Deep Generative Models

Some images are from Fei-fei Li's course @Standford http://cs231n.stanford.edu/ and Towards Data Science https://towardsdatascience.com

Outline

- Generative models
- Variational Autoencoder (VAE)
- Generative Adversarial Network (GAN)

Generative Models

Given training data, generate new samples from the same distribution

Training data ~ $p_{data}(x)$

Generated samples ~ $p_{model}(x)$

Want to learn $p_{model}(x)$ similar to $p_{data}(x)$

-Explicit density estimation: explicitly define and solve for $\mathbf{p}_{model}(x)$

-Implicit density estimation: learn model that can sample from $\mathbf{p}_{model}(x)$ w/o explicitly defining it

Discriminative Model vs Generative Model

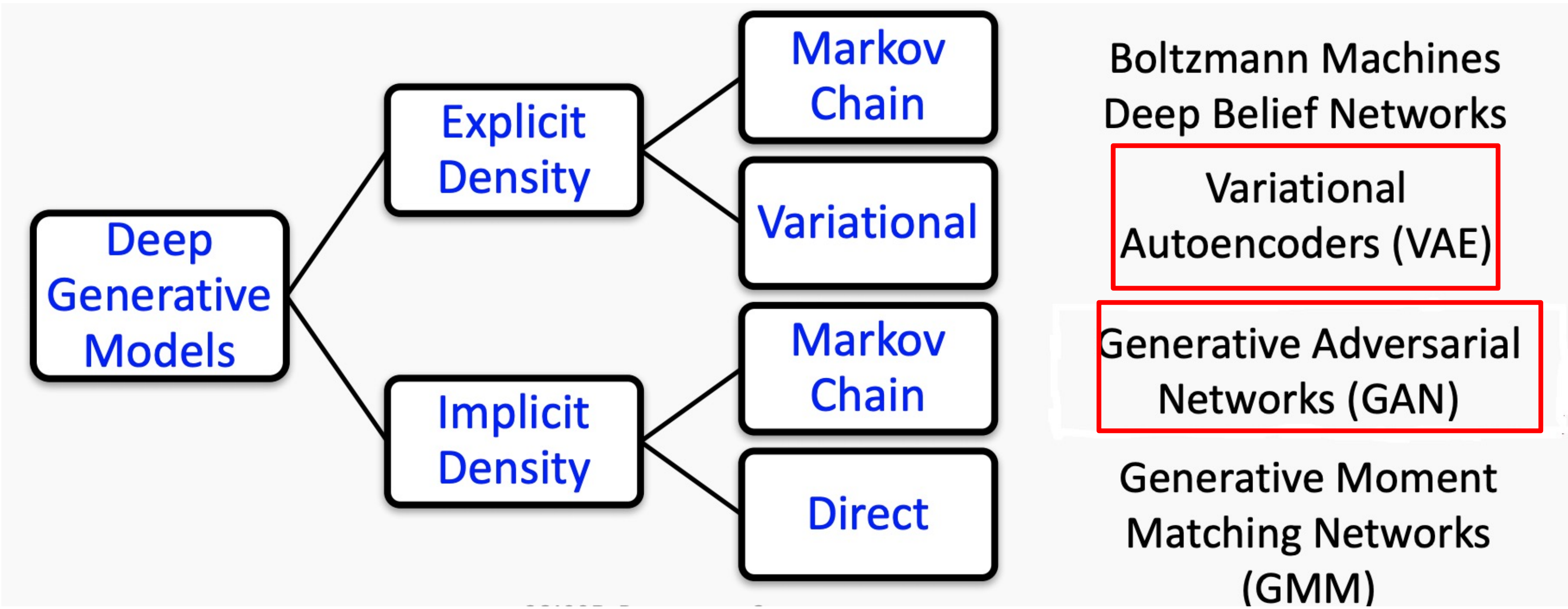
Discriminative Model

- Given an image **X**, predict a label Y
- Estimates P(Y|**X**)

Generative Model

- Model P(**X**)
- Can generate new images

Deep Generative Models



Generative Models

Generate examples in image datasets

Generate realistic photos

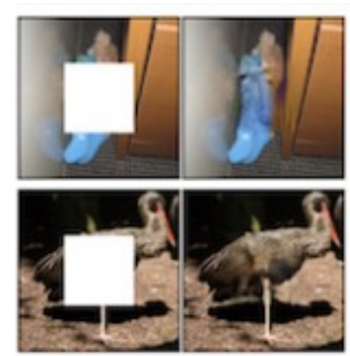


Image inpainting

Image-to-image translation

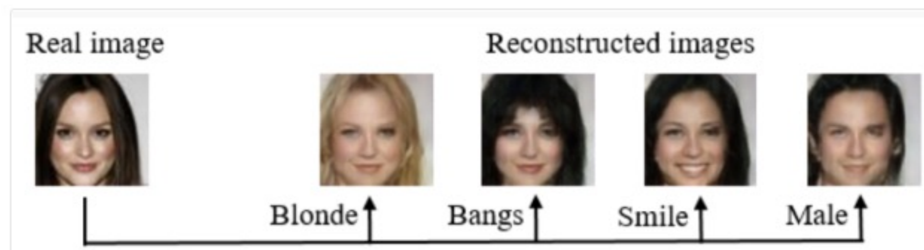
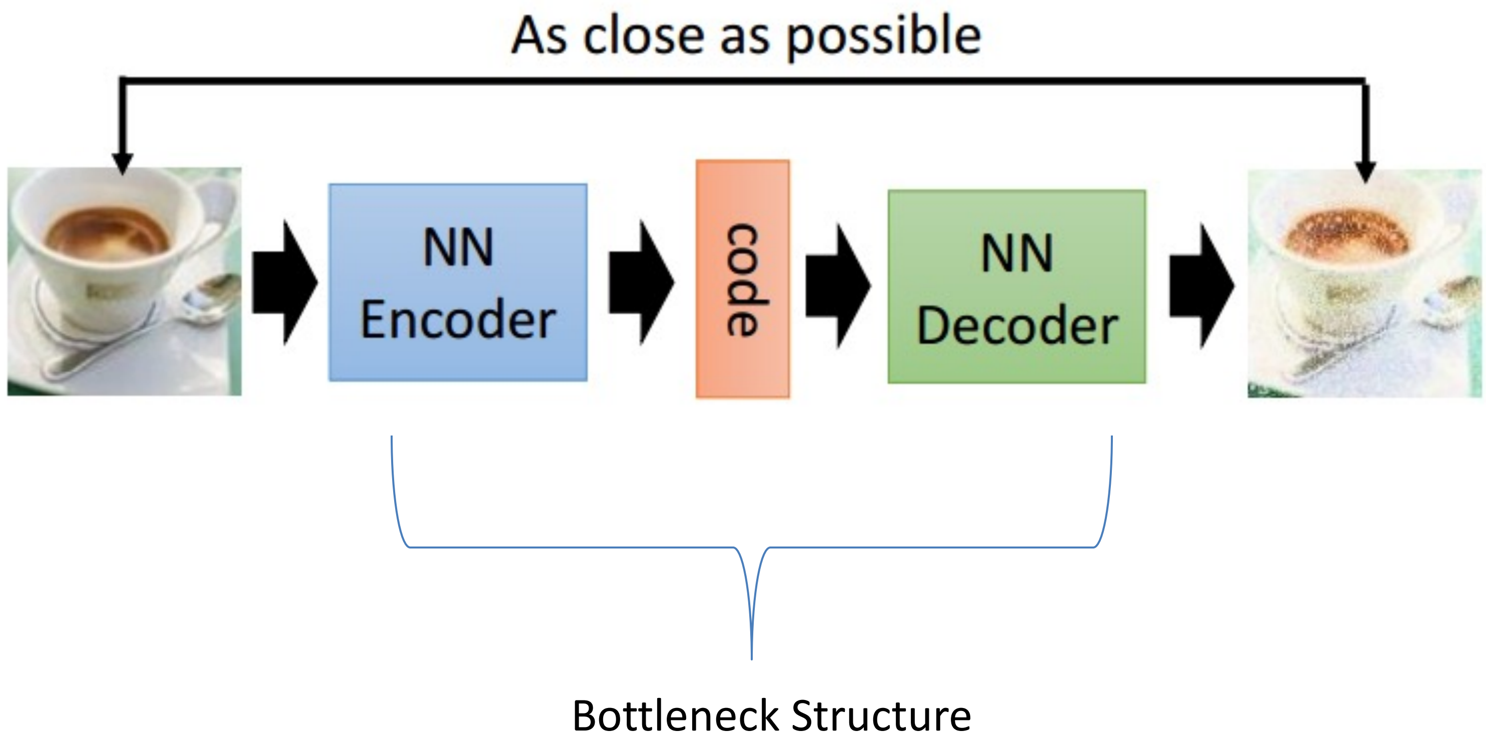


Photo editing

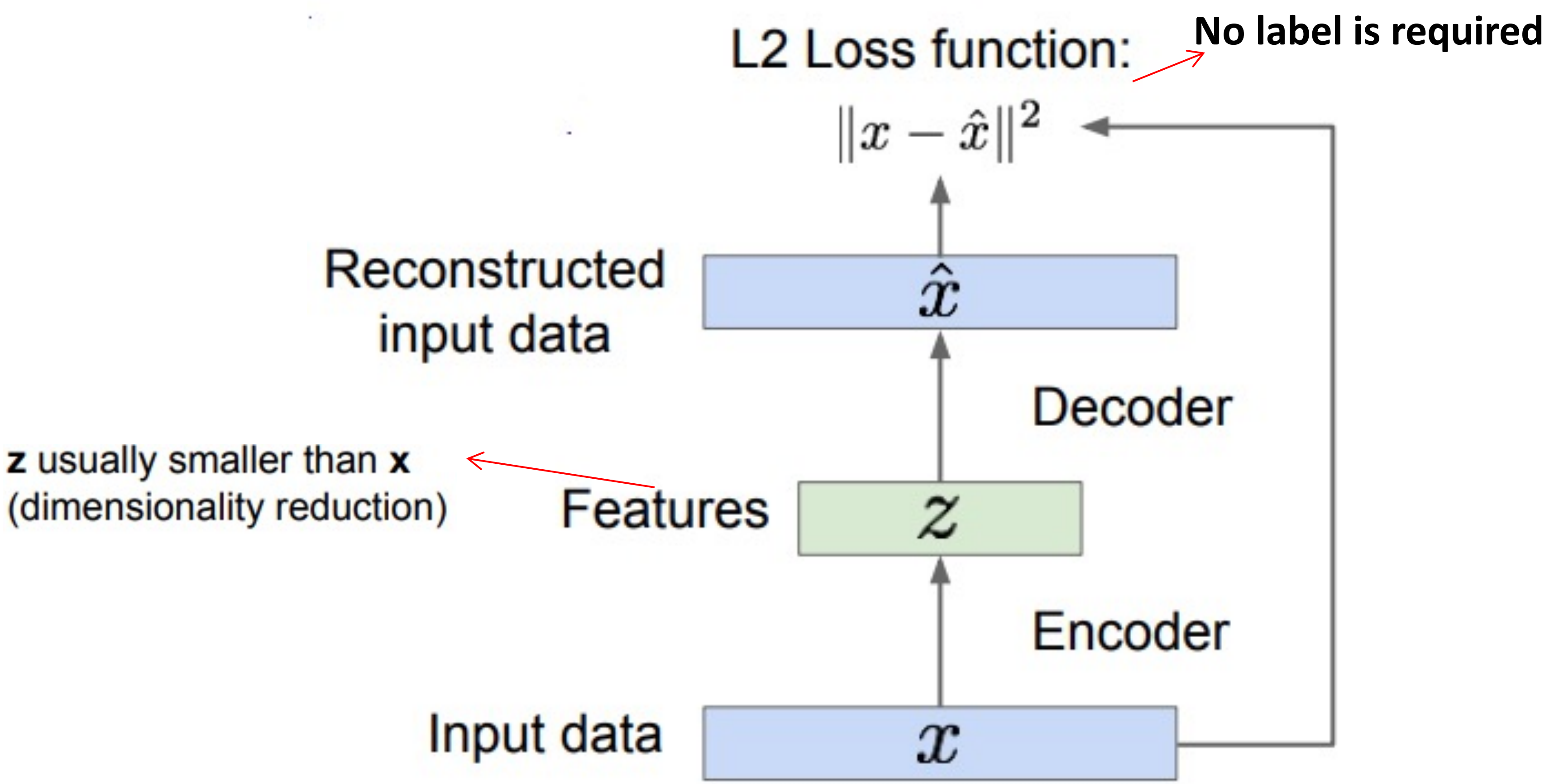
Outline

- Generative models
- Variational Autoencoder (VAE)
- Generative Adversarial Network (GAN)

Auto-encoder (AE): original model



Auto-encoder (AE): original model



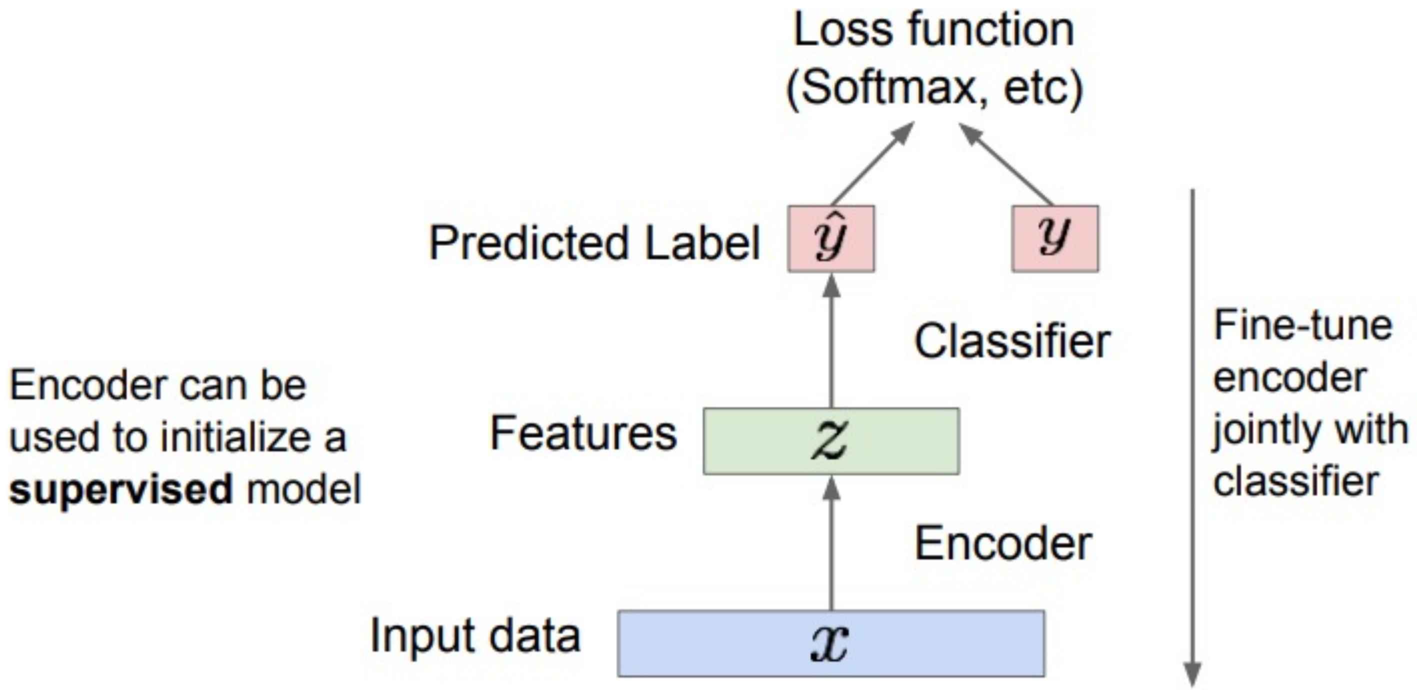
z is a latent variable that has smaller dimension than **x**

Autoencoder (AE): original model

- Autoencoder is a multi-layer neural network. The only difference is that the size of its output layer is the same as the input layer.
- The reconstruction objective does not involve any label.
- The generated latent representation z could be used as features extracted in an unsupervised way.
- The feature z could also be used to initialize a supervised model.

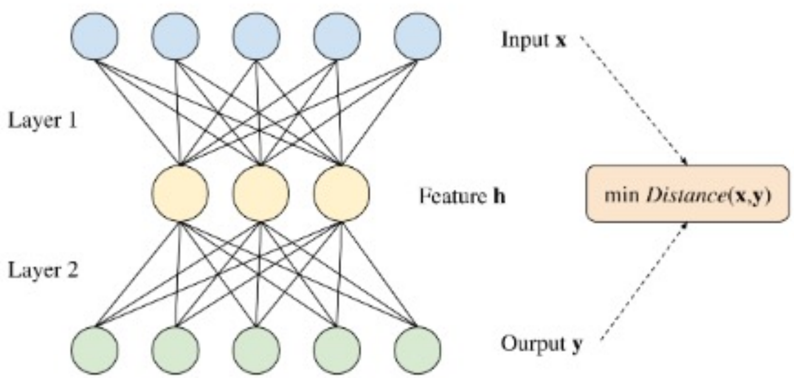
Auto-encoder (AE): original model

Autoencoder could be trained as a good initialization in a supervised model.

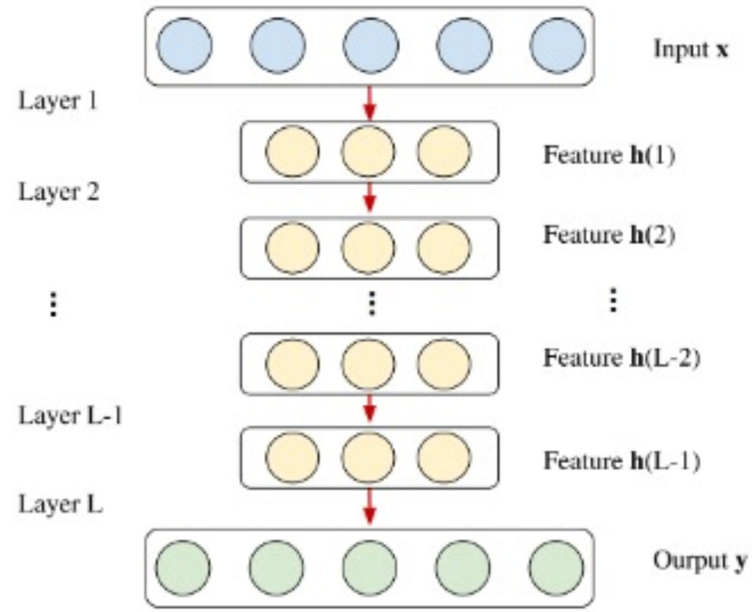


Autoencoder (AE): stacked autoencoder

Autoencoder could be stacked to form a deep model and generate a hierarchy of latent features.



Original AE

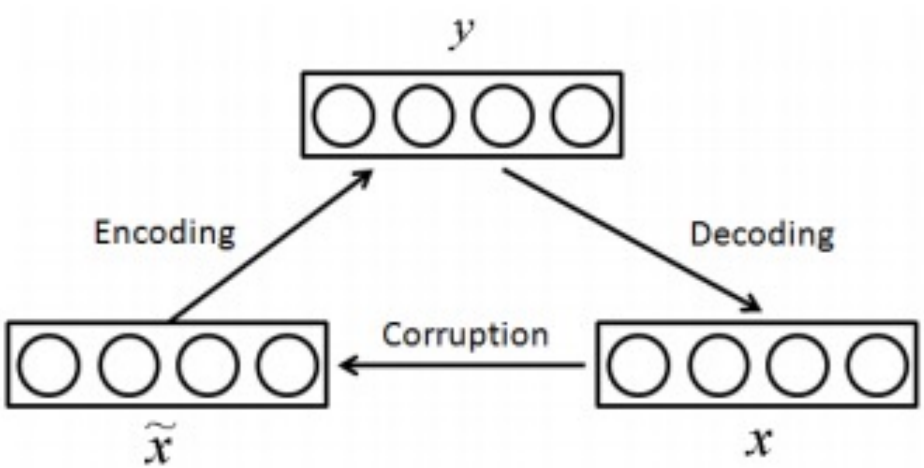


Stacked AE

Auto-encoder (AE): denoising autoencoder

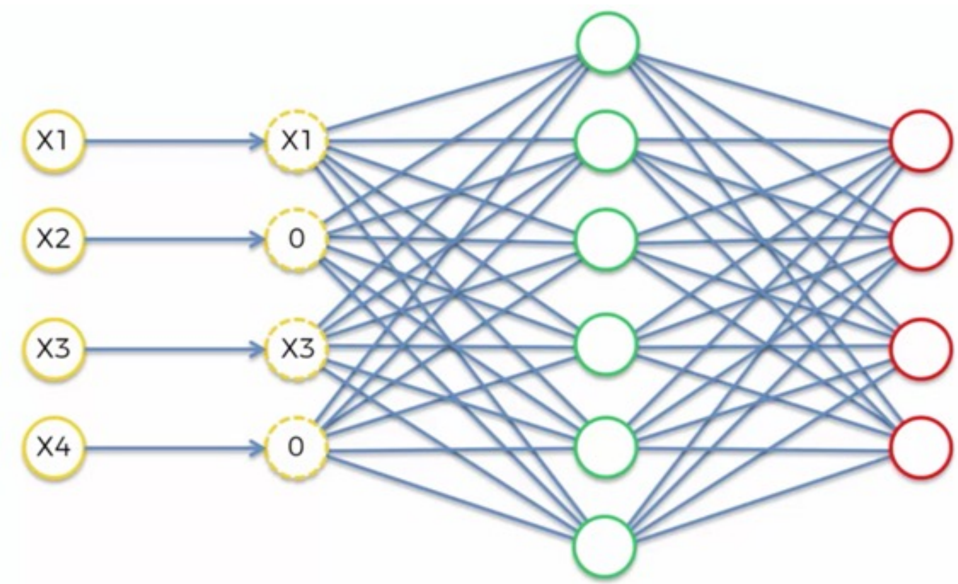
Denoising autoencoder:

In order to force the hidden layer to discover robust features and prevent it from simply learning the identity function, we want to train the autoencoder to reconstruct the input from a corrupted version of it



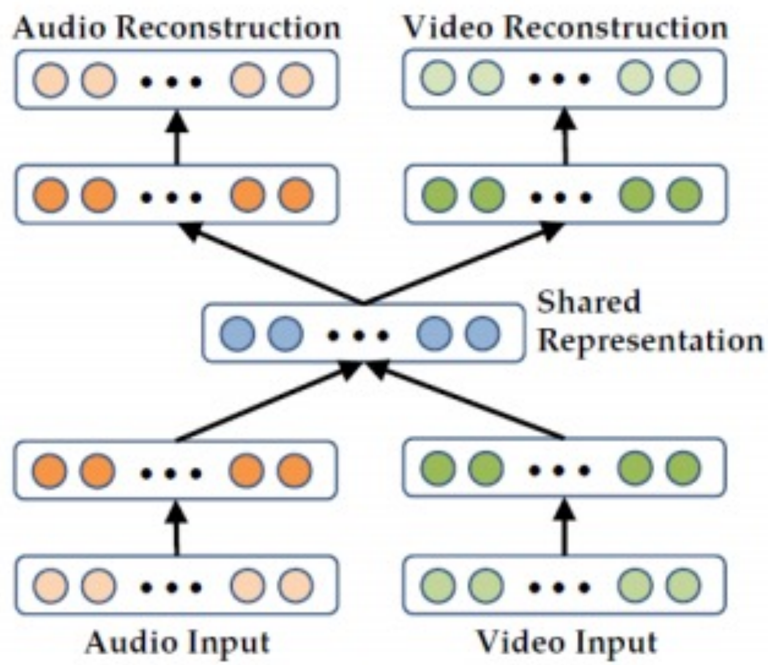
Auto-encoder (AE): denoising autoencoder

- To convert the autoencoder to a denoising autoencoder, we can randomly sets some of the inputs (as many as half of them) to zero.



- The input can be corrupted in other ways

Auto-encoder (AE): multimodal autoencoder



Bimodal Deep Autoencoder

Example:
One third of training data has only video
One third of training data has only audio
One third of training data has both video and audio

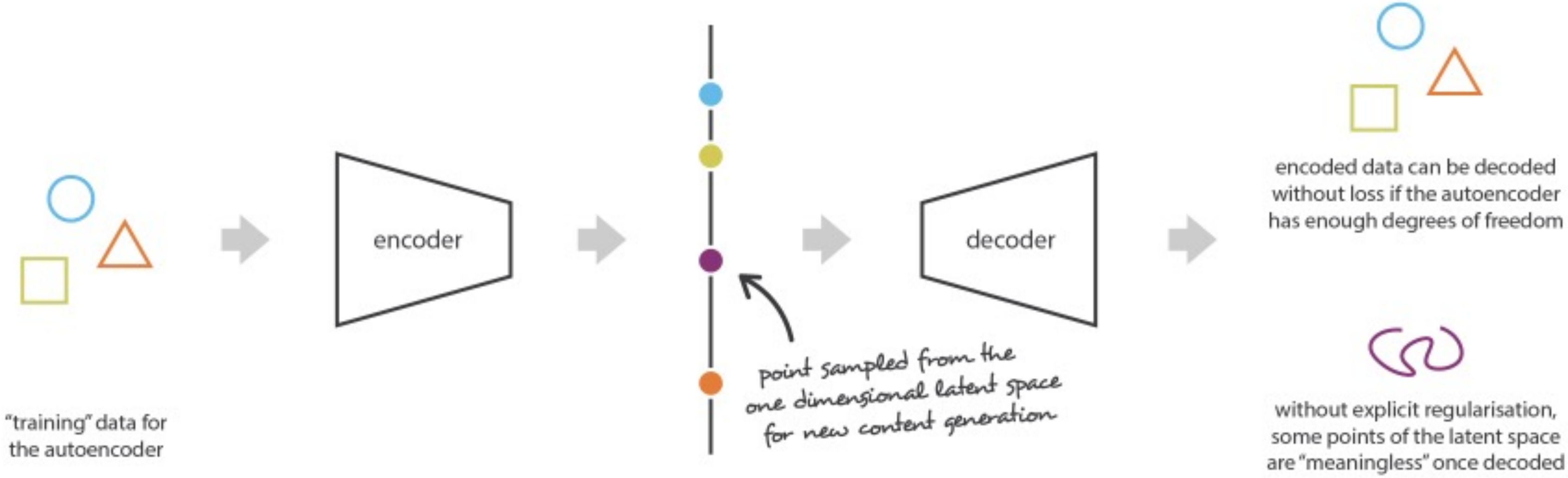
Variational Auto-encoder (VAE)

- Autoencoder gives low reconstruction error on test examples from the same distribution as the training examples, but generally high reconstruction error on samples randomly chosen from the input space.

- Variational auto-encoder:

  - Probabilistic spin on autoencoders - enable sampling from the model to generate data

Variational Auto-encoder (VAE)



Irregular latent space prevent us from using autoencoder for new content generation.

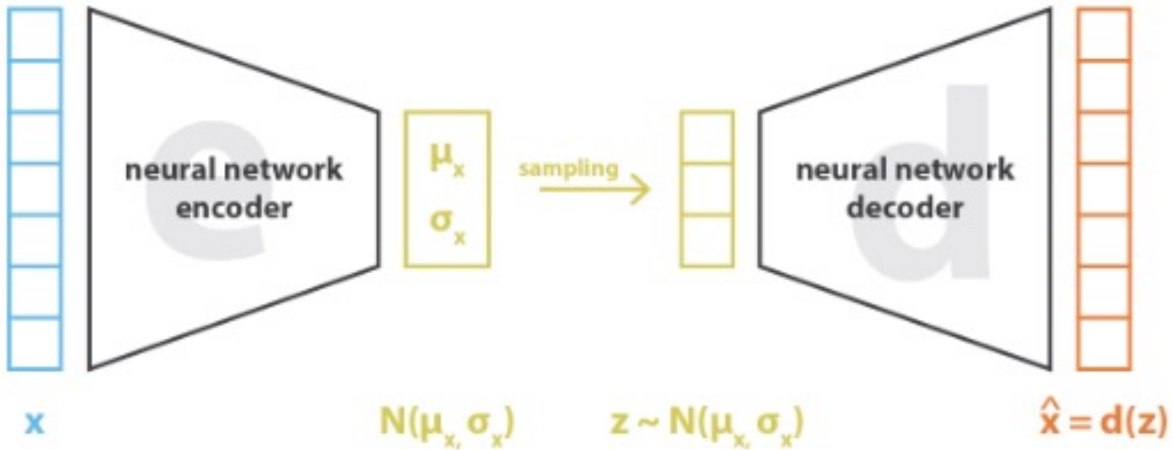
Variational Auto-encoder (VAE)

Difference between autoencoder (deterministic) and variational autoencoder (probabilistic).

Instead of encoding an input as a single point, VAE encodes it as a distribution over the latent space.

Variational Auto-encoder (VAE)

VAE could be treated as “regularized” autoencoder.

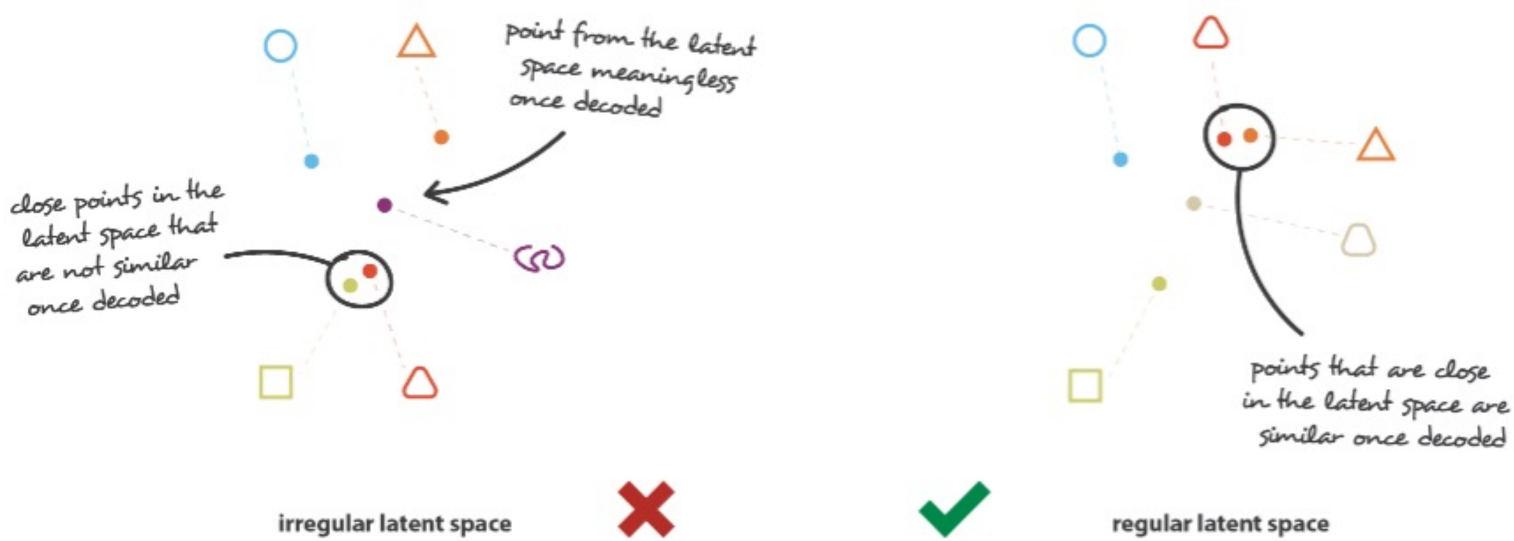


$$\text{loss} = \| x - \hat{x} \|^2 - KL[\, N(\mu_x, \sigma_x), N(0, I)\,] = \boxed{\| x - d(z) \|^2} - \boxed{KL[\, N(\mu_x, \sigma_x), N(0, I)\,]}$$

Reconstruction term

Regularization term

Variational Auto-encoder (VAE)



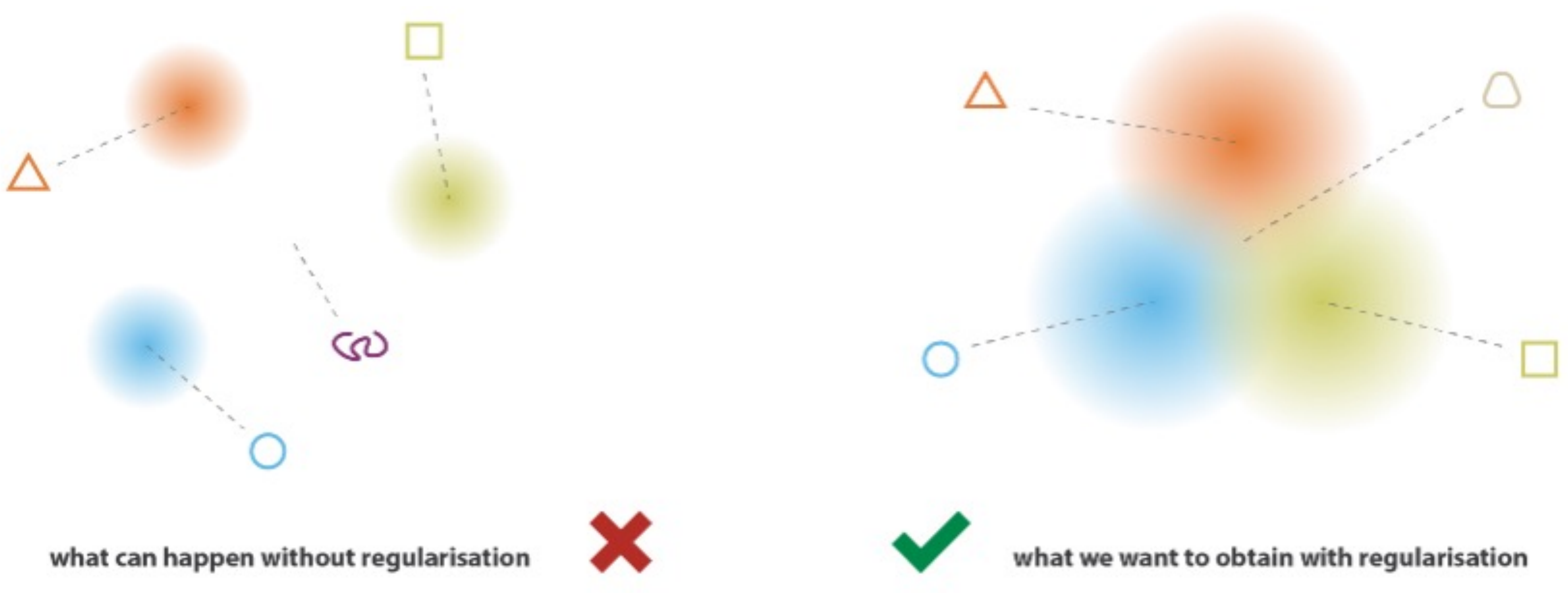
Difference between a "regular" and an "irregular" latent space.

A "regular" latent space requires :

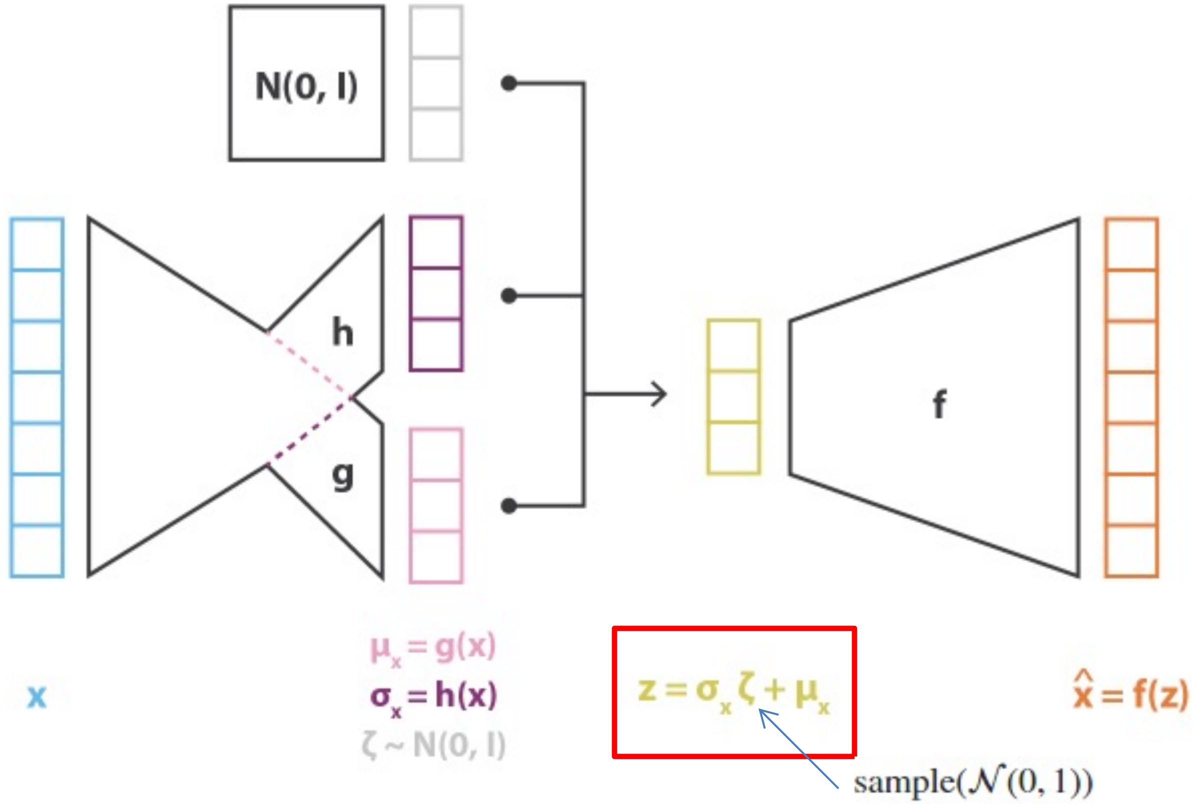
(1) continuity : wo close points should not give two completely different contents once decoded

(2) completeness : a point sampled from the latent space should give "meaningful" content once decoded

Variational Auto-encoder (VAE)



Variational Auto-encoder (VAE)



$$\text{loss} = C\,\|x-\hat{x}\|^2 - \mathrm{KL}[\,N(\mu_x,\sigma_x), N(0,I)\,] = C\,\|x-f(z)\|^2 - \mathrm{KL}[\,N(g(x), h(x)), N(0,I)\,]$$

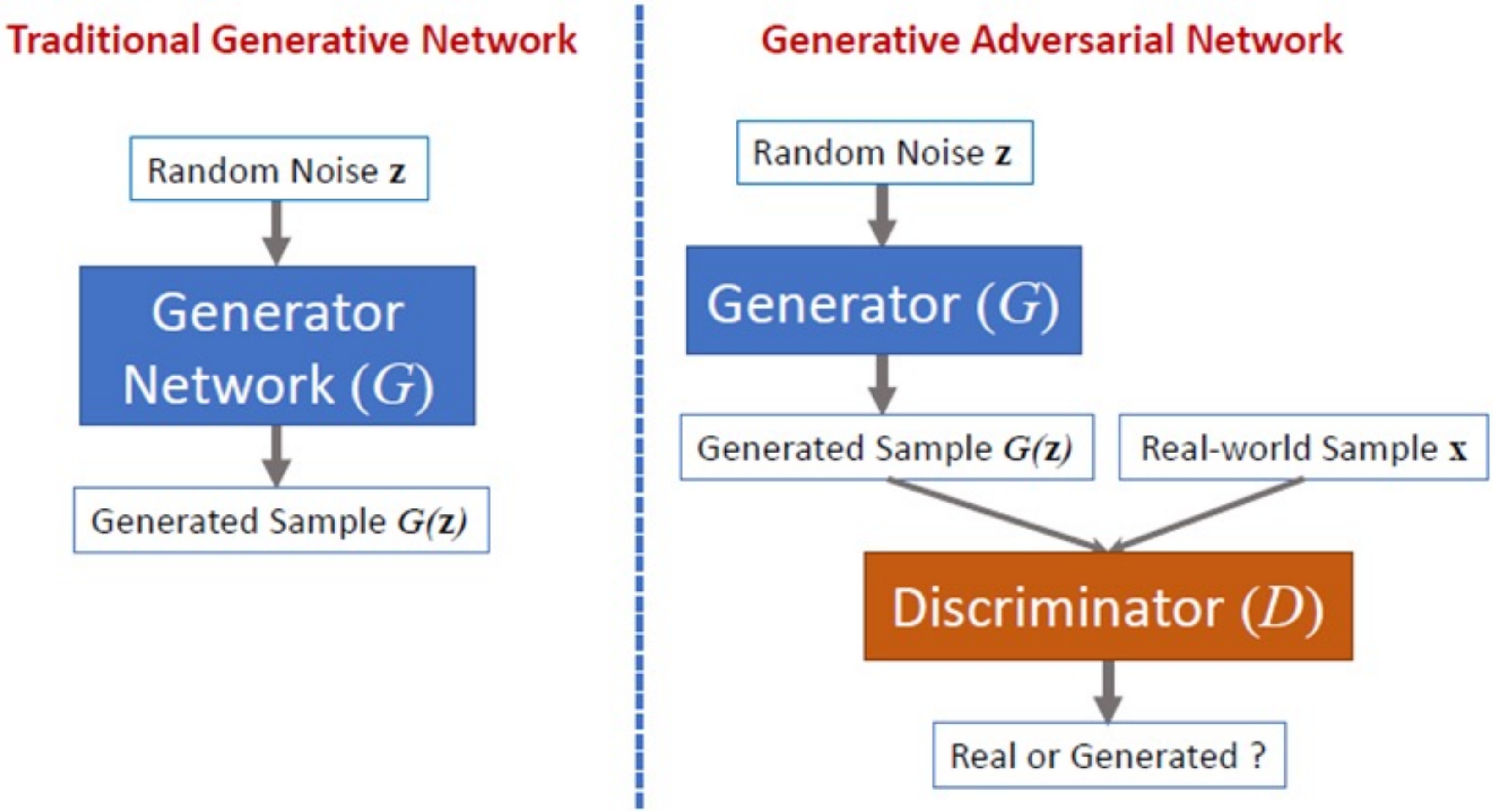
Variational Auto-encoder (VAE)

Generating celebrity-look like photos

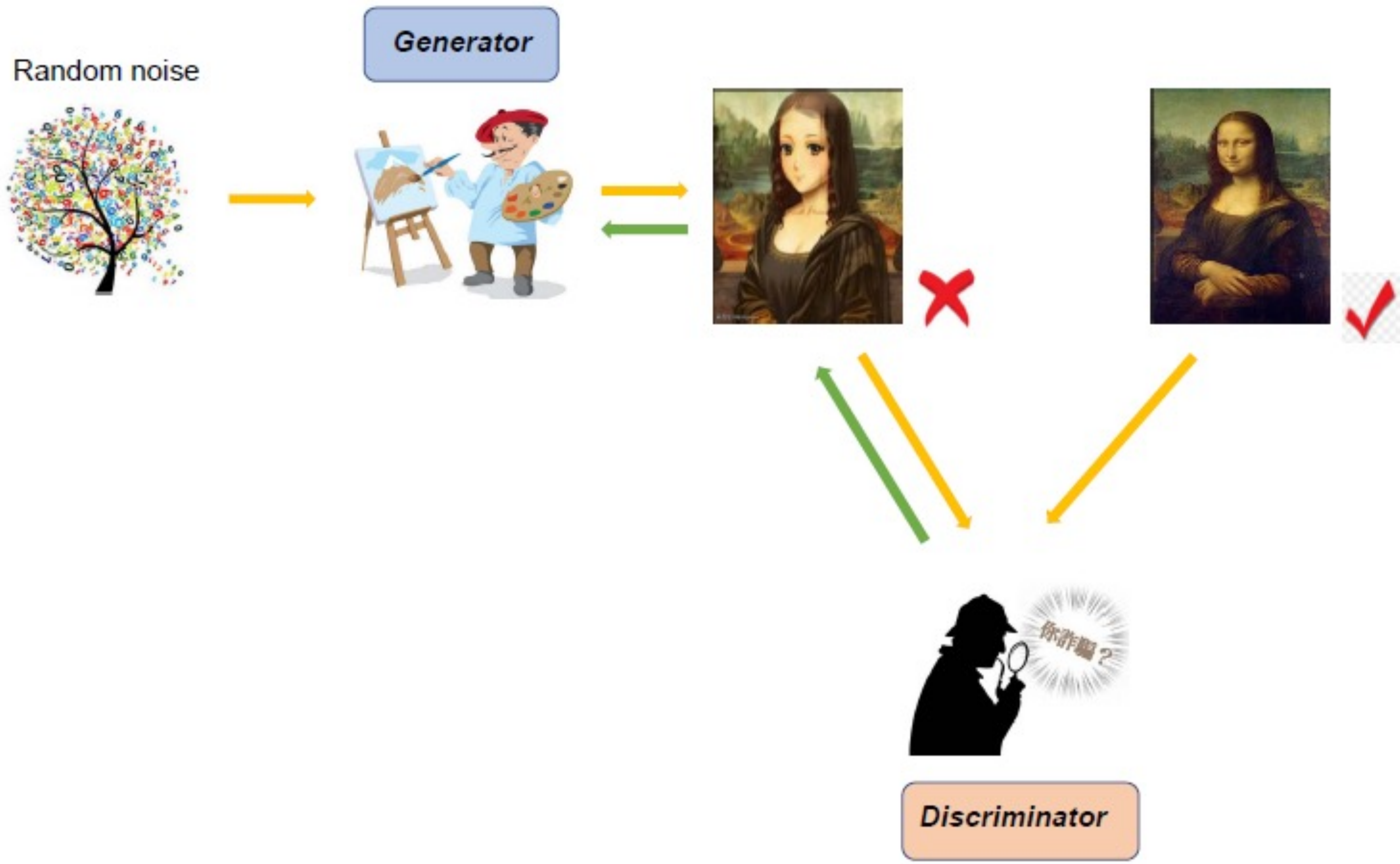
Outline

- Generative models
- Variational Autoencoder (VAE)
- Generative Adversarial Network (GAN)

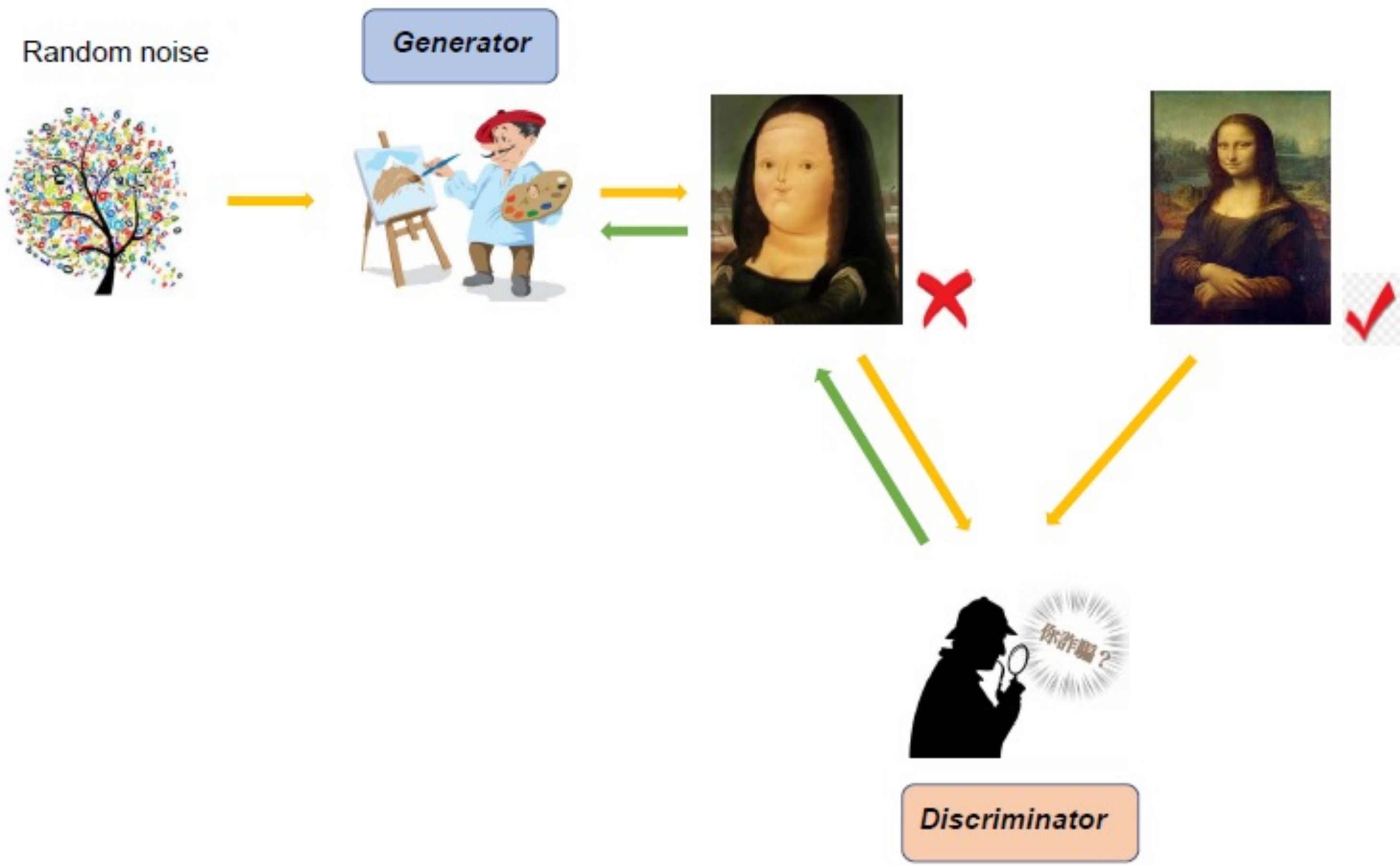
Generative Adversarial Network (GAN)



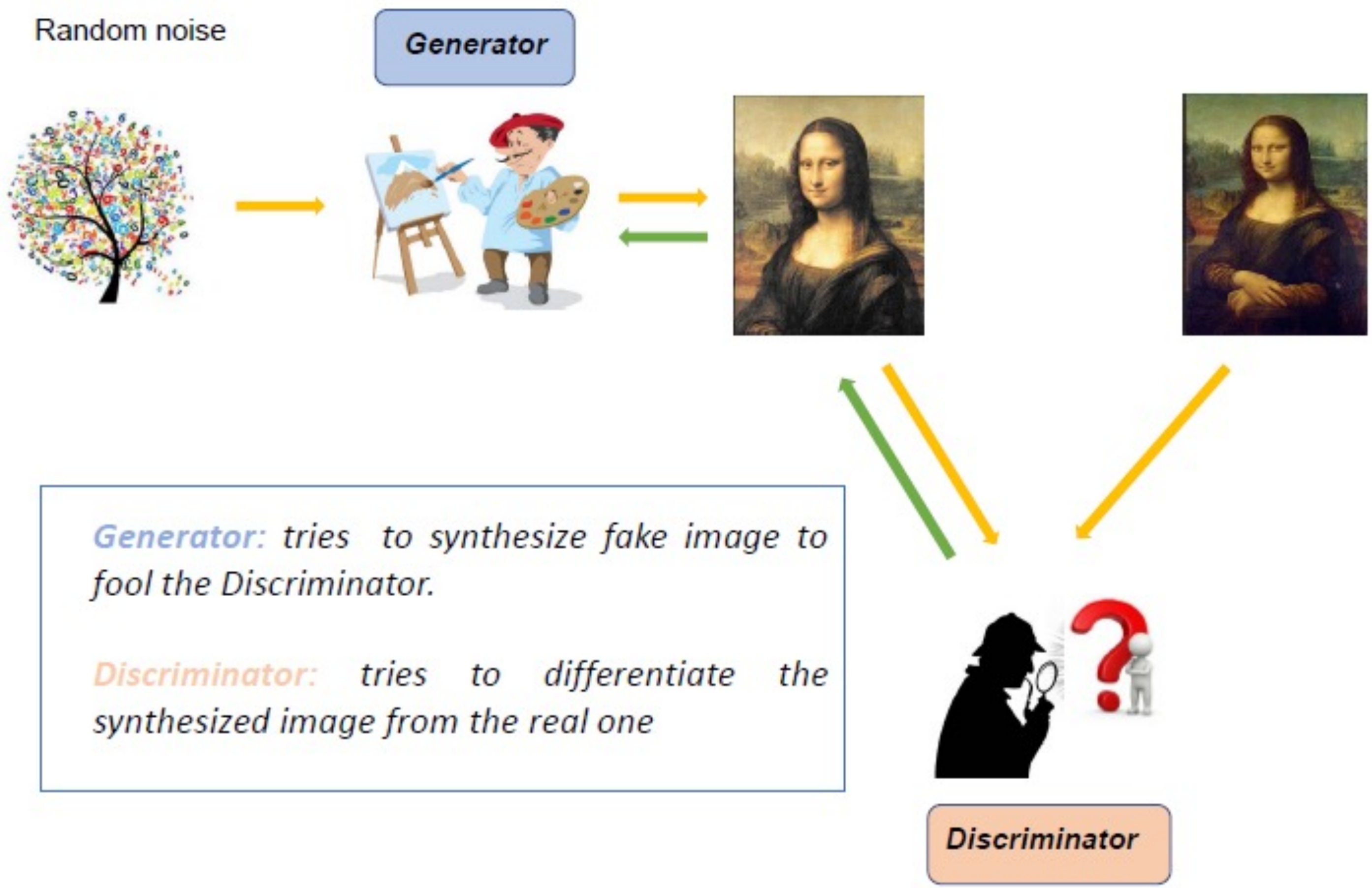
Generative Adversarial Network (GAN)



Generative Adversarial Network (GAN)



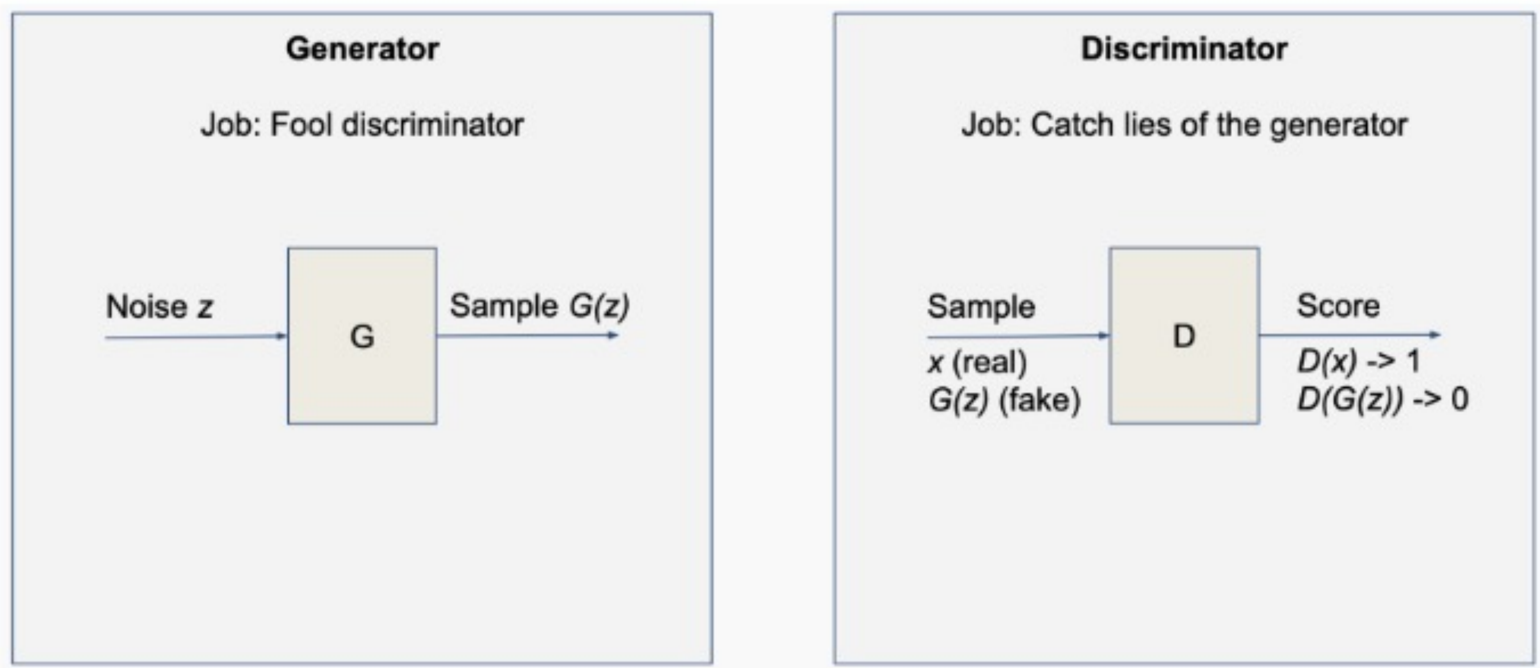
Generative Adversarial Network (GAN)



Generator: tries to synthesize fake image to fool the Discriminator.

Discriminator: tries to differentiate the synthesized image from the real one

Generative Adversarial Network (GAN)



Generative Adversarial Network (GAN)

- The Discriminator is trying to maximize its reward $\boldsymbol{V(D, G)}$
- The Generator is trying to minimize Discriminator's reward (or maximize its loss)

$$V(D,G) = \boxed{\mathbb{E}_{x\sim p(x)}[\log D(x)]} + \boxed{\mathbb{E}_{z\sim q(z)}\left[\log\left(1 - D(G(z))\right)\right]}$$

The overall objective of GAN is:

$$\min_G \max_D V(D,G) = \min_G \max_D \mathbb{E}_{x\sim p(x)}[\log D(x)] + \boxed{\mathbb{E}_{z\sim q(z)}\left[\log\left(1 - D(G(z))\right)\right]}$$

Generative Adversarial Network (GAN)

Conditional GAN:

GANs could be extended to conditional model if both the generator and the discriminator are conditioned on extra information $\boldsymbol{y}$

$$\min_G \max_D V(D,G) = \min_G \max_D \mathbb{E}_{x\sim p(x)}[\log D(\boldsymbol{x}|\boldsymbol{y})] + \mathbb{E}_{z\sim q(z)}[\log(1 - D(G(\boldsymbol{z}|\boldsymbol{y})))]$$

Generative Adversarial Network (GAN)

Pix2Pix GAN:

Translation between image pairs:
The input image **x** and the target image **y** has one-to-one correspondence

Edges to Photo

Day to Night

BW to Color

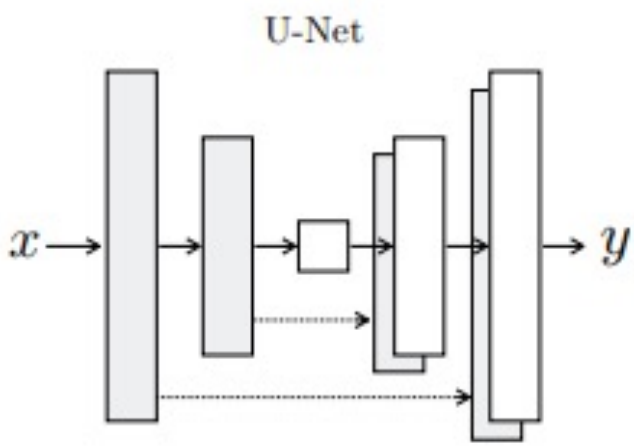
Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros , Image-to-Image Translation with Conditional Adversarial Networks, CVPR, 2017.

Generative Adversarial Network (GAN)

Generator loss:

$$\mathcal{L}^{G}_{cGAN} = \mathbb{E}_{\mathbf{x} \sim P_{data}(\mathbf{x})}[\log(1 - D(\mathbf{x}, G(\mathbf{x})))$$
$$+ \boxed{\lambda_{l1} \mathbb{E}_{\mathbf{x},\mathbf{y} \sim P_{data}(\mathbf{x},\mathbf{y})}[\|y - G(\mathbf{x})\|_1]}$$

Pixel –to-pixel
intensity loss



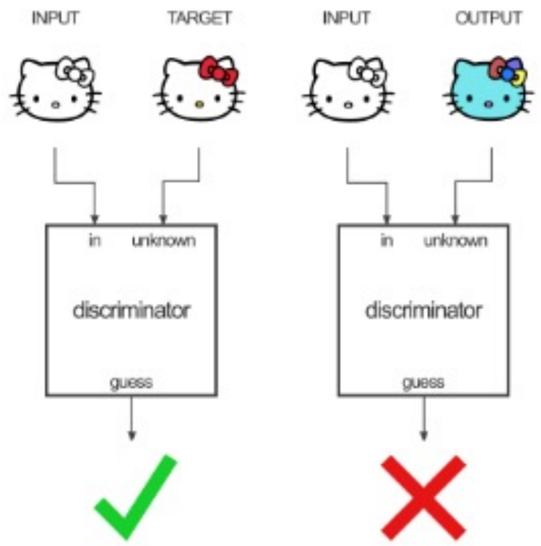
Generator

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros , Image-to-Image Translation with Conditional Adversarial Networks, CVPR, 2017.

Generative Adversarial Network (GAN)

Discriminator loss:

$$\mathcal{L}_{cGAN}^{D} = -\mathbb{E}_{\mathbf{x},\mathbf{y}\sim P_{data}(\mathbf{x},y)}[\log D(\mathbf{x},\mathbf{y})] - \mathbb{E}_{\mathbf{x}\sim P_{data}(\mathbf{x})}[\log(1 - D(\mathbf{x}, G(\mathbf{x})))].$$



Overall loss:

$$\mathcal{L}_{cGAN} = \mathcal{L}_{cGAN}^{G} + \mathcal{L}_{cGAN}^{D}$$

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros , Image-to-Image Translation with Conditional Adversarial Networks, CVPR, 2017.

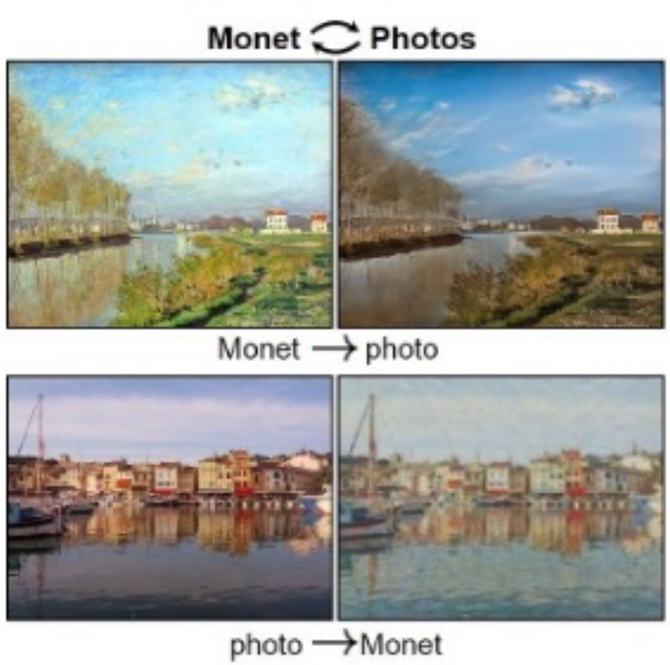
Generative Adversarial Network (GAN)

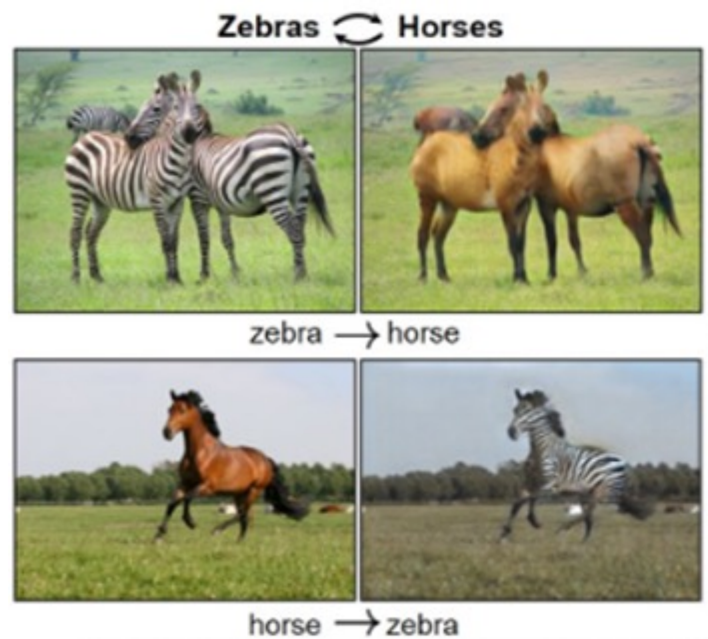
Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros , Image-to-Image Translation with Conditional Adversarial Networks, CVPR, 2017.

Generative Adversarial Network (GAN)

CycleGAN:

Translation between **unpaired** images:
The input image **x** and the target image **y** do **not** have correspondence

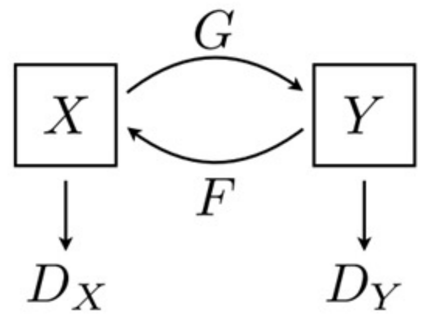




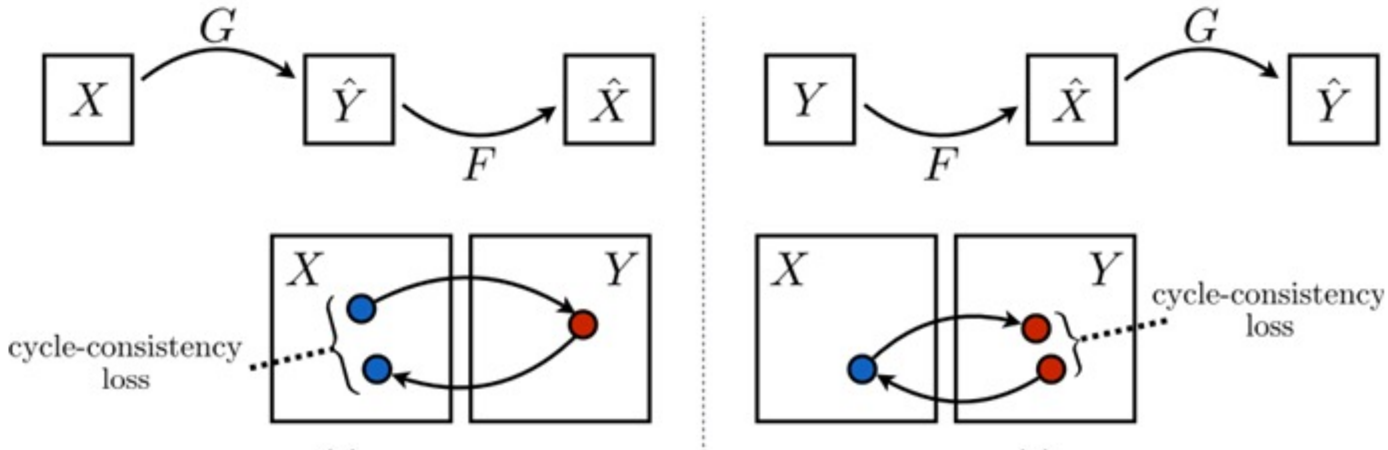
Jun-Yan Zhu*, Taesung Park*, Phillip Isola, Alexei A. Efros. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, ICCV 2017

Generative Adversarial Network (GAN)

CycleGAN:



(a)



(b) (c)

Jun-Yan Zhu*, Taesung Park*, Phillip Isola, Alexei A. Efros. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, ICCV2017

Generative Adversarial Network (GAN)

CycleGAN:

- Two generators,

G learns the mapping $X \rightarrow Y$

F learns the mapping $Y \rightarrow X$

- Two discriminators

D_X differentiates $\hat{\mathbf{x}} = F(G(\mathbf{x}))$ from $\mathbf{x}$

D_Y differentiates $\hat{\mathbf{y}} = G(F(\mathbf{y}))$ from $\mathbf{y}$

Generative Adversarial Network (GAN)

CycleGAN

Loss function:

$$\mathcal{L}(G, F, D_X, D_Y) = \underbrace{\mathcal{L}_{GAN}^{G,D_Y,X,Y} + \mathcal{L}_{GAN}^{G,D_X,Y,X}}_{\text{Common GAN loss}} + \underbrace{\boxed{\lambda \mathcal{L}_{cycle}^{G,F}}}_{\text{Cycle consistency}}$$

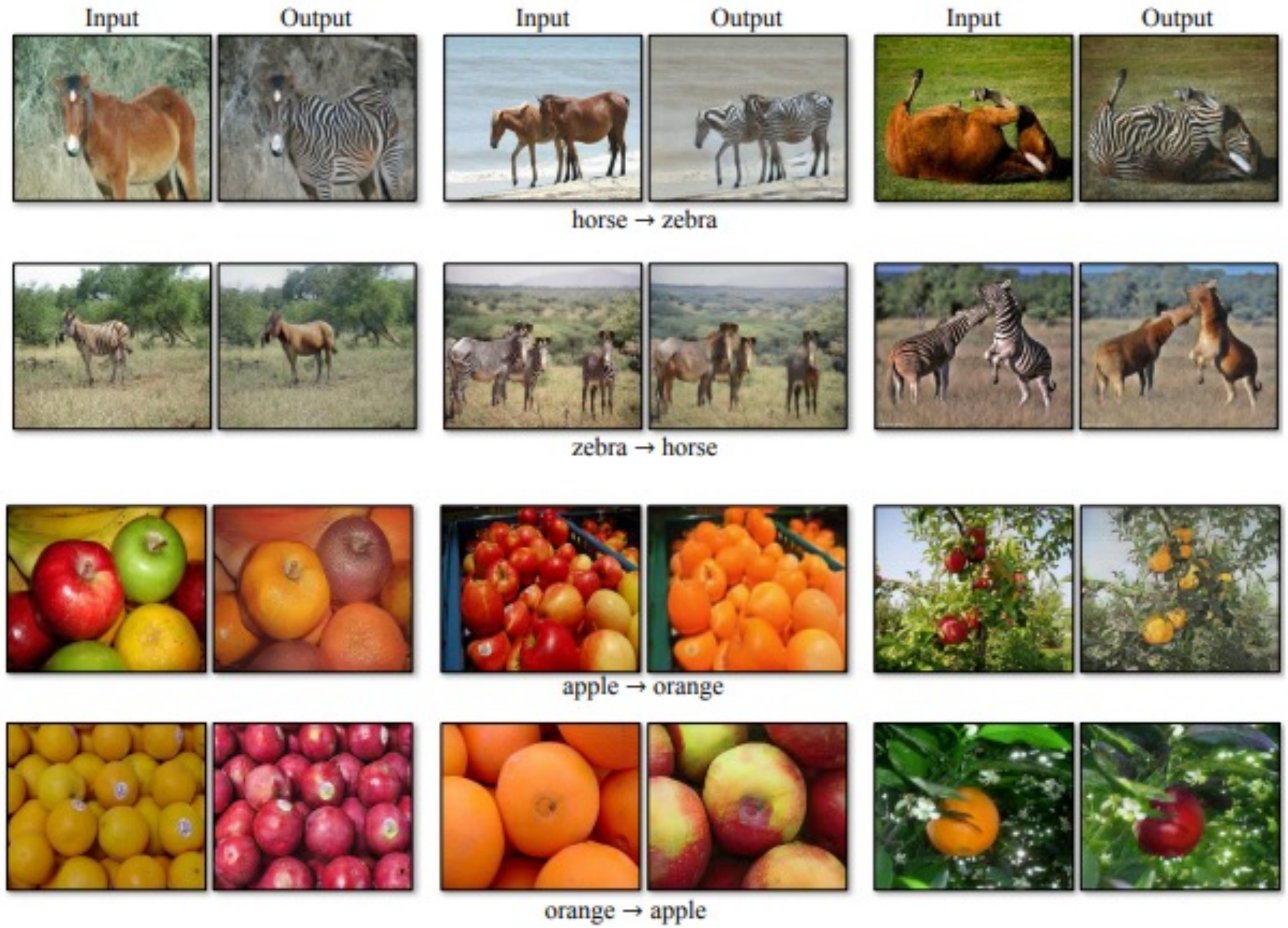
Cycle consistency term:

$$\mathcal{L}_{cycle}^{G,F} = \mathbb{E}_{\mathbf{x} \sim P_{data}(\mathbf{x})}[\|F(G(\mathbf{x})) - \mathbf{x}\|_1] + \mathbb{E}_{\mathbf{y} \sim P_{data}(\mathbf{y})}[\|G(F(\mathbf{y})) - \mathbf{y}\|_1]$$

Generative Adversarial Network (GAN)

Jun-Yan Zhu*, Taesung Park*, Phillip Isola, Alexei A. Efros. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, ICCV2017

Generative Adversarial Network (GAN)



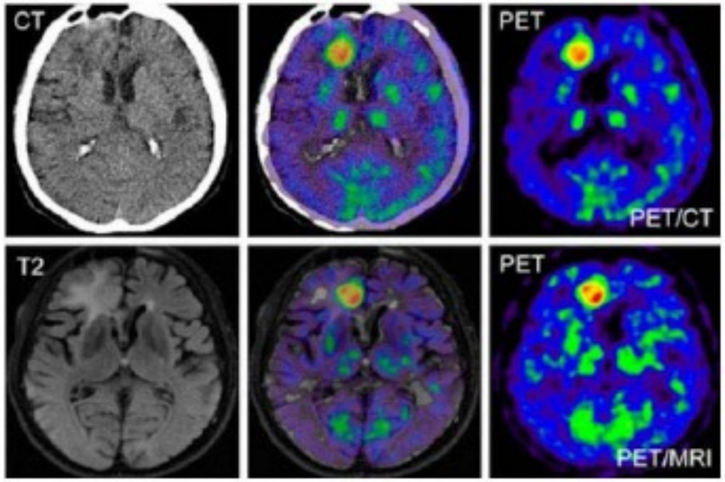
Jun-Yan Zhu*, Taesung Park*, Phillip Isola, Alexei A. Efros. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, ICCV2017

A Case Study

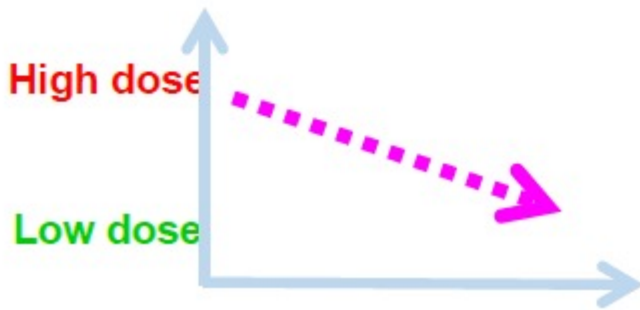
Y. Wang, B. Yu, L. Wang, C. Zu, D.S. Lalush, W. Lin, X. Wu, J. Zhou, D. Shen*, and **L. Zhou***, "**3D Conditional Generative Adversarial Networks for High-quality PET Image Estimation at Low Dose",** Neuroimage, 2018

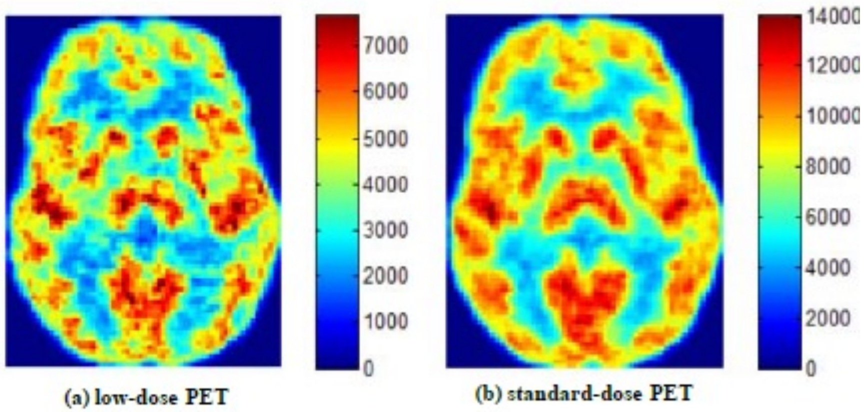
Estimating PET images from low dose

- Positron emission tomography (PET) can produce a 3D image of the functional processes in the body, i.e., using PET/CT.
  → Widely used in various clinical applications, such as cancer diagnosis.



Estimating PET images from low dose

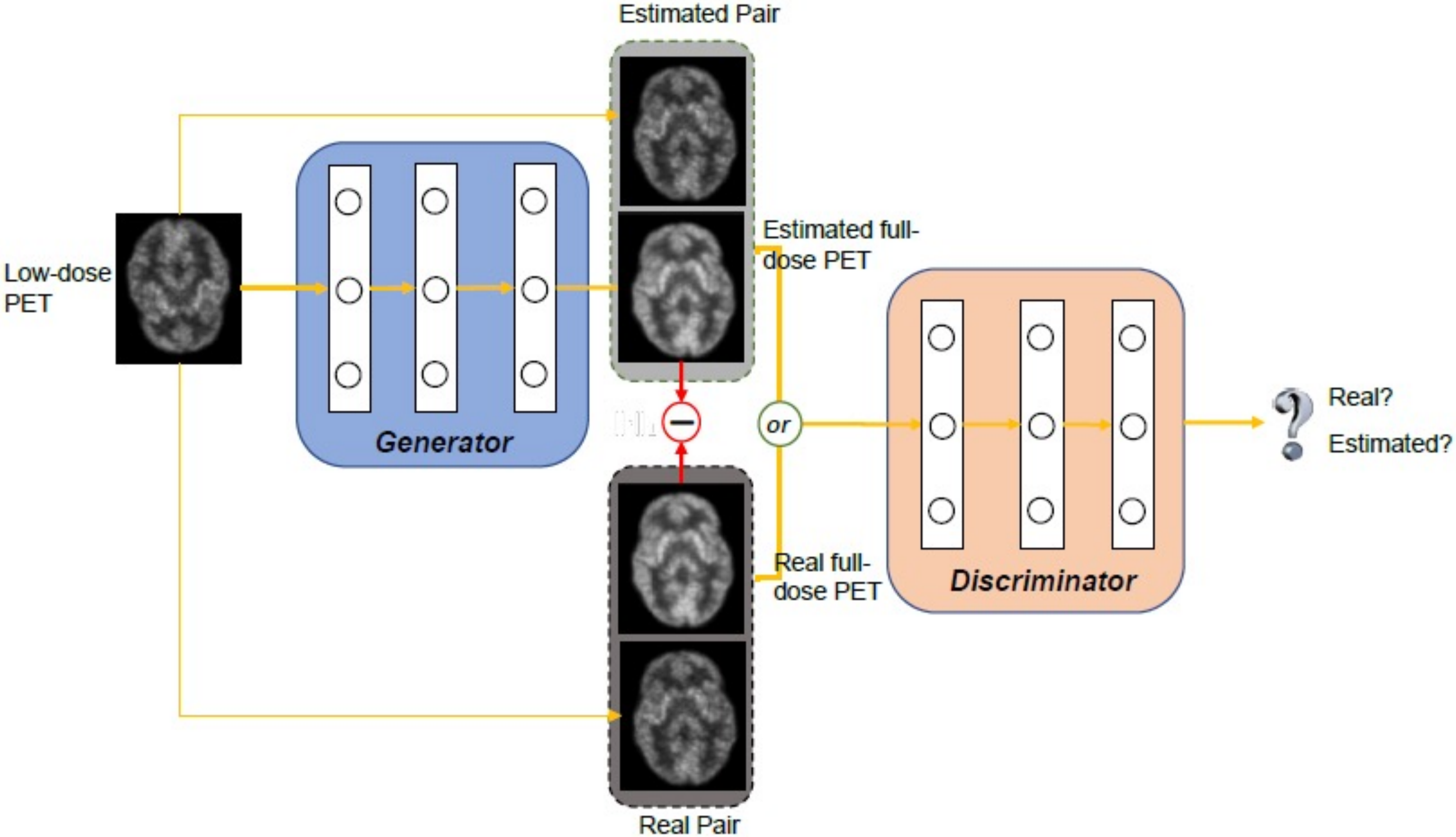




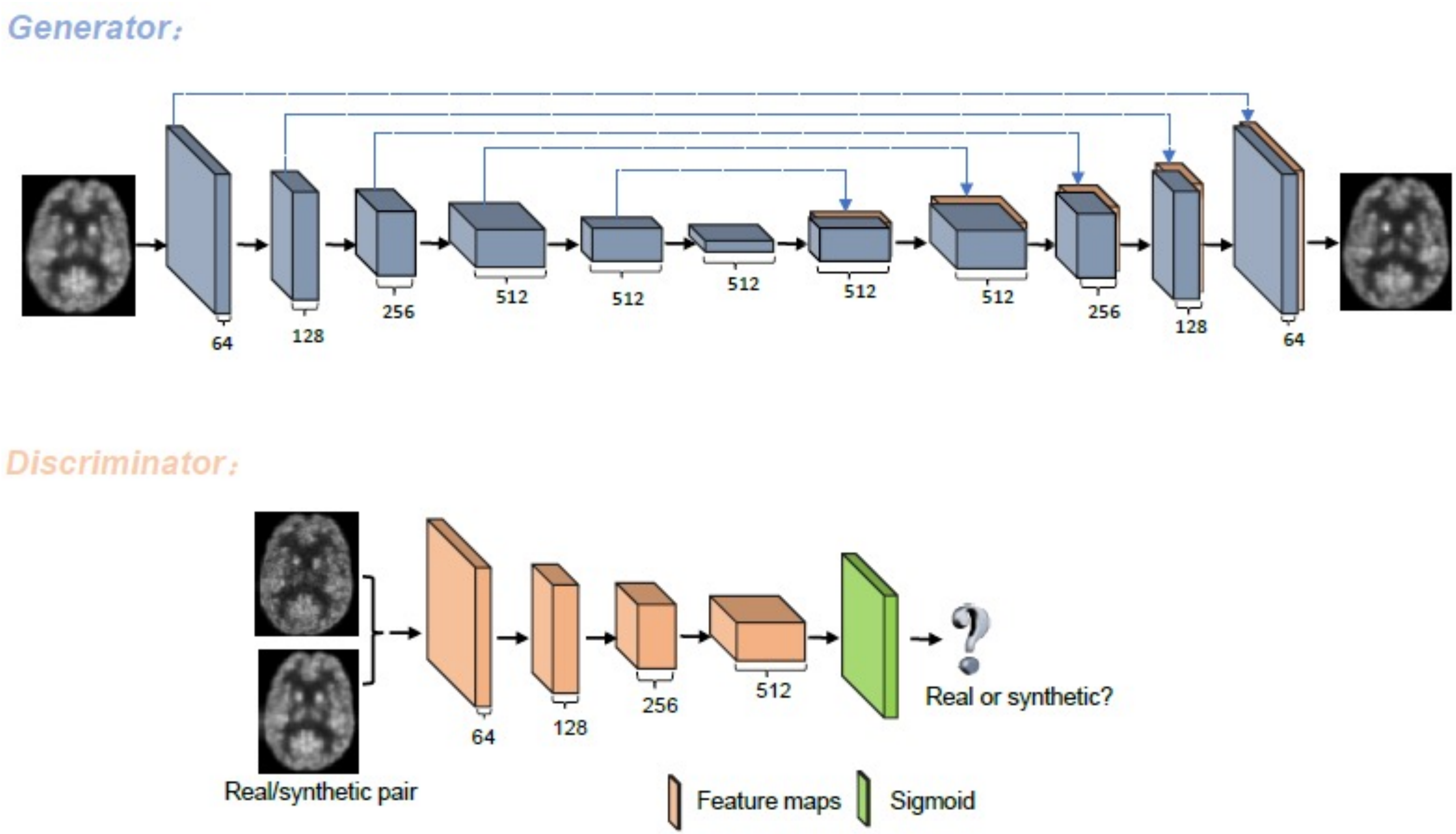
Comparison between low-dose (L-PET) and full-dose PET (F-PET) images

High quality PET image plays an essential role in accurately diagnosing diseases/ disorders and in assessing the response to therapy.

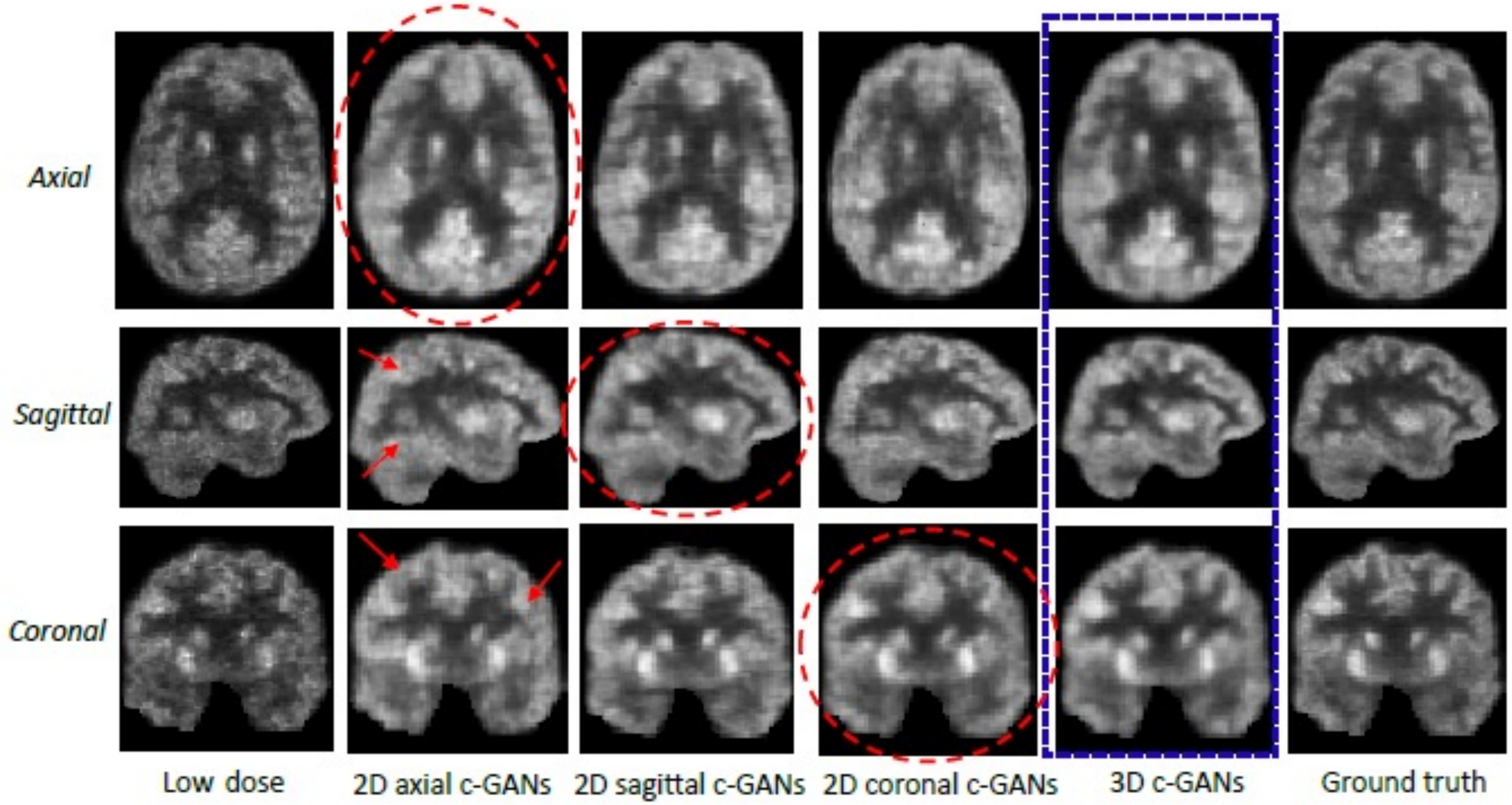
3D c-GAN based method



Estimating PET images from low dose



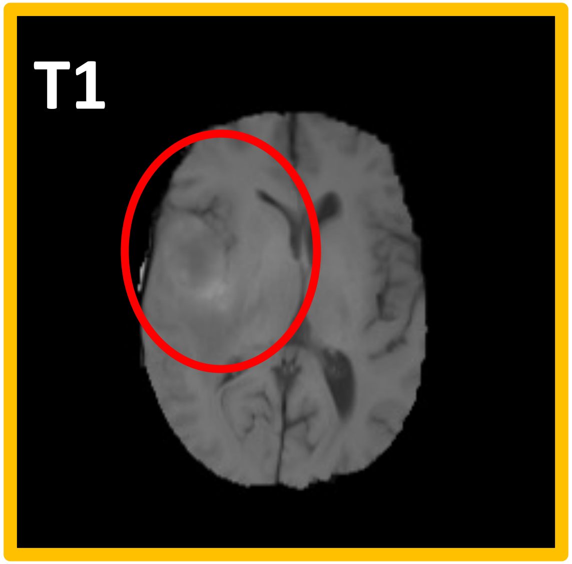
Estimating PET images from low dose

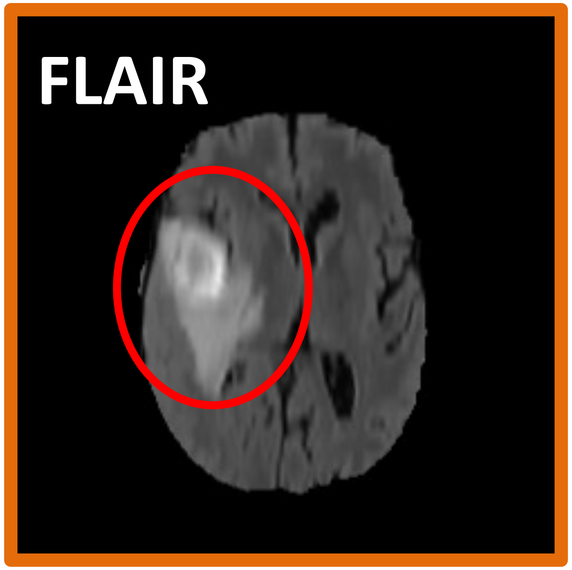


A Case Study II

B. Yu, **L. Zhou**, L. Wang, Y. Shi, J. Fripp, and P. Bourgeat, **"Ea-GANs: Edge-aware Generative Adversarial Networks for Cross-modality MR Image Synthesis**", IEEE Transactions on Medical Imaging (IEEE T-MI), 2019

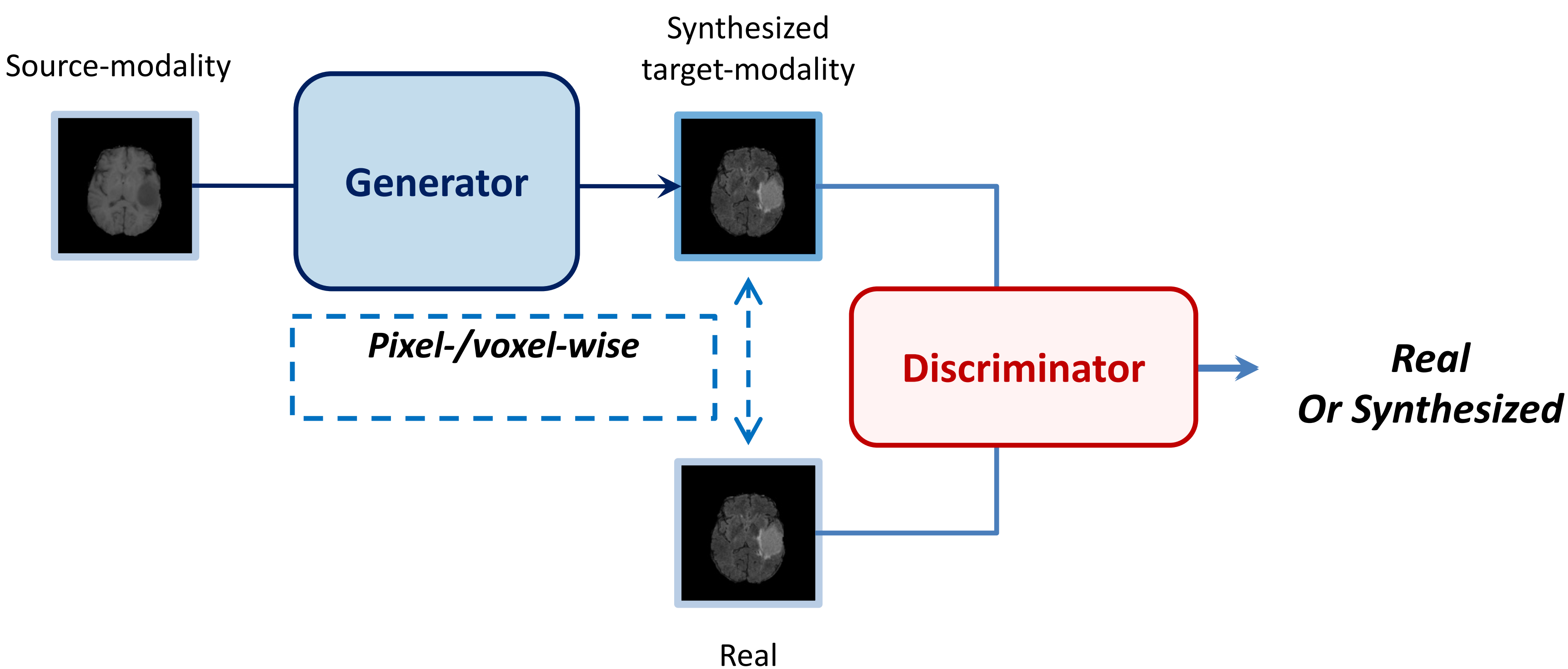
Cross MR modality synthesis





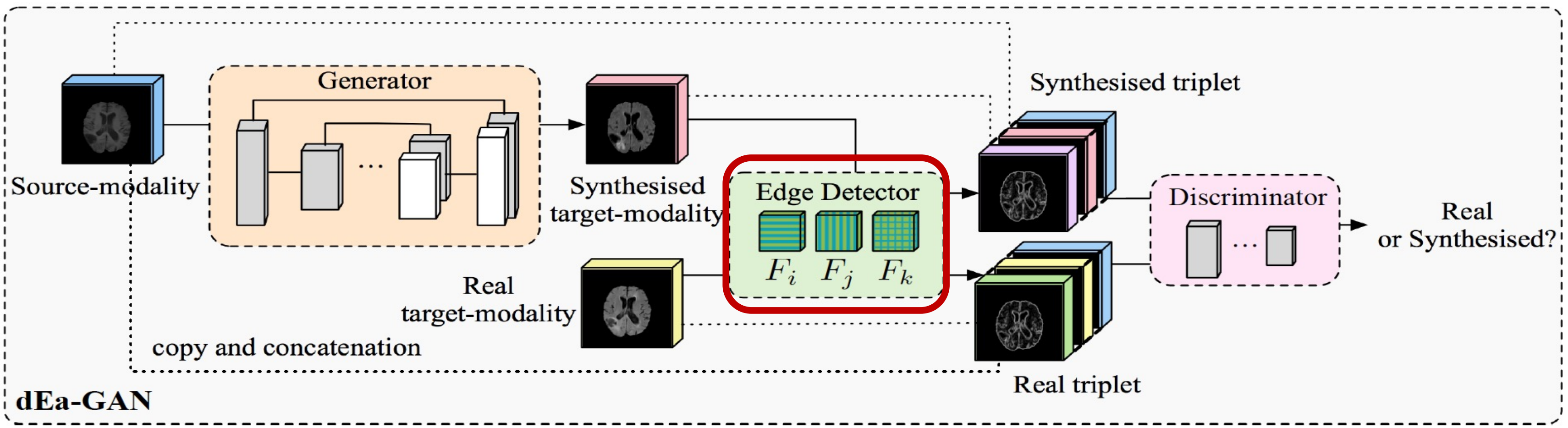
Cross MR modality synthesis

Existing Methods



Cross MR modality synthesis

- ***Discriminator-induced Ea-GAN (dEa-GAN)***
  - incorporate edge information into training both the **generator** and the **discriminator**



Cross MR modality synthesis