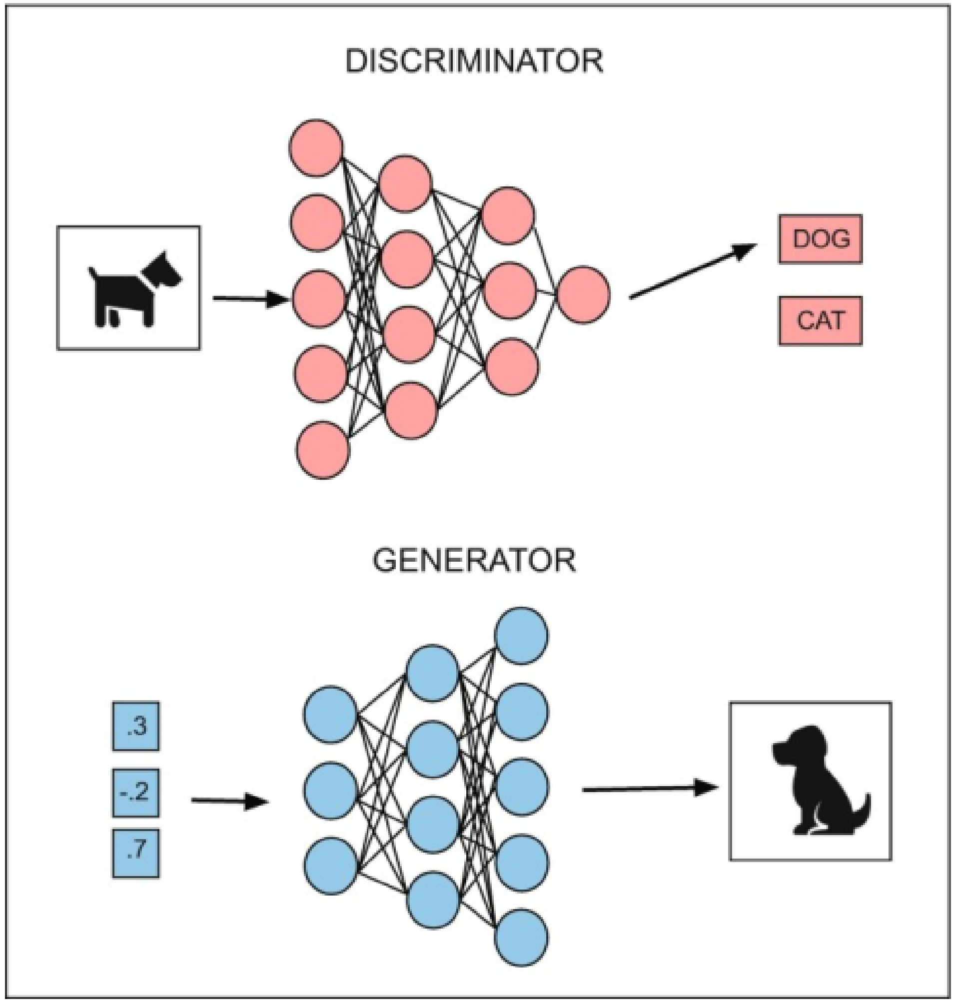


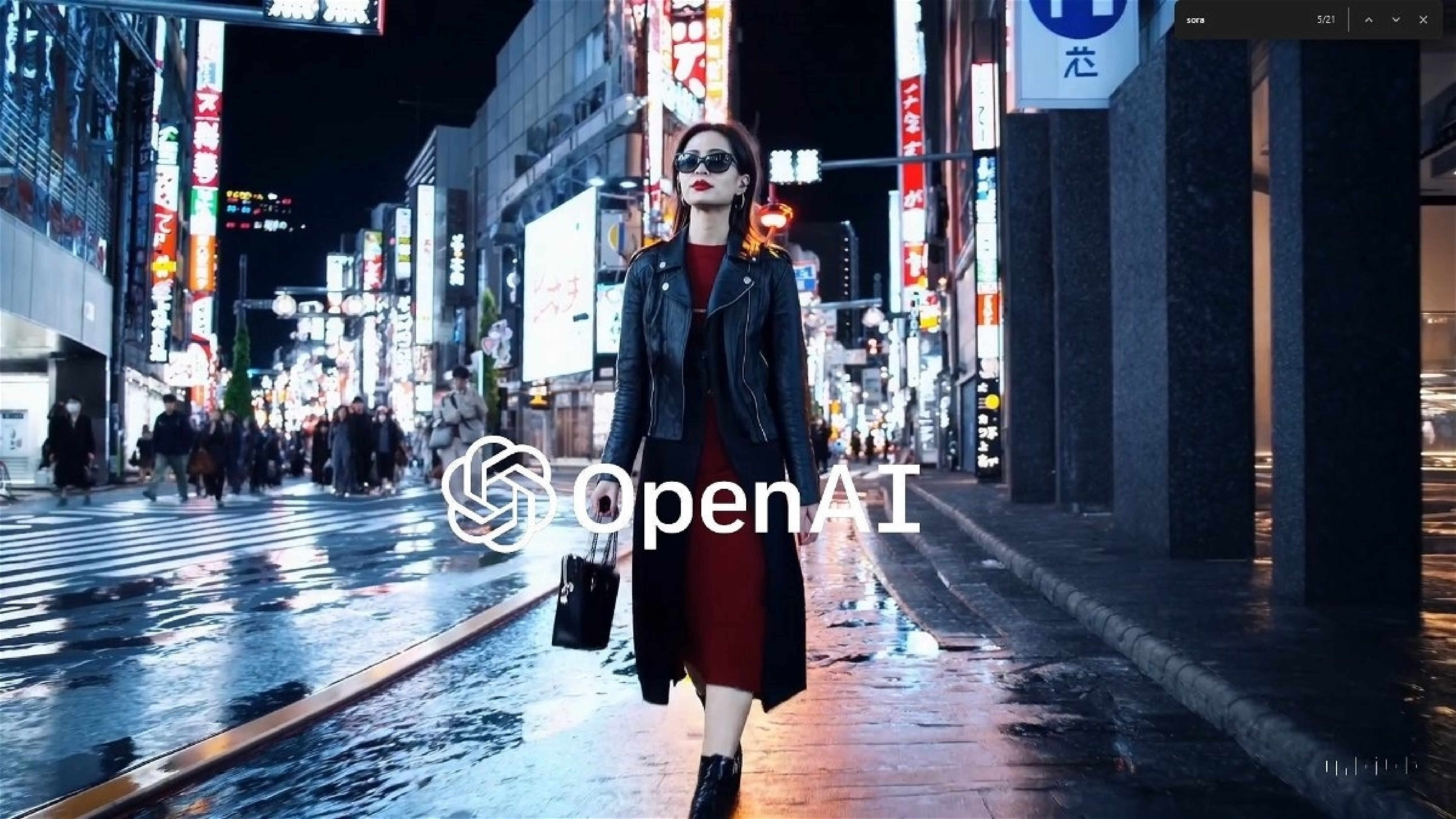
Overview of Generative Models

张江

北京师范大学系统科学学院教授

集智俱乐部、集智学园创始人

集智研究中心理事长


sora
5/21
OpenAI

What is AIGC?

a Hacker and a witch sit together, Hacker working on a portable laptop computer, head covered by the hood, fire flare come out of computer screen, deep dark night, dark forest with tall trees with twisted branches, magical sparkling light in the air, 8K, sharp focus, studio photo, intricate details, highly detailed, by greg rutkowski"


Rokey的Blog


Rokey

Rokey的B

Super Resolution Reconstruction

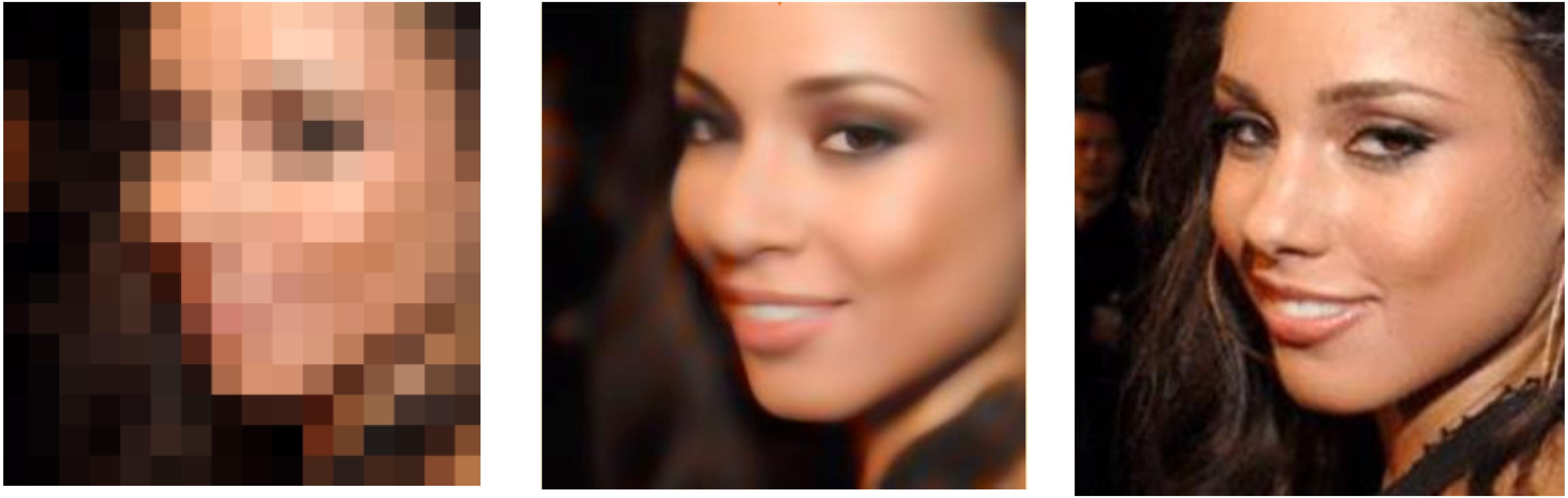
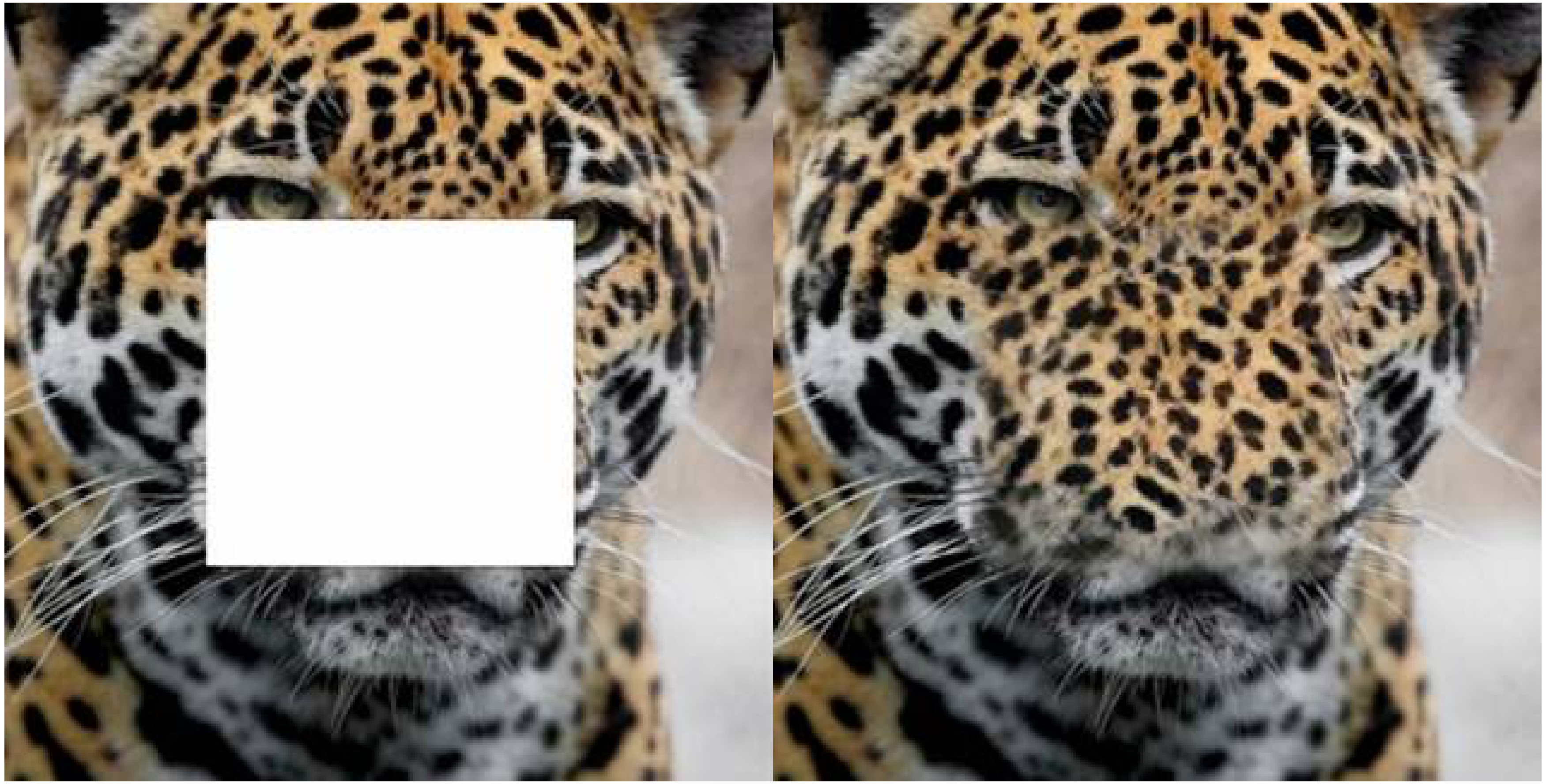
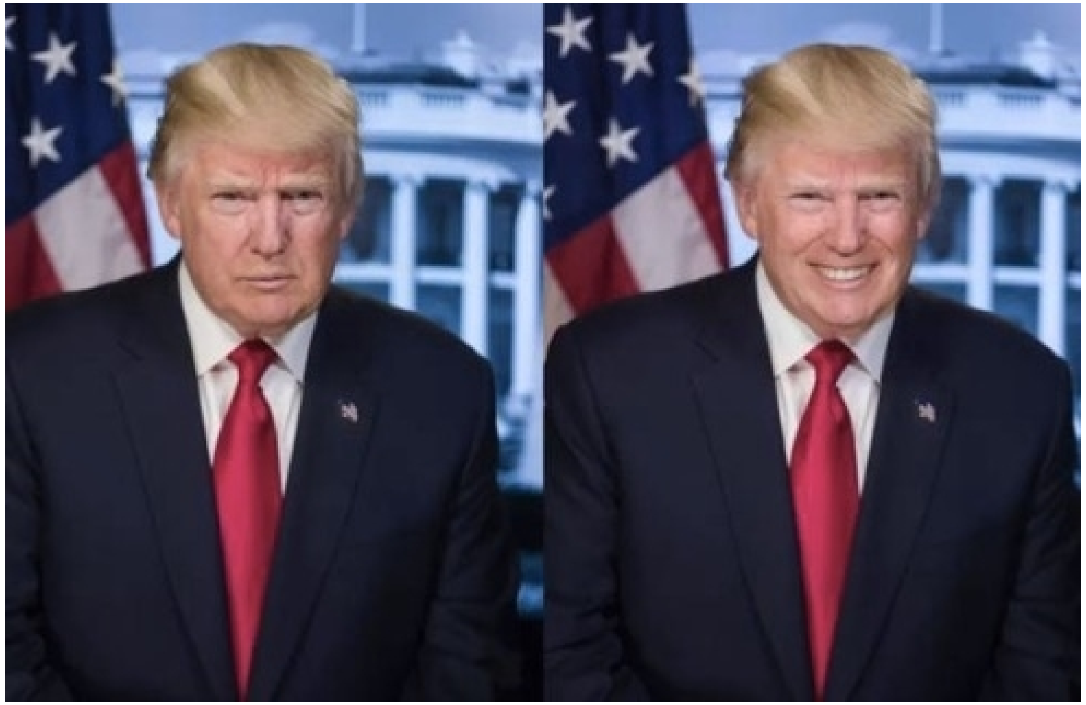
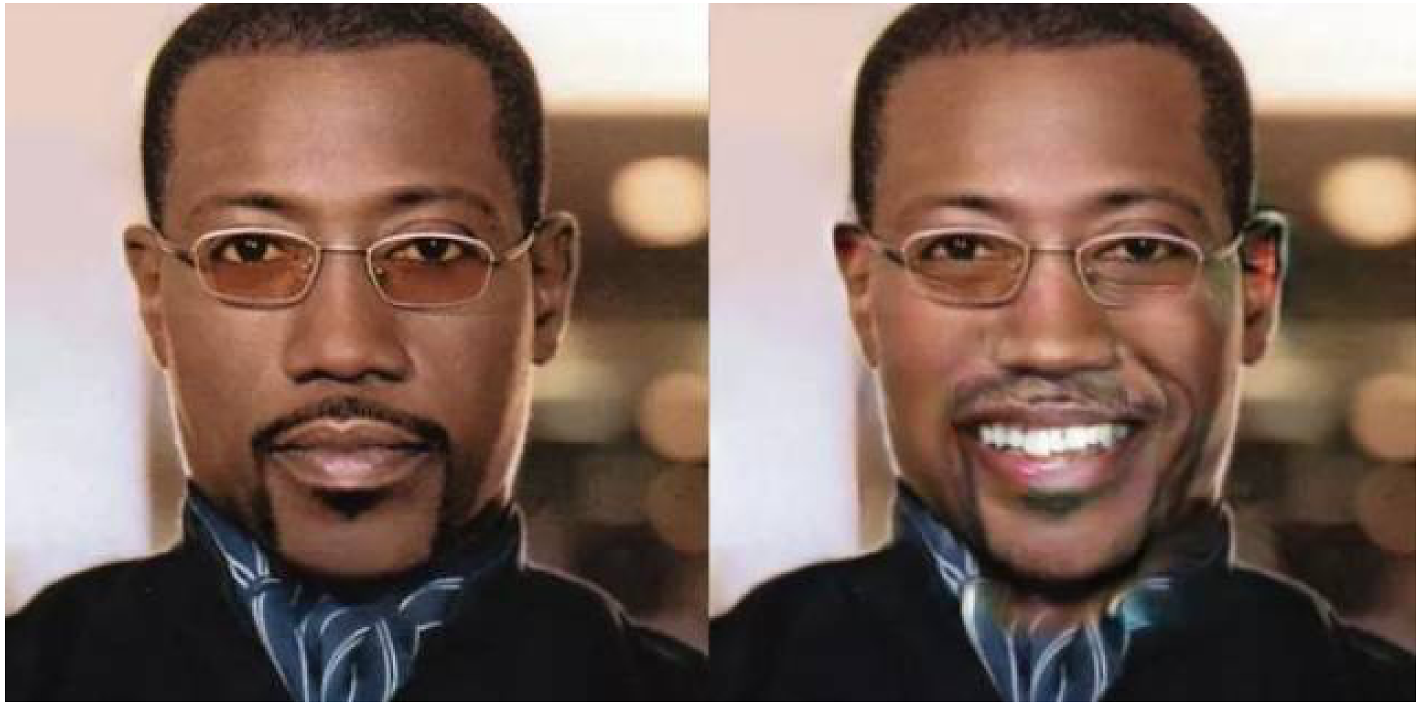


Image Completion



FaceAPP





Replacing object in video

Text-to-Video

very close up of Penguin riding wave on yellow surfboard.

企鹅在黄色冲浪板上乘风破浪的特写镜头

penguin rides surf yellow surfboard unto beach. Penguin leaves yellow surfboard and keeps walking.

企鹅乘着黄色的冲浪板到海滩上。企鹅离开了黄色的冲浪板，继续往前走。

Penguin quickly walking on beach and camera following. Penguin waves to camera. Feet go by camera in foreground.

企鹅在海滩上快速行走，摄像机跟随。企鹅向摄像机挥手。前景中的摄像机拍下企鹅的双脚。

A penguin runs into a 100 colorful bouncy balls.

一只企鹅撞上了100个彩色弹力球。

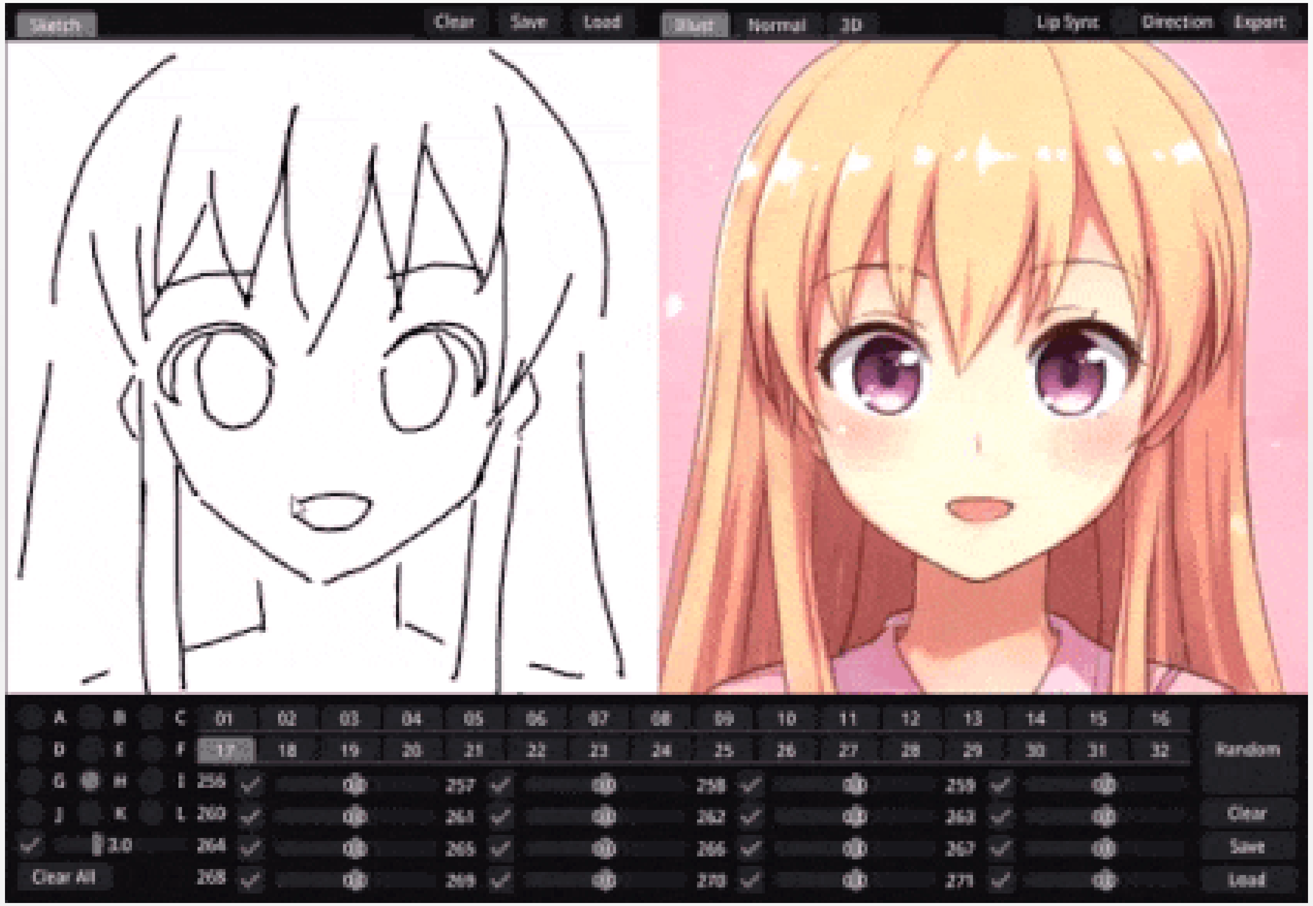
slow zoom out. penguin sitting on bird nest with a single colorful egg.

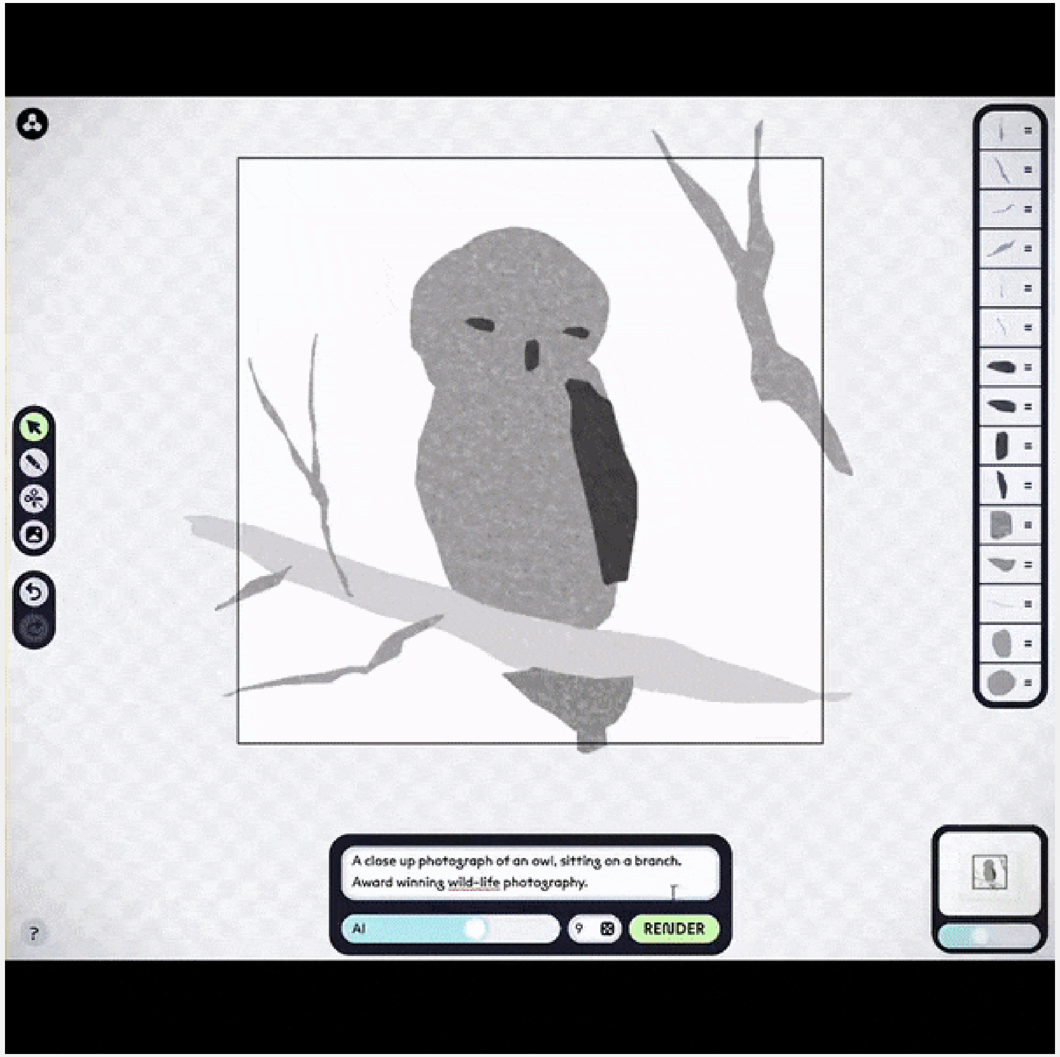
镜头慢慢拉远，企鹅带着一个彩蛋坐在鸟巢上。

zoom out. Aerial view of penguin sitting on bird nest in rainbow antarctic glacier.

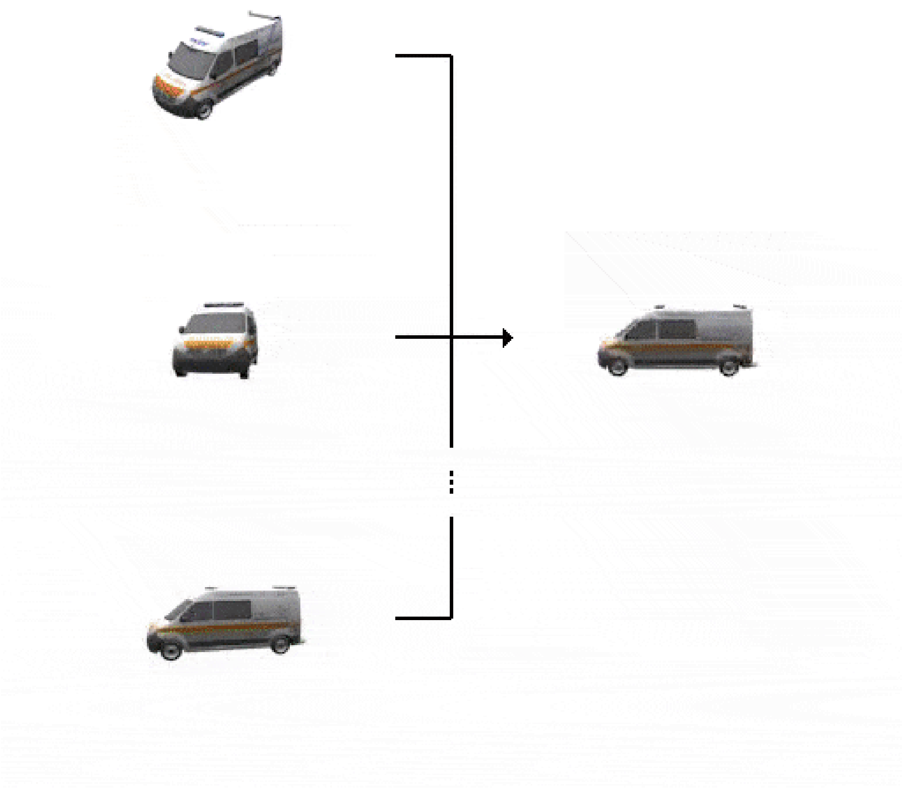
镜头拉远。彩虹照耀南极冰川、企鹅坐在鸟巢上的鸟瞰图。

Art assistant



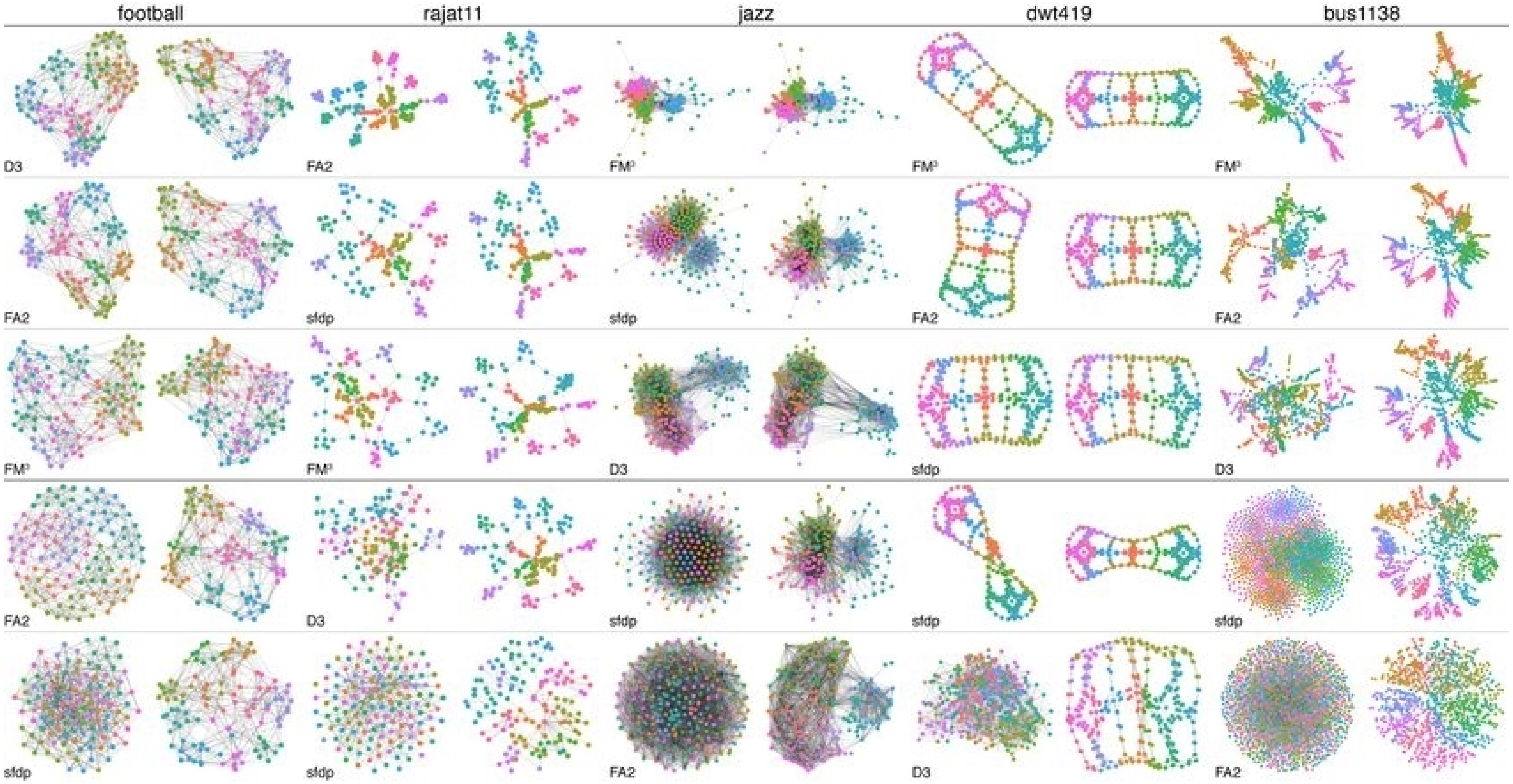


2D→3D



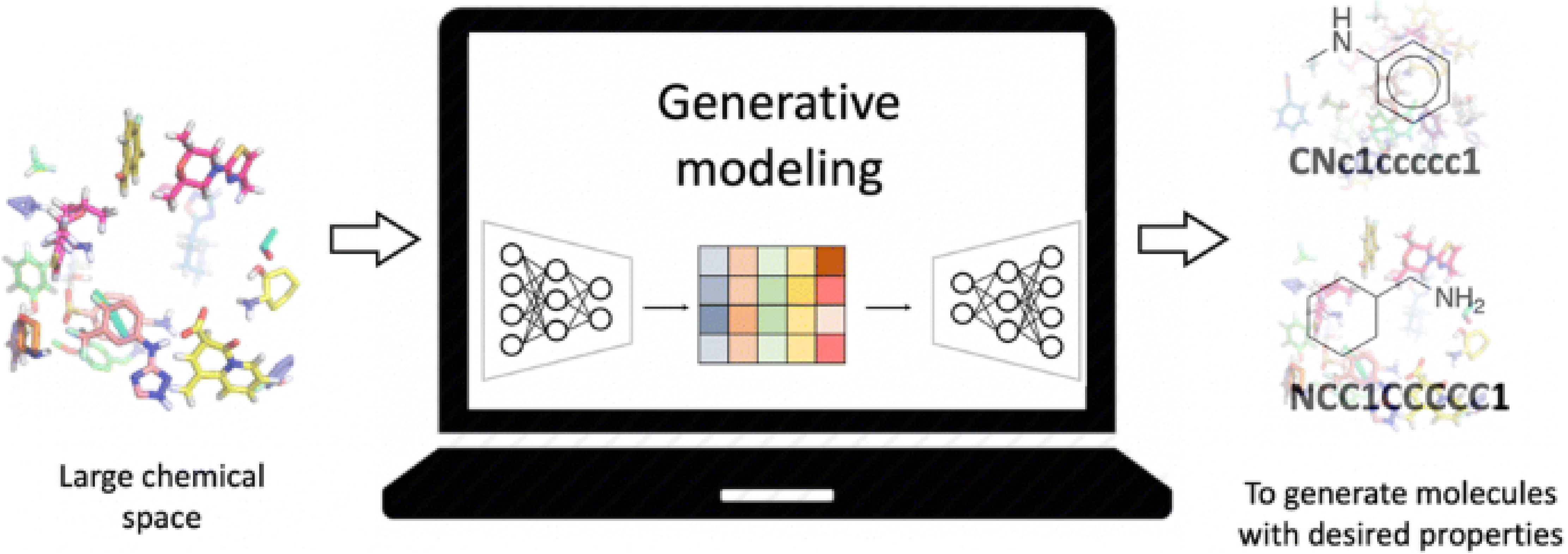
https://3d-diffusion.github.io https://dreamfusion3d.github.io/

Graph Generation



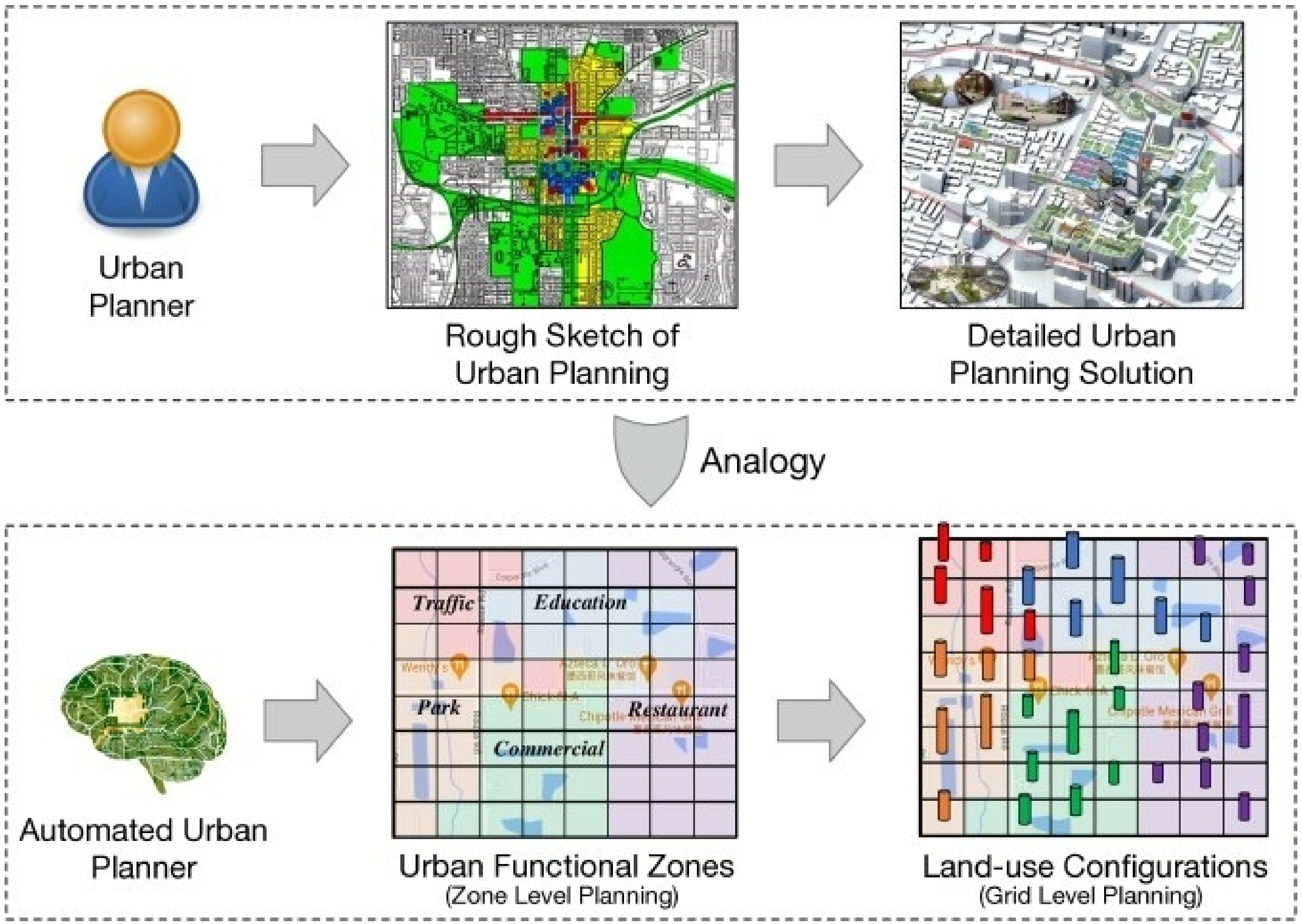
https://arxiv.org/pdf/2007.06686.pdf

Molecular Structures



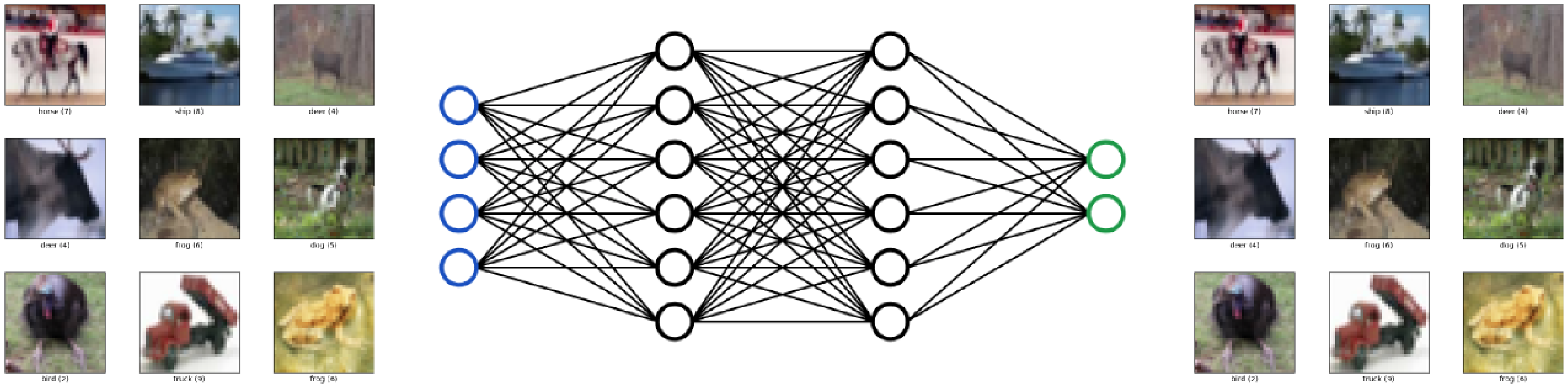
https://link.springer.com/article/10.1007/s00894-021-04674-8

City Planning



https://deepai.org/publication/human-instructed-deep-hierarchical-generative-learning-for-automated-urban-planning

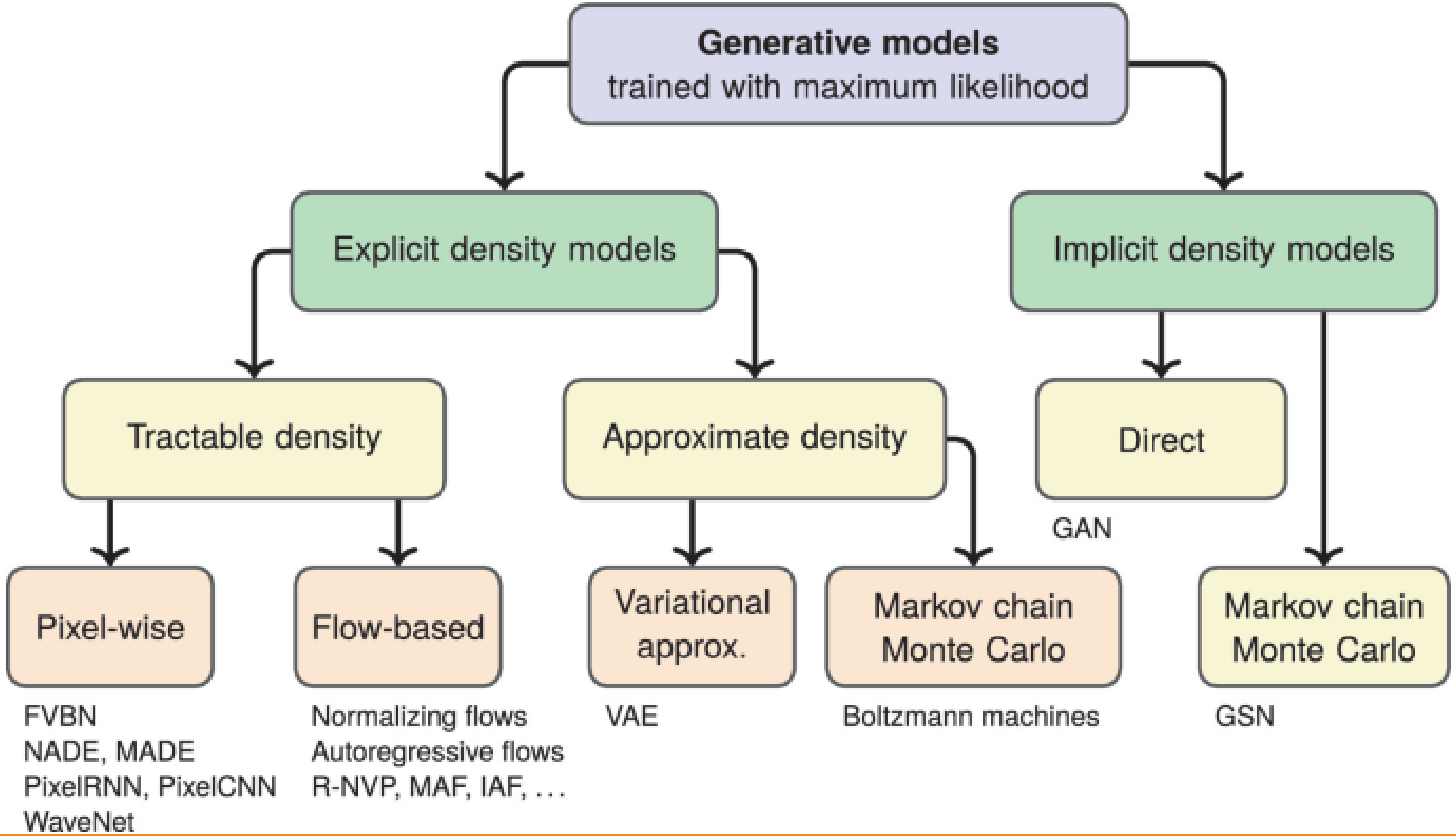
What is generative model?



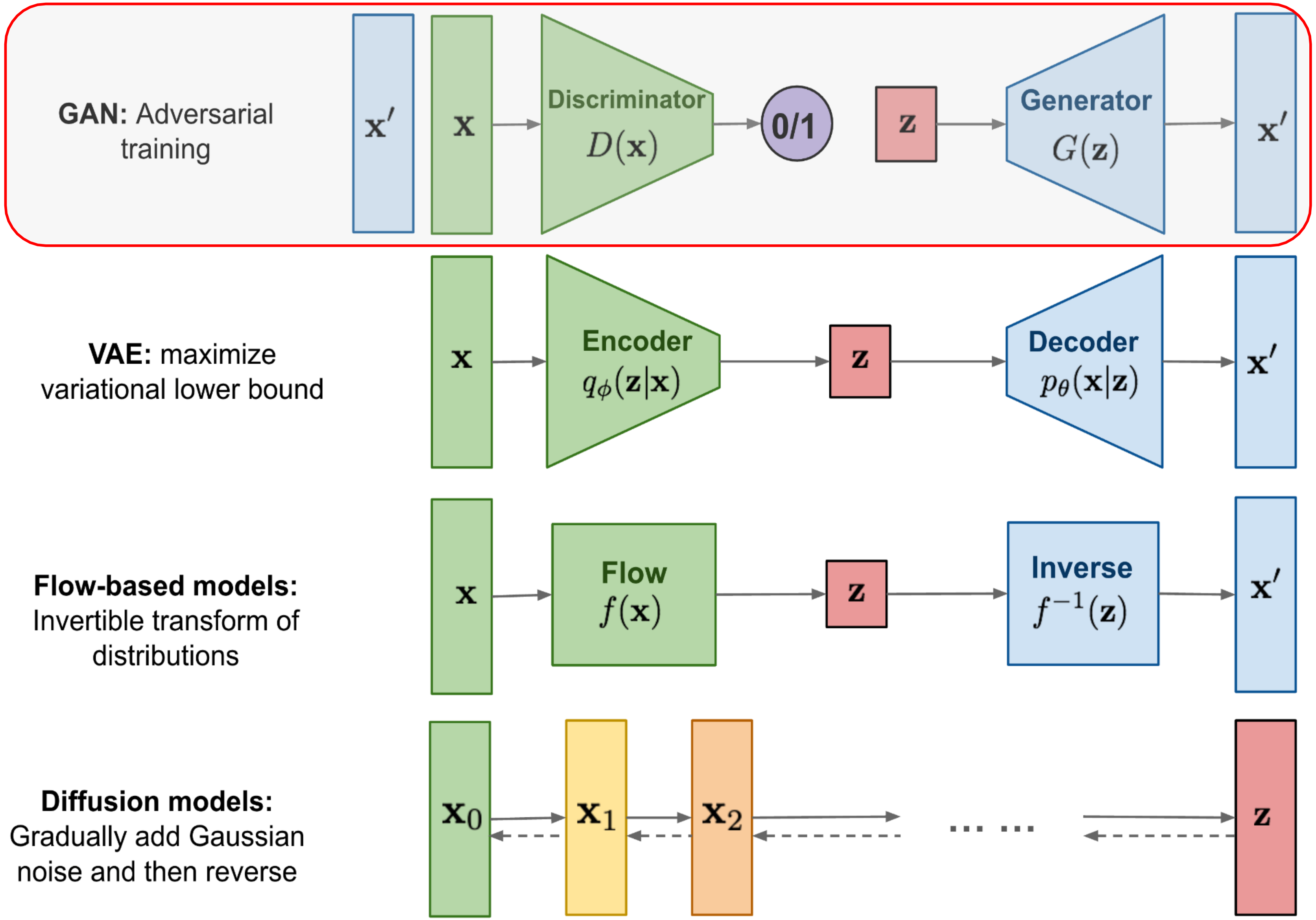
training data ~ p_data(x)   learning p_model(x)   generating new samples

1. learning p_model(x) that approximate p_data(x)
2. generating new samples x ~ p_model(x)

Generative Model Zoo

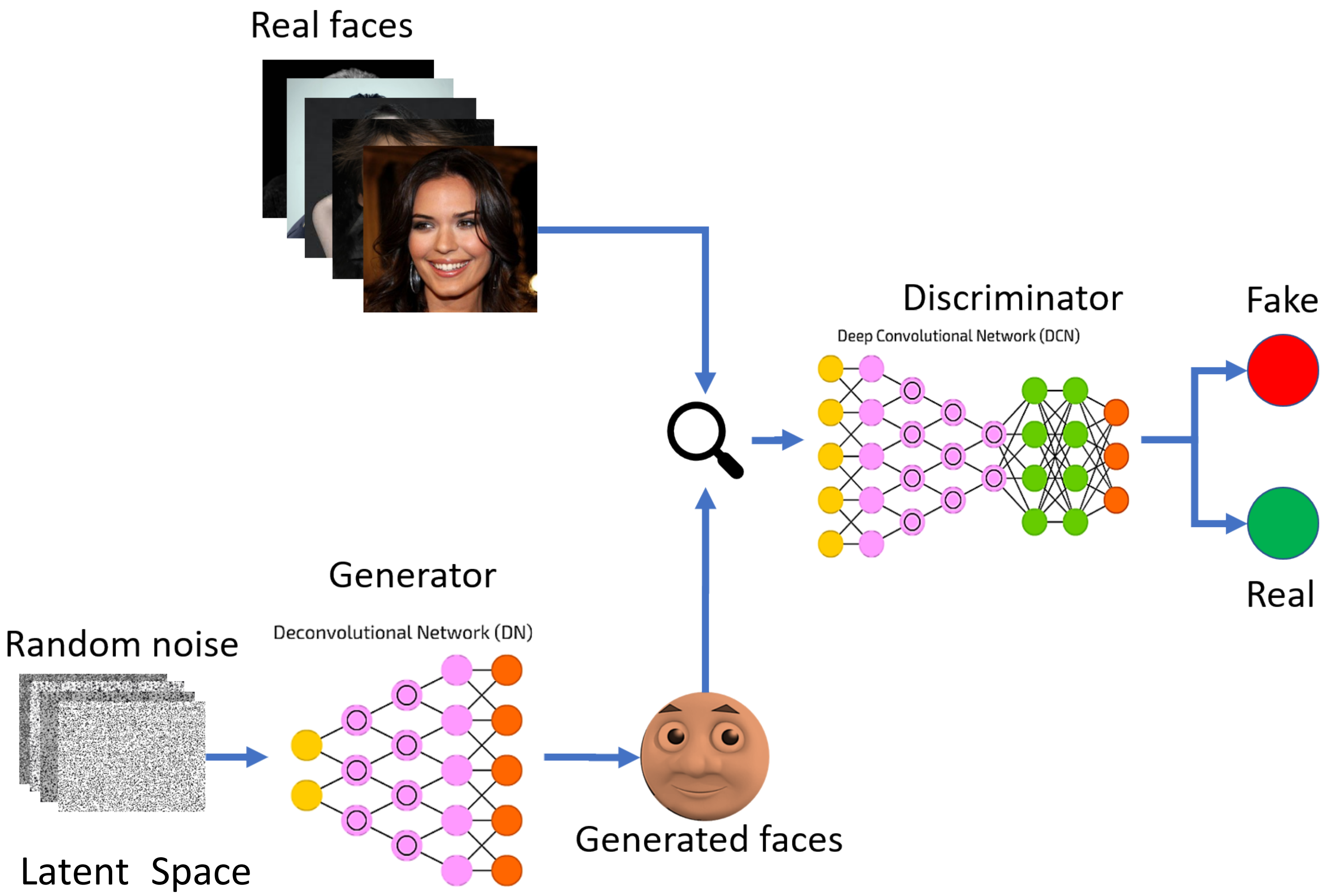


Generative Model Zoo



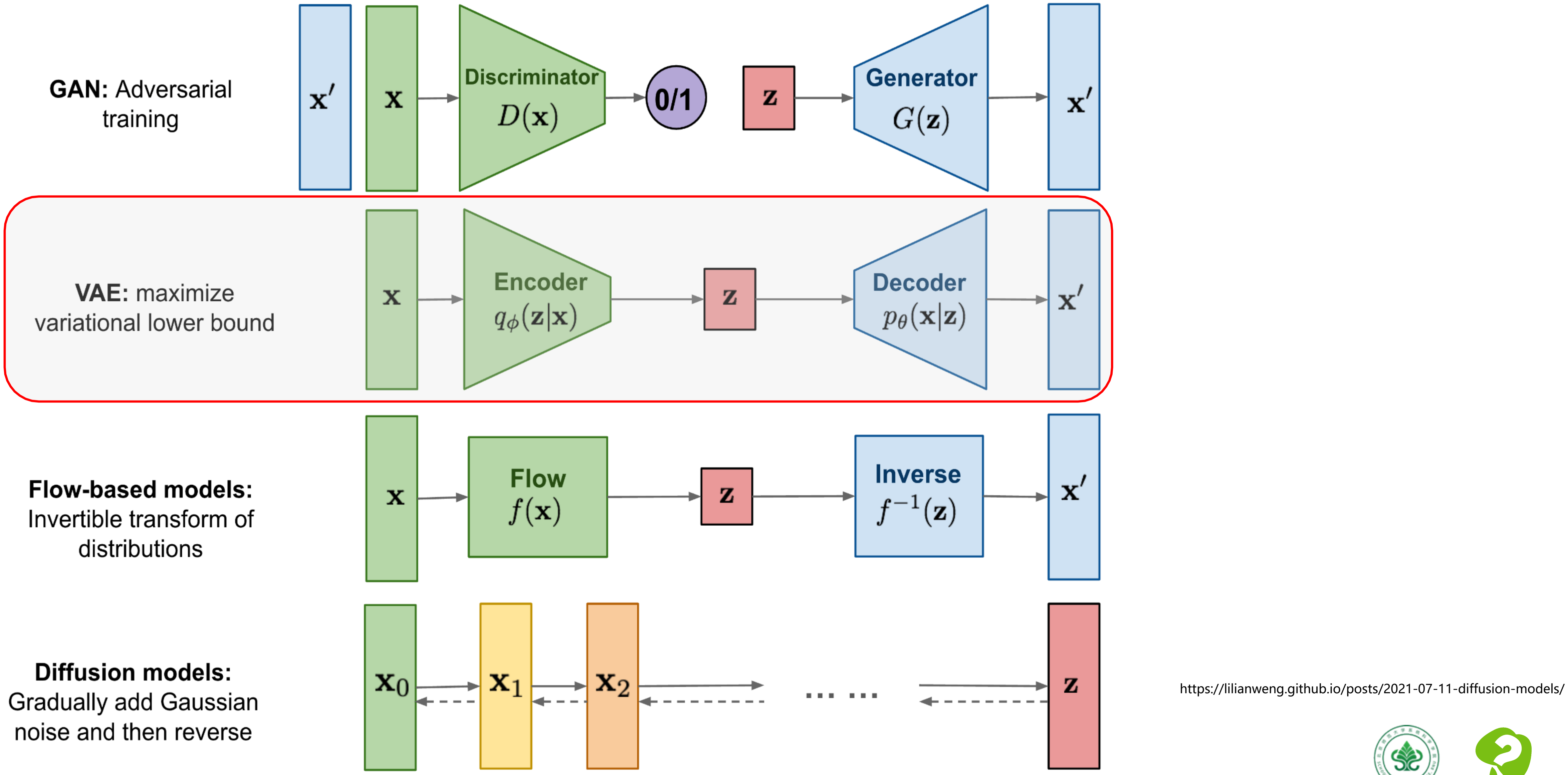
https://lilianweng.github.io/posts/2021-07-11-diffusion-models/

GAN

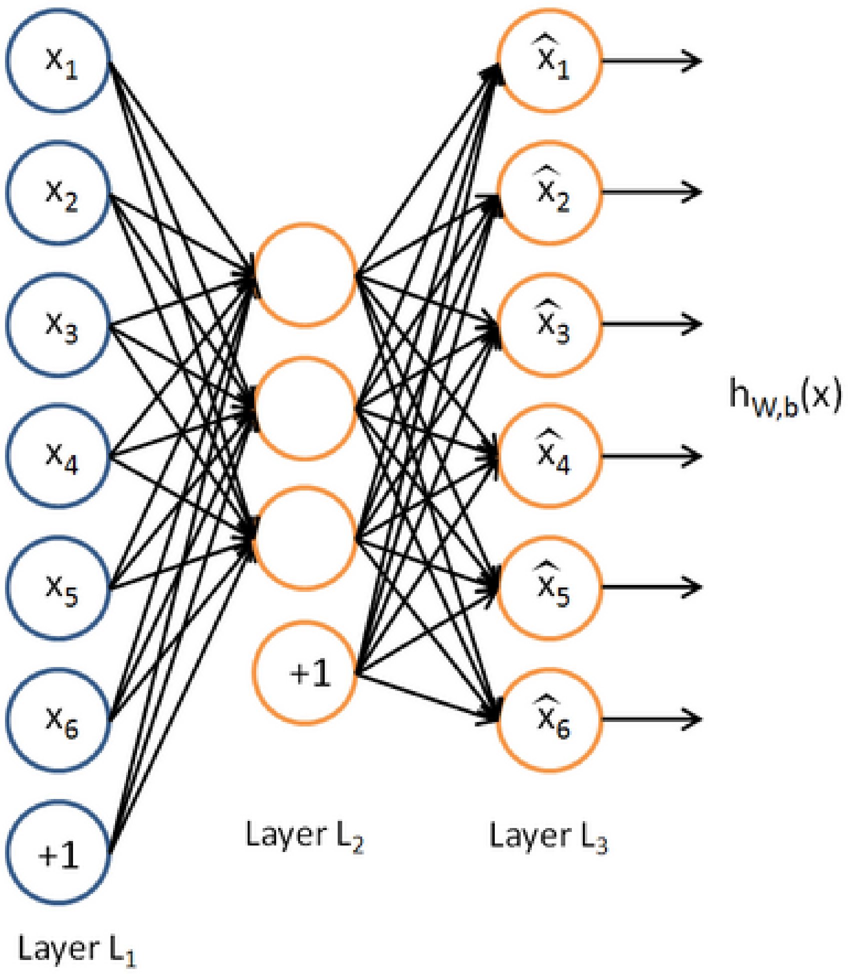


https://arxiv.org/pdf/1406.2661v1.pdf

Generative Model Zoo

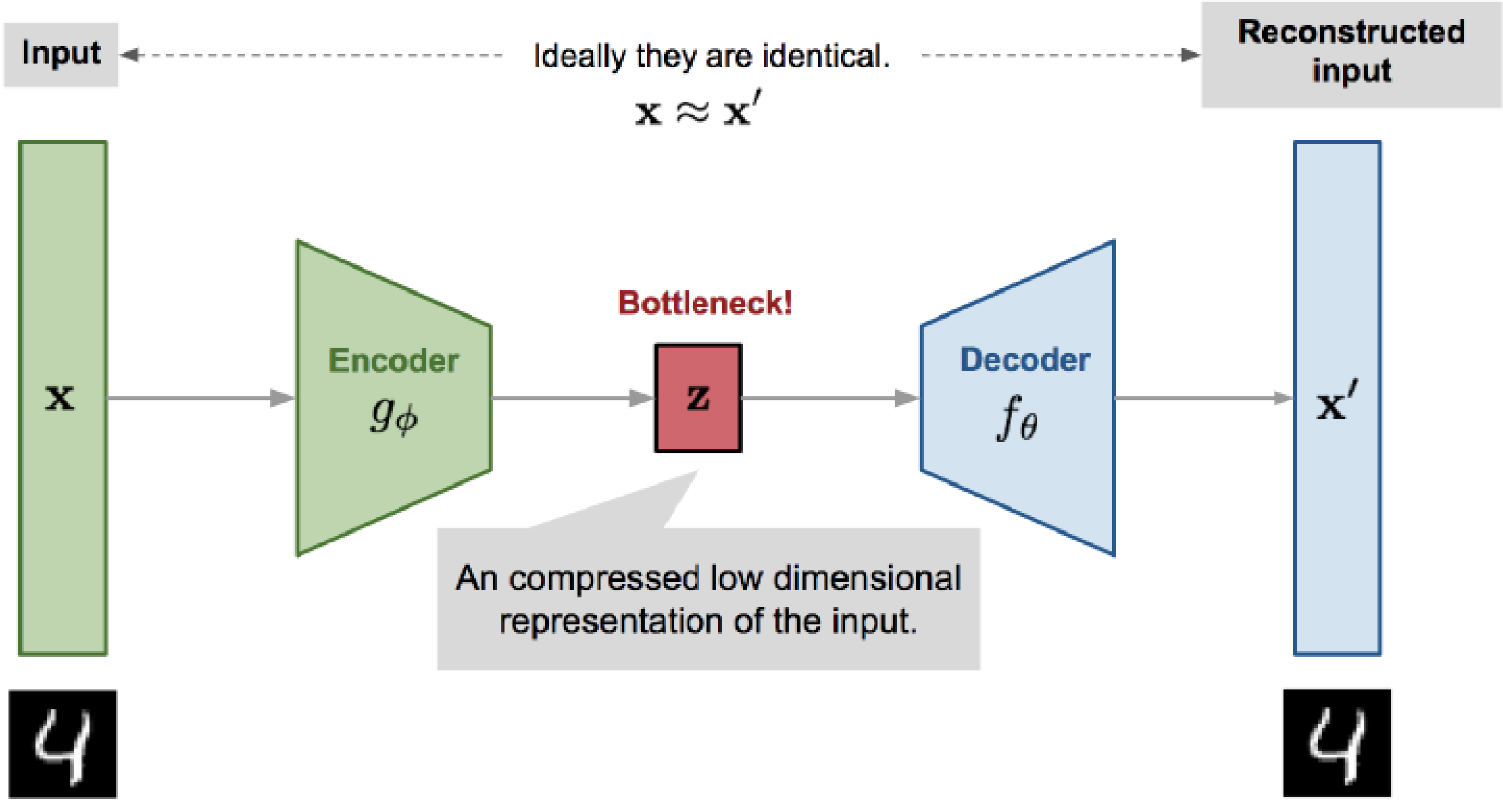


Autoencoder



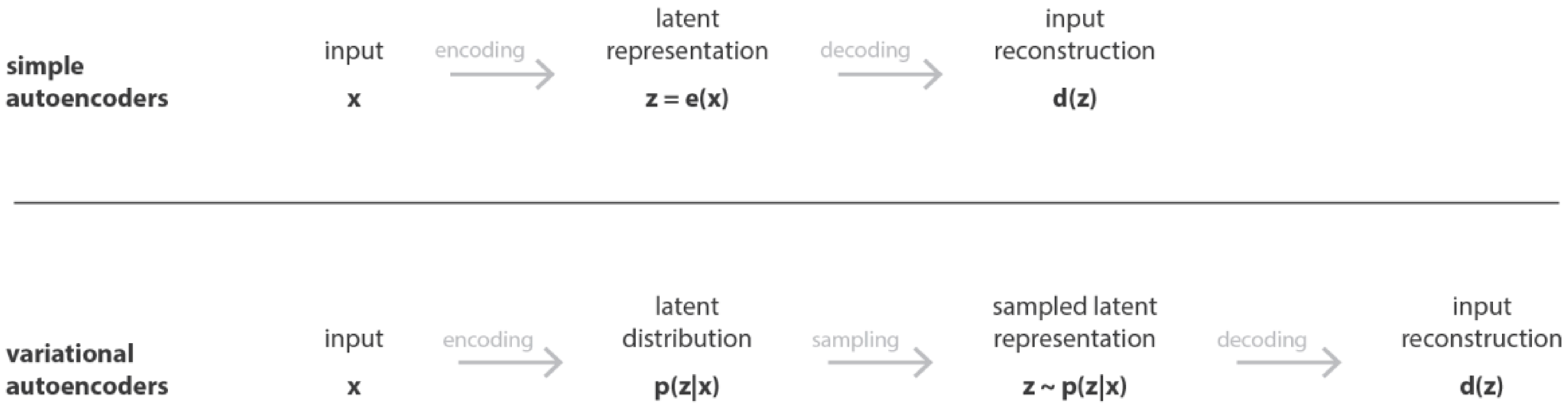
- Unsupervised learning
- Feed-back
  - Reconstruct inpu $h_\theta(x) \approx x$
  - BP Learning
- Constraints

Variational Auto Encoder

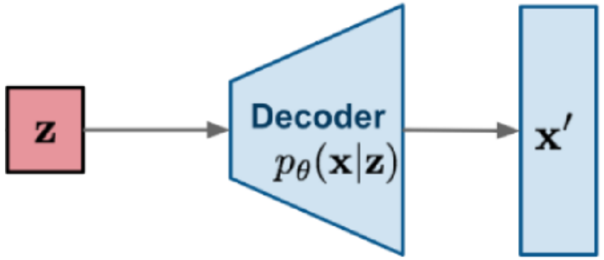


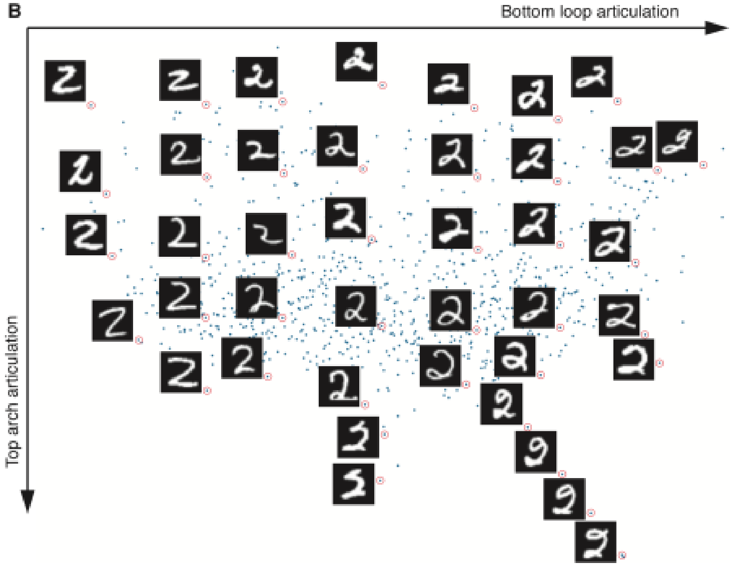
https://lilianweng.github.io/posts/2018-08-12-vae/

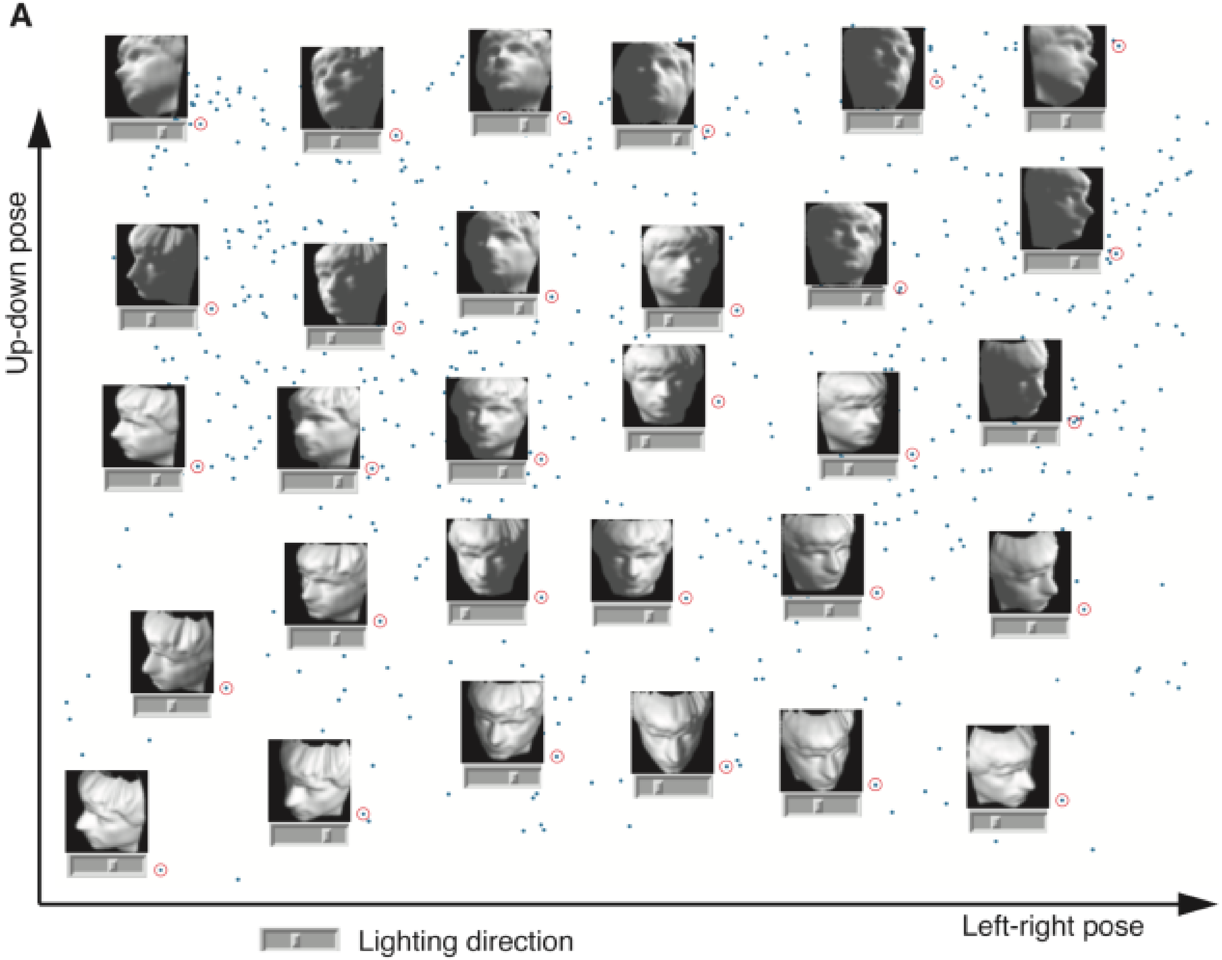
What is the Difference



Latent Variable



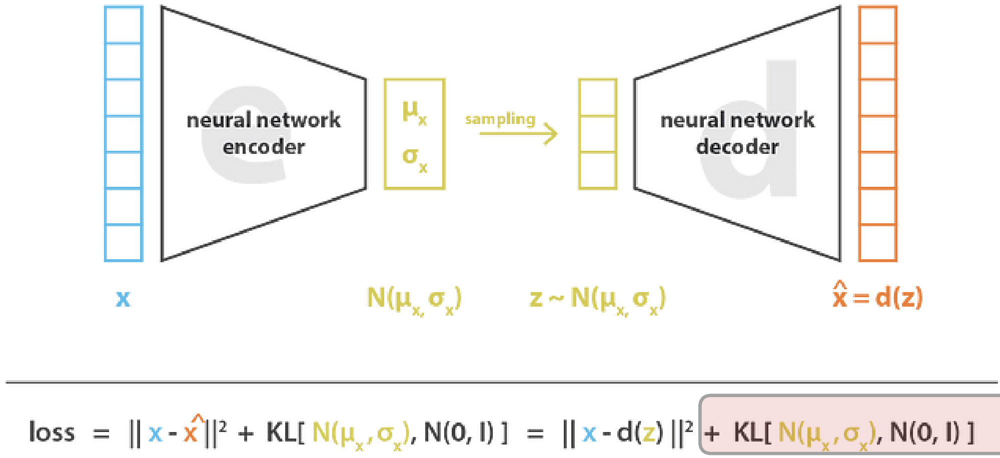




How to design loss function?

Two important requirements:

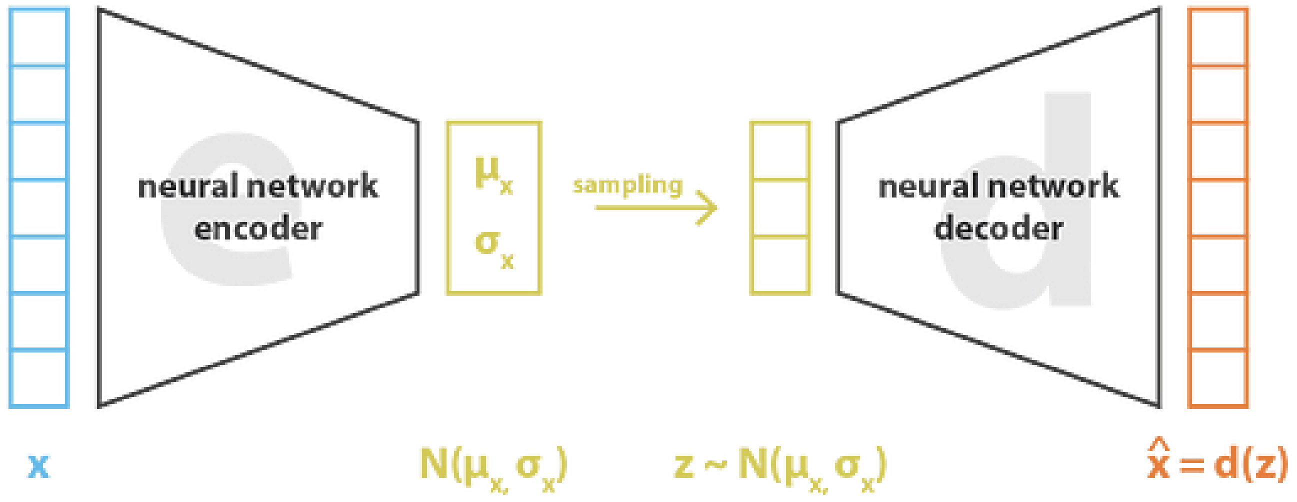
- **Continuity**
- **Completeness**



•***continuity***: two close points in the latent space should give close contents when decoded.
•***completeness***: a point sampled from the latent space should give "meaningful" content once decoded.

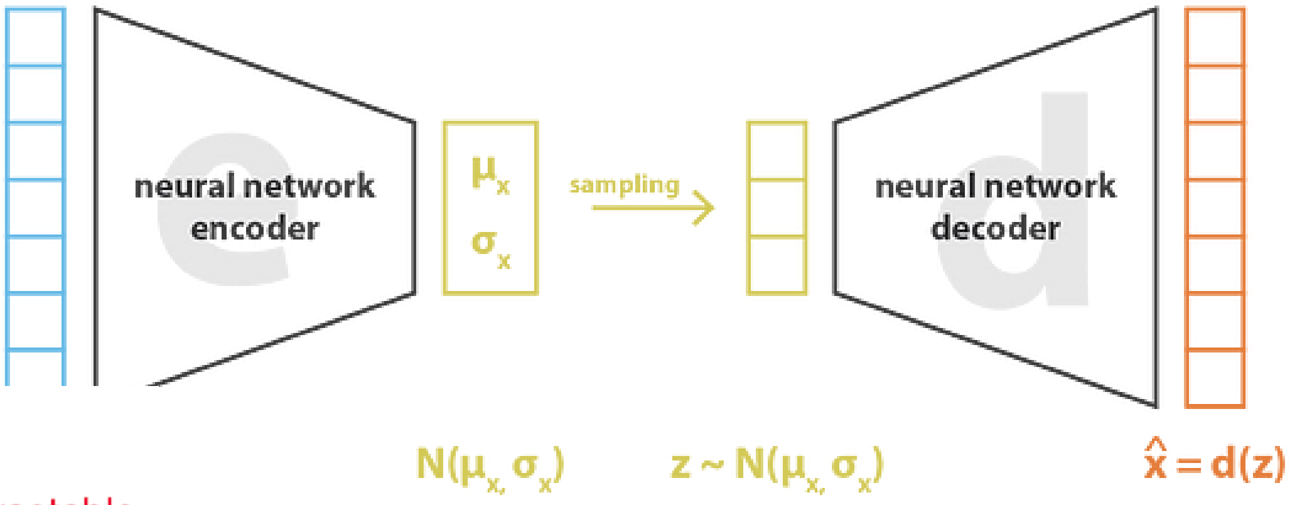
https://creatis-myriad.github.io/tutorials/2022-09-12-tutorial-vae.html

Maximum Likelihood Principle



$$\max p_\theta(x) \Longrightarrow \text{loss} = \| x - \hat{x} \|^2 + \mathrm{KL}[\, N(\mu_x, \sigma_x), N(0, I)\,] = \| x - d(z) \|^2 + \mathrm{KL}[\, N(\mu_x, \sigma_x), N(0, I)\,]$$

Maximum Likelihood Principle

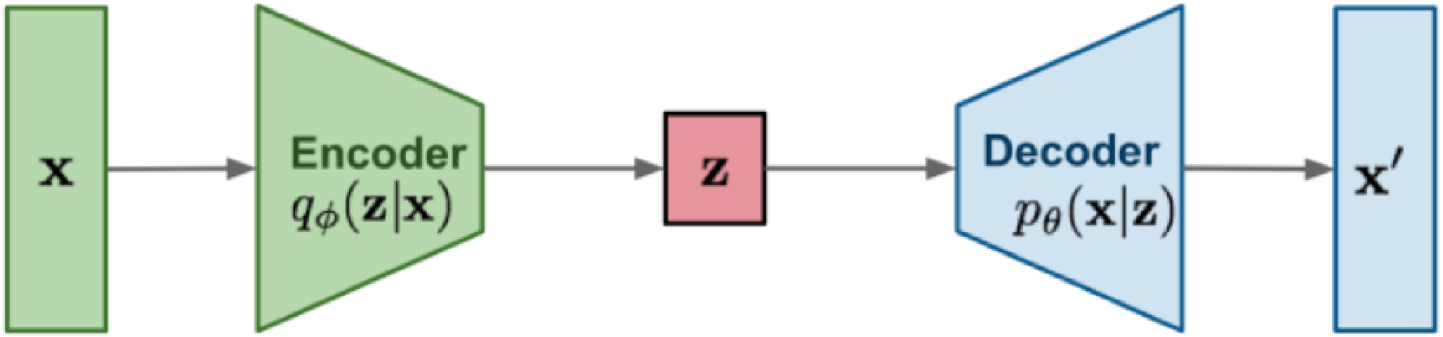


$$p_\theta(\mathbf{x}) = \boxed{\int p_\theta(\mathbf{x}\,|\,\mathbf{z})p(\mathbf{z})d\mathbf{z}}$$

intractable

$$\max p_\theta(x) \Rightarrow \text{loss} = \|x - \hat{x}\|^2 + \mathrm{KL}[N(\mu_x, \sigma_x), N(0, I)] = \|x - d(z)\|^2 + \mathrm{KL}[N(\mu_x, \sigma_x), N(0, I)]$$

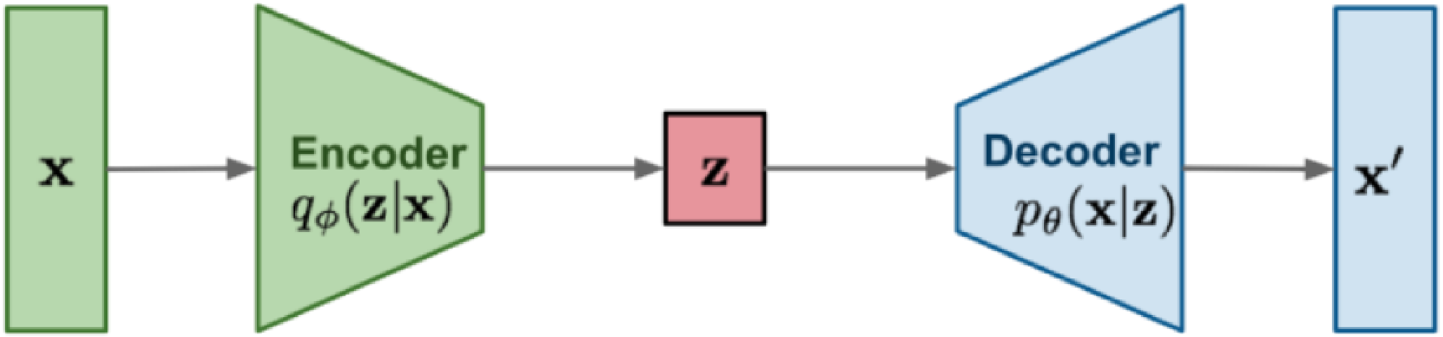
Variation



$$
\begin{aligned}
\ln p_\theta(\boldsymbol{x}) &= \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})} \ln p_\theta(\boldsymbol{x}) \\
&= \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})} \ln \left[\frac{p_\theta(\boldsymbol{x},\boldsymbol{z})}{p_\theta(\boldsymbol{z}|\boldsymbol{x})}\right] \\
&= \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})} \ln \left[\frac{p_\theta(\boldsymbol{x},\boldsymbol{z})\,\boxed{q_\phi(\boldsymbol{z}|\boldsymbol{x})}}{q_\phi(\boldsymbol{z}|\boldsymbol{x})\,p_\theta(\boldsymbol{z}|\boldsymbol{x})}\right] \\
&= \underbrace{\mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})} \ln \left[\frac{p_\theta(\boldsymbol{x},\boldsymbol{z})}{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\right]}_{\text{Evidence Lower BOund, ELBO}} + D_{\mathrm{KL}}\left(q_\phi(\boldsymbol{z}|\boldsymbol{x}) \| p_\theta(\boldsymbol{z}|\boldsymbol{x})\right)
\end{aligned}
$$

Variational
Distribution

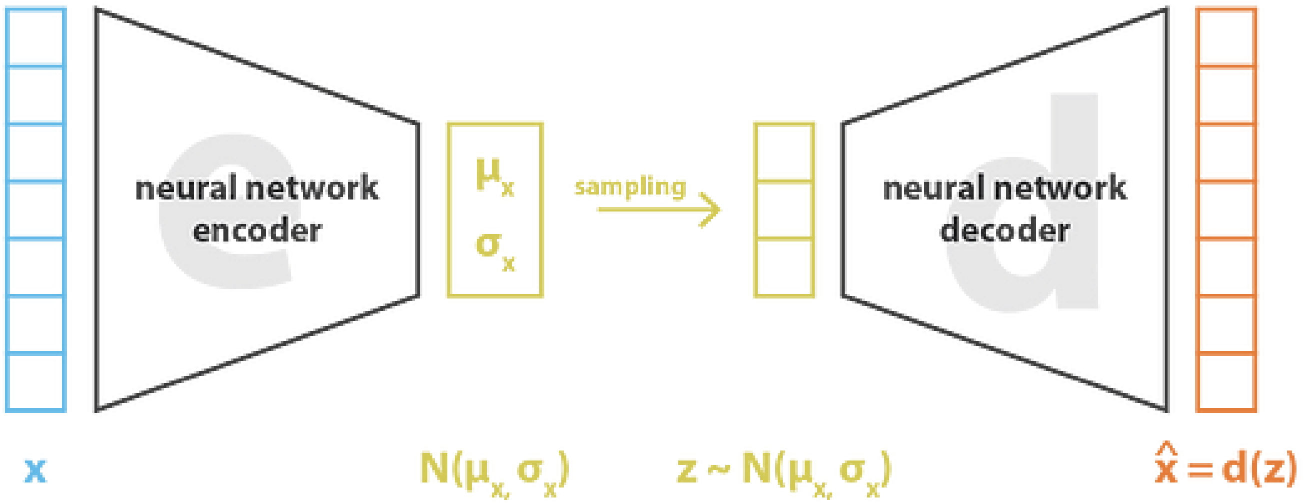
Variation



$$
\begin{aligned}
\ln p_\theta(\boldsymbol{x}) &= \mathbb{E}_{q_\phi(z|x)} \ln p_\theta(\boldsymbol{x}) \\
&= \mathbb{E}_{q_\phi(z|x)} \ln \left[\frac{p_\theta(\boldsymbol{x},\boldsymbol{z})}{p_\theta(\boldsymbol{z}|\boldsymbol{x})}\right] \\
&= \mathbb{E}_{q_\phi(z|x)} \ln \left[\frac{p_\theta(\boldsymbol{x},\boldsymbol{z})}{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\frac{q_\phi(\boldsymbol{z}|\boldsymbol{x})}{p_\theta(\boldsymbol{z}|\boldsymbol{x})}\right] \\
&= \underbrace{\mathbb{E}_{q_\phi(z|x)} \ln \left[\frac{p_\theta(\boldsymbol{x},\boldsymbol{z})}{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\right]}_{\text{Evidence Lower BOund, ELBO}} + D_{\mathrm{KL}}\left(q_\phi(\boldsymbol{z}|\boldsymbol{x}) \| p_\theta(\boldsymbol{z}|\boldsymbol{x})\right)
\end{aligned}
$$

$$
\begin{aligned}
\ln p_\theta(\boldsymbol{x}) &\geq \mathbb{E}_{q_\phi(z|x)} \ln \left[\frac{p_\theta(\boldsymbol{x},\boldsymbol{z})}{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\right] \\
&\geq \mathbb{E}_{q_\phi(z|x)} \ln \left[\frac{p_\theta(\boldsymbol{x}|\boldsymbol{z})p(\boldsymbol{z})}{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\right] \\
&\geq \underbrace{\mathbb{E}_{q_\phi(z|x)} \ln p_\theta(\boldsymbol{x}|\boldsymbol{z})}_{\text{reconstruction error}} - \underbrace{D_{\mathrm{KL}}\left(q_\phi(\boldsymbol{z}|\boldsymbol{x}) \| p(\boldsymbol{z})\right)}_{\text{regularizer}}
\end{aligned}
$$

How to design loss function?



$$p_\theta(x|z) = \frac{1}{Z}\exp\frac{(x-\hat{x})^2}{2\sigma^2}$$

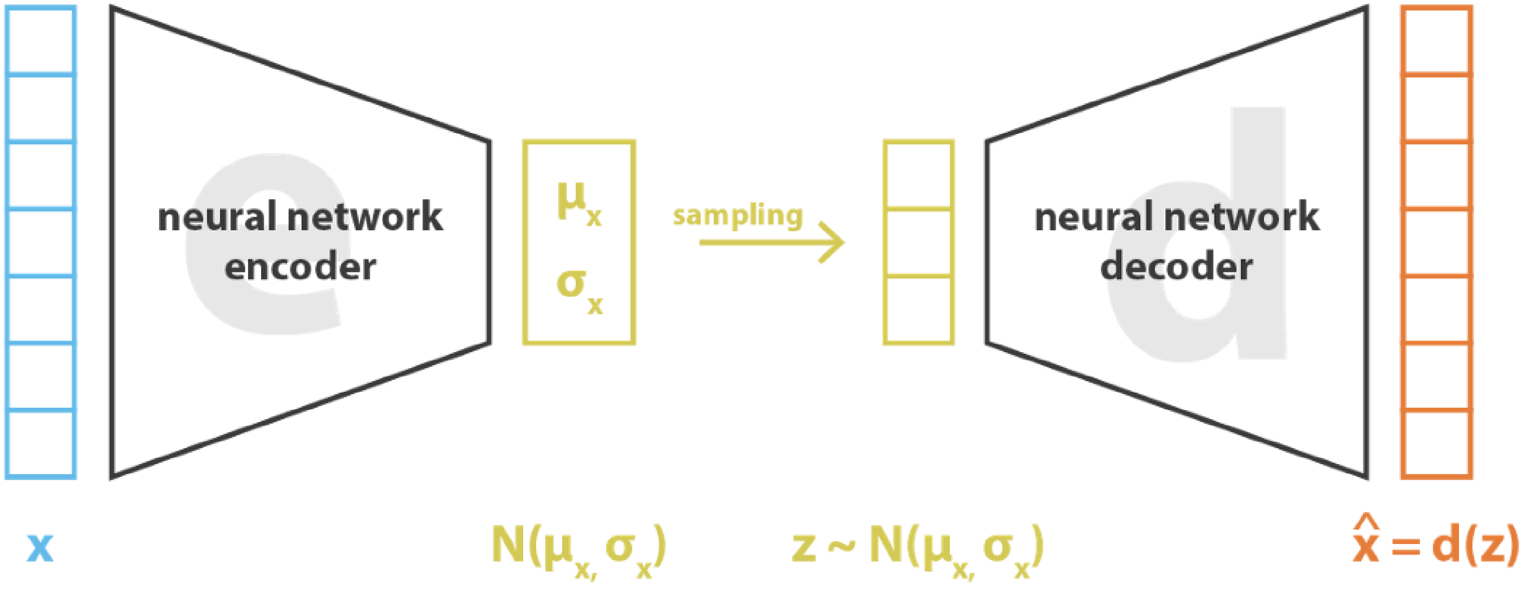
$$p_\theta(x|z) = \frac{1}{Z}\exp\frac{(x-\hat{x})^2}{2\sigma^2} \qquad \sigma^2 = 1/2$$

$$q_\phi(z|x) = \frac{1}{Z}\exp\frac{(z-\mu_X)^2}{2\sigma_X{}^2}$$

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

ELBO

Latent Variable as Gaussian

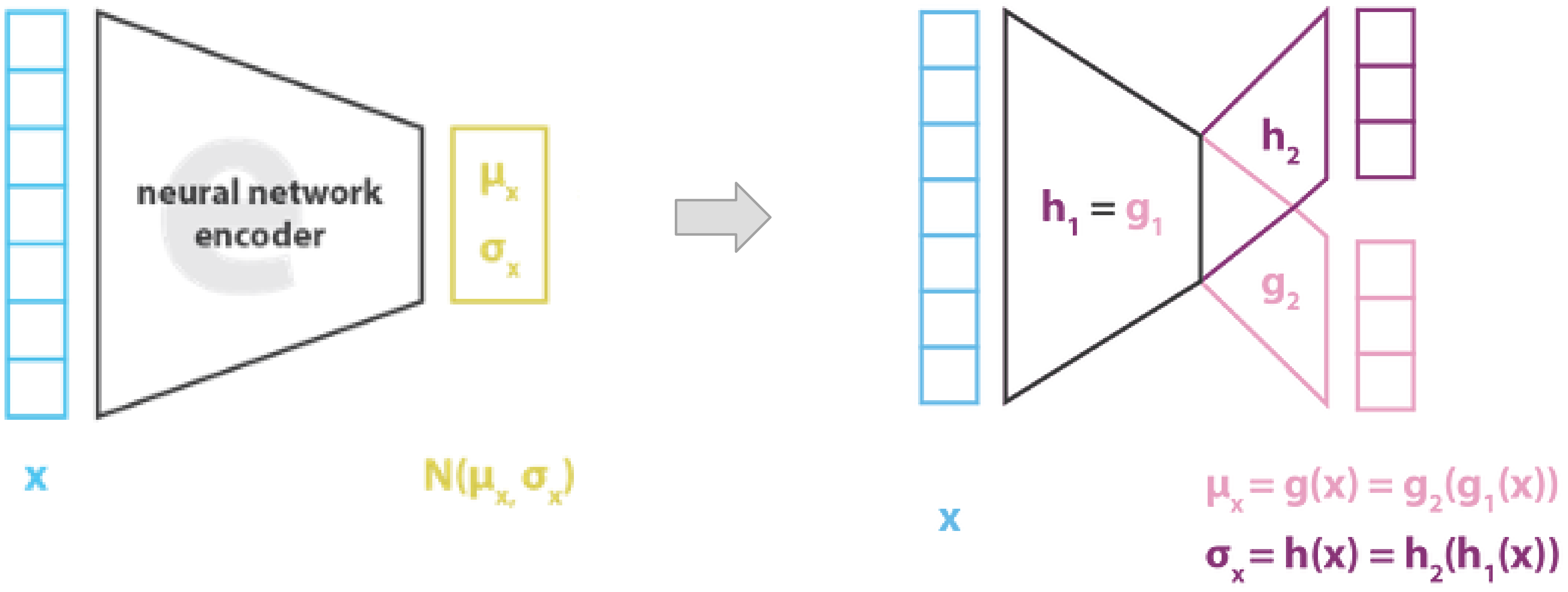


$$p_\theta(x|z) = \frac{1}{Z}\exp\frac{(x-\hat{x})^2}{2\sigma^2} \qquad \sigma^2 = 1/2$$

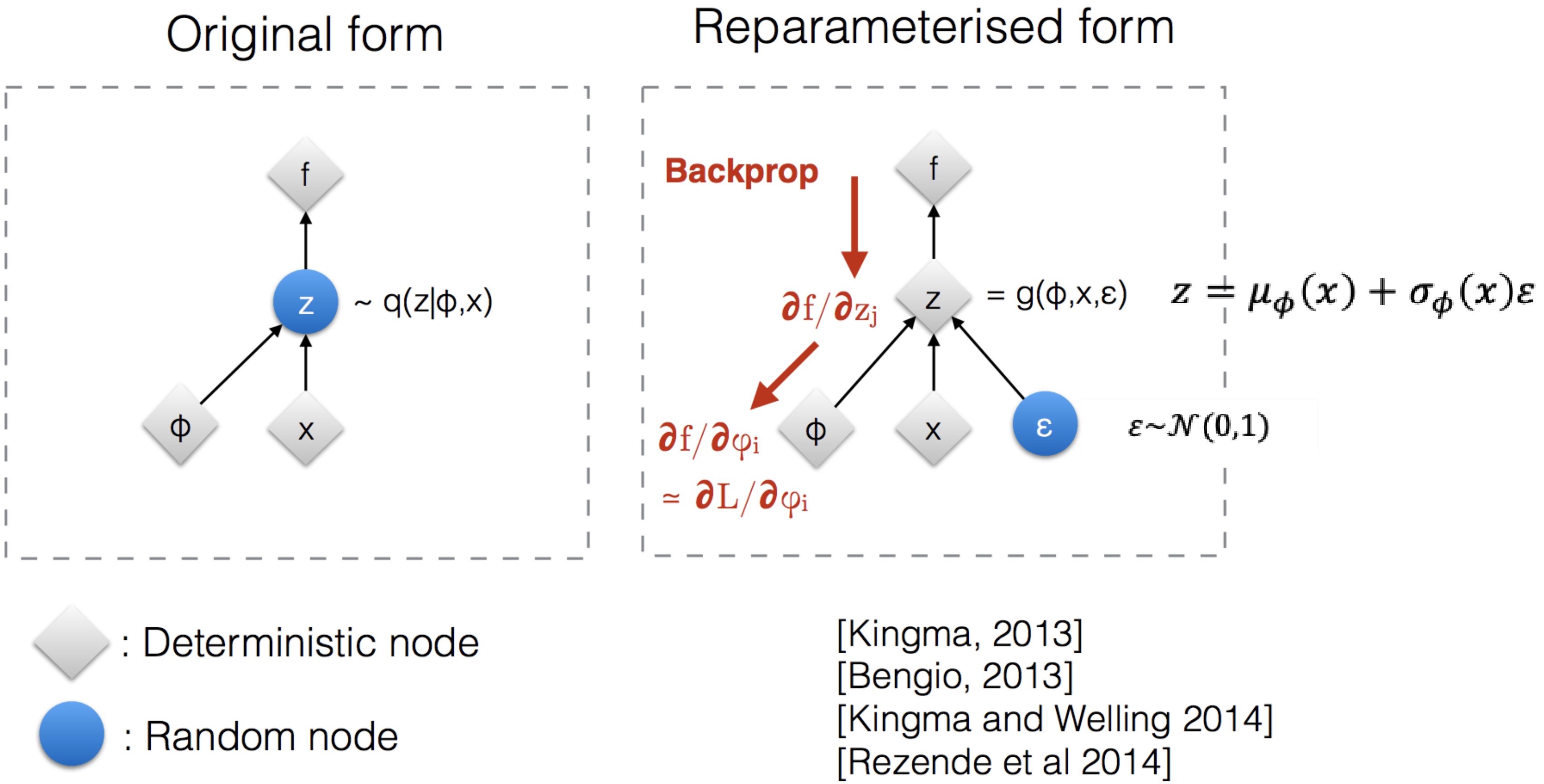
$$q_\phi(z|x) = \frac{1}{Z}\exp\frac{(z-\mu_X)^2}{2{\sigma_X}^2}$$

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

How to generate random noise?



Reparameter Technique



[Kingma, 2013]
[Bengio, 2013]
[Kingma and Welling 2014]
[Rezende et al 2014]