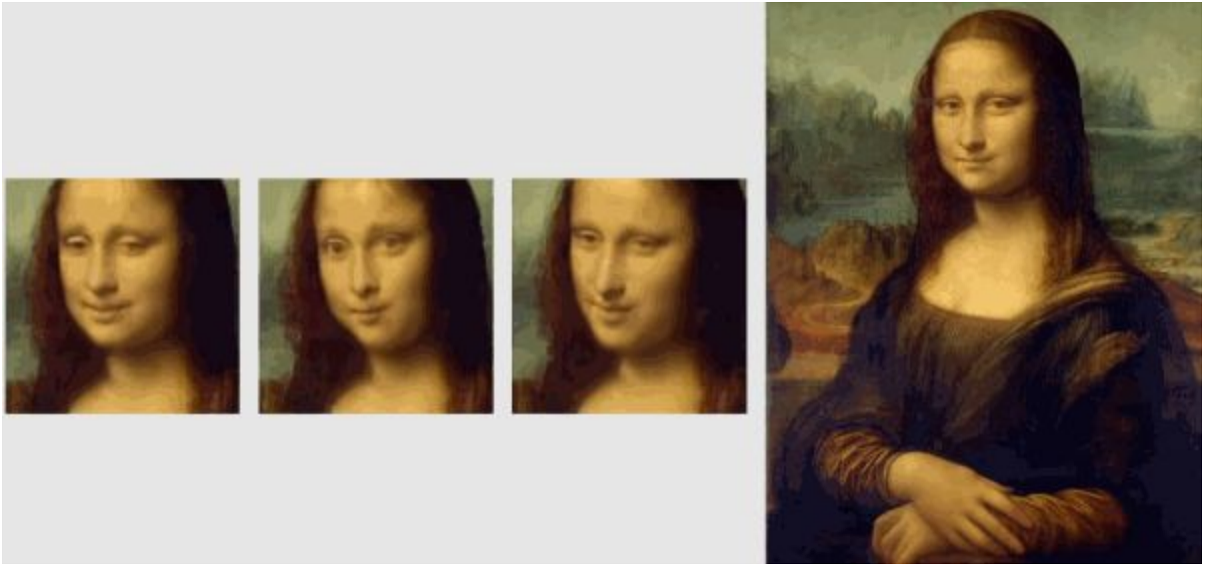
(b) MidiNet model 2

(c) MidiNet model 3

Figure 3. Example result of the melodies (of 8 bars) generated by different implementations of MidiNet.

MIDINET: A CONVOLUTIONAL GENERATIVE ADVERSARIAL NETWORK FOR SYMBOLIC-DOMAIN MUSIC GENERATION.

GANs



Few-Shot Adversarial Learning of Realistic Neural Talking Head Models

GANs: DeepFake