GENERATIVE MODELLING

Param Hanji • Nov 2023

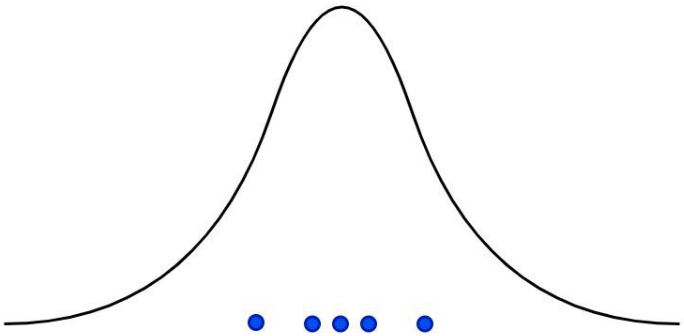
UNIVERSITY OF CAMBRIDGE

Outline

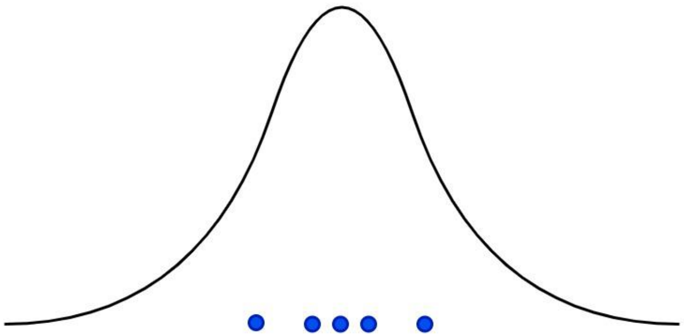
- Basic operations
- Problems relevant to machine perception
- Commonly used models
  - VAE
  - GAN
  - Normalizing flows
  - Diffusion models
- Evaluation

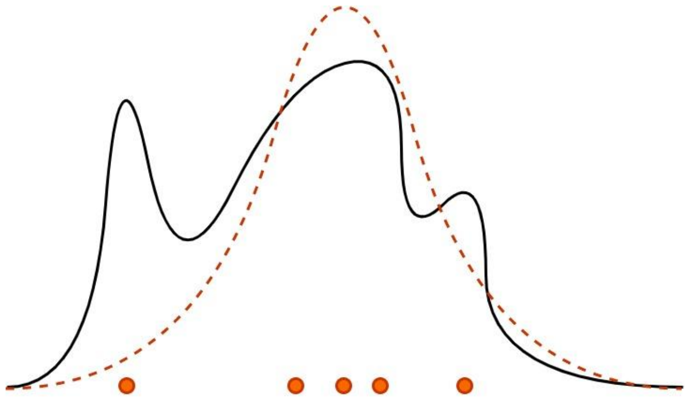
Images as samples

Operation 1: Sample generation



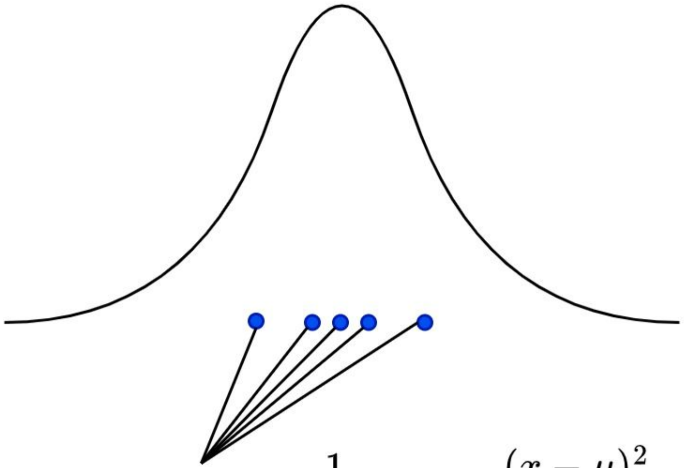
Operation 1: Sample generation



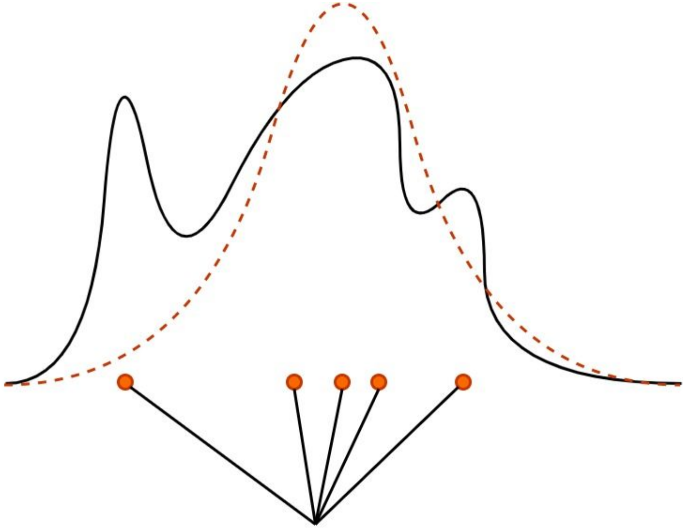


UNIVERSITY OF CAMBRIDGE

Operation 2: Density estimation



$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp - \frac{(x-\mu)^2}{2\sigma^2}$$



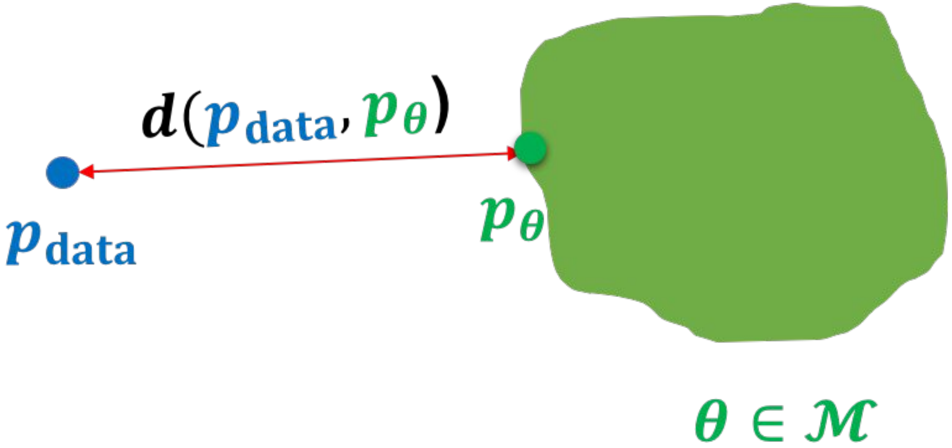
$$p(x) = ?$$

$$q_\theta(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp - \frac{(x-\mu)^2}{2\sigma^2}$$

UNIVERSITY OF CAMBRIDGE

Training

$$\mathbf{x}^{(j)} \sim p_{\text{data}}$$
$$j = 1, 2, \ldots, |\mathcal{D}|$$

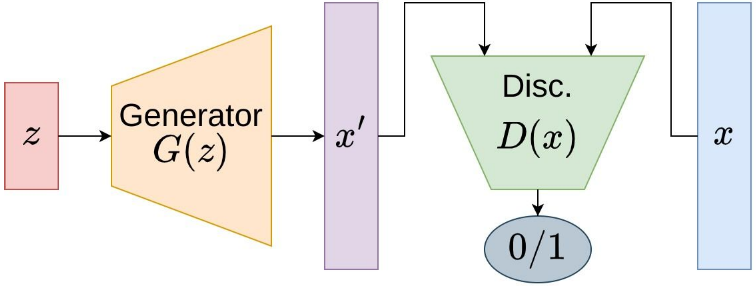


Model family

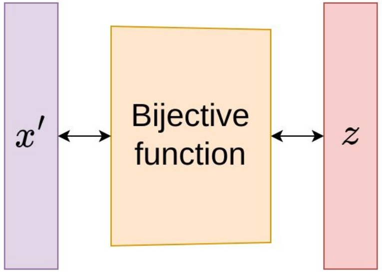
Domains

- Computer vision
- Computer graphics
- Text generation
- Medical imaging
- Audio synthesis
- Astrophysics

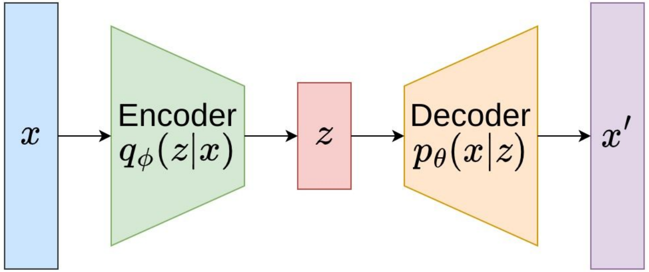
Deep generative modelling



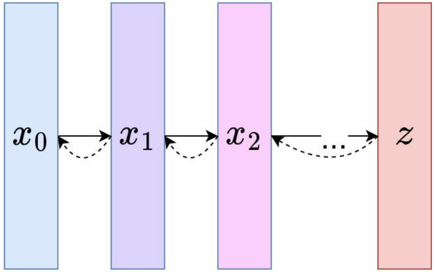
Generative adversarial network



Normalizing flow

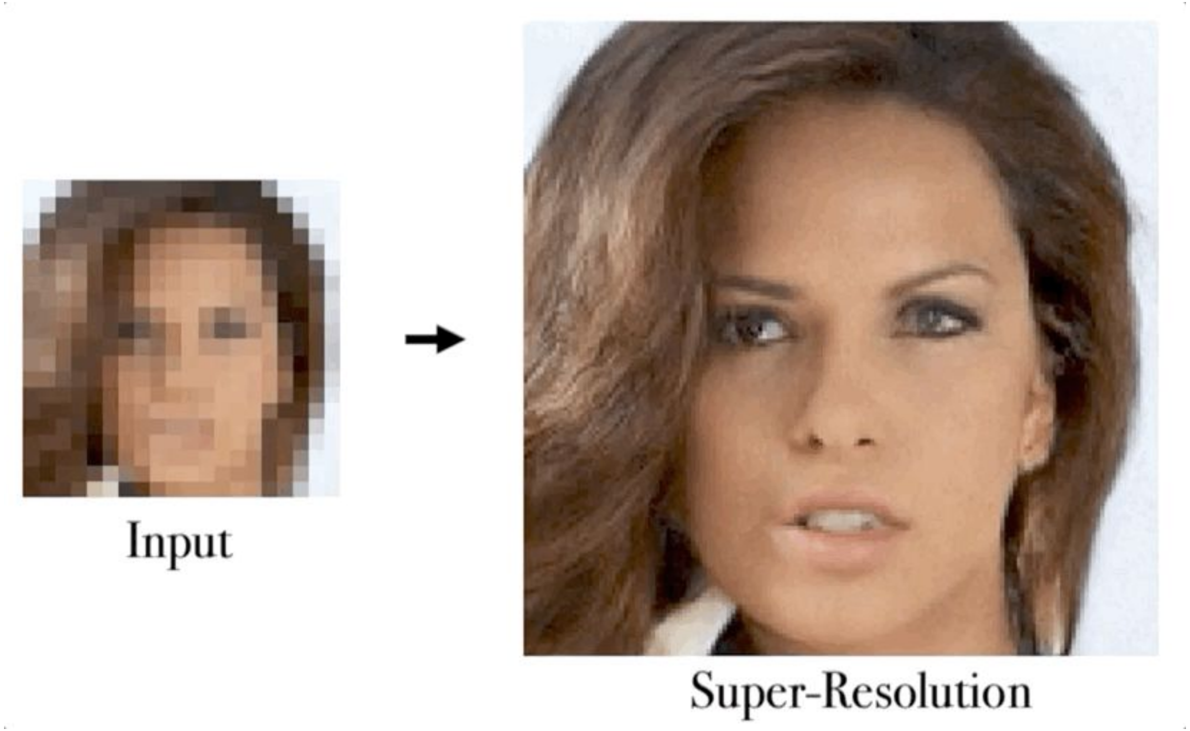


Variational autoencoder



Diffusion method

Inverse problems

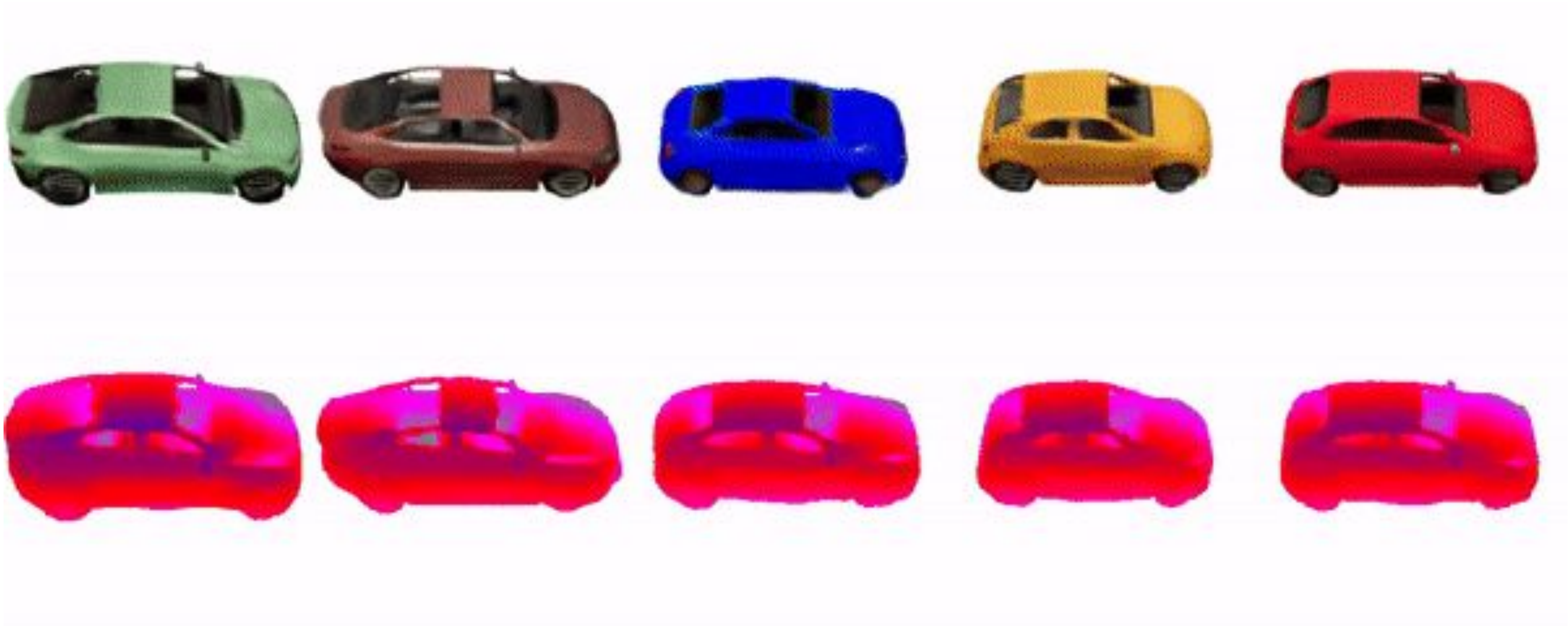


Lugmayr, Andreas, et al. "Srflow: Learning the super-resolution space with normalizing flow." *European conference on computer vision*. Springer, Cham, 2020.

Inverse imaging

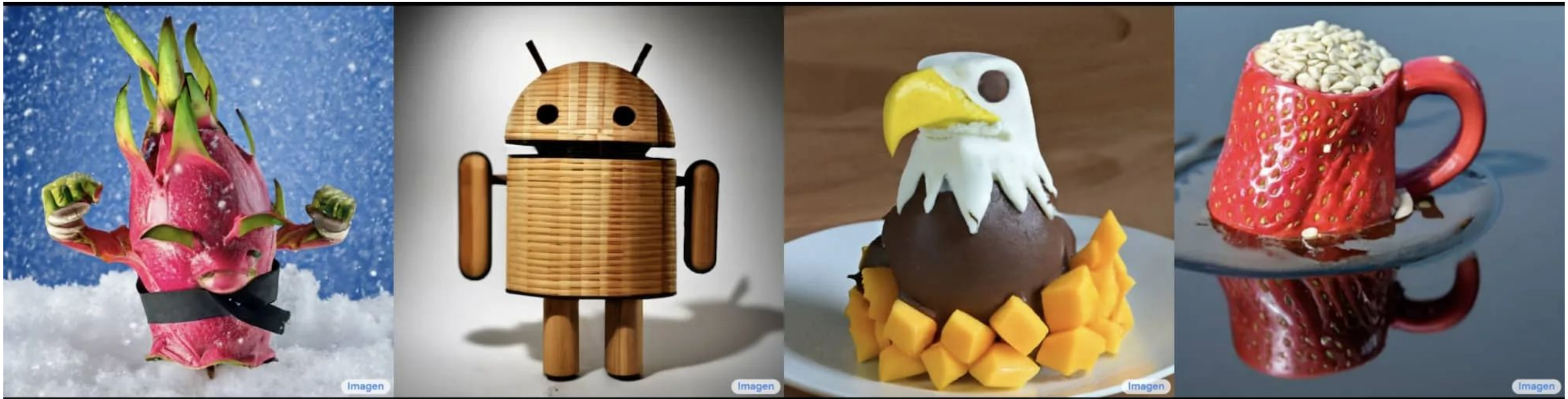
- Superresolution
- Denoising
- Colorization
- Inpainting
- Uncropping
- Debluring
- Single-image HDR

Inverse graphics



Schwarz, Katja, et al. "Graf: Generative radiance fields for 3d-aware image synthesis." *Advances in Neural Information Processing Systems* 33 (2020): 20154-20166.

Text-conditional generation

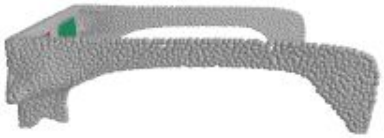


A dragon fruit wearing karate belt in the snow. | Android Mascot made from bamboo. | A bald eagle made of chocolate powder, mango, and whipped cream. | A strawberry mug filled with white sesame seeds. The mug is floating in a dark chocolate sea.

UNIVERSITY OF CAMBRIDGE

Saharia, Chitwan, et al. "Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding." *arXiv preprint arXiv:2205.11487* (2022).

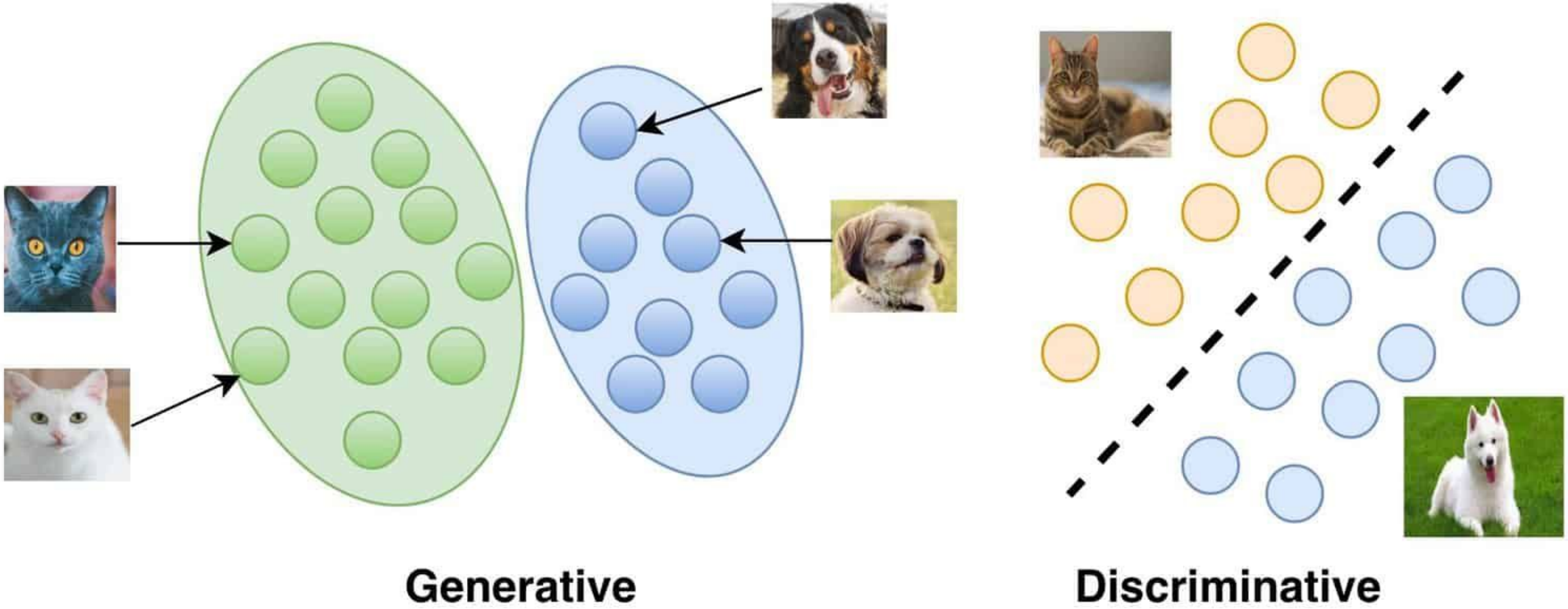
Text-conditional 3D



UNIVERSITY OF CAMBRIDGE

Nichol, Alex, et al. "Point-e: A system for generating 3d point clouds from complex prompts." *arXiv preprint arXiv:2212.08751* (2022).

Classification



Mackowiak, Radek, et al. "Generative classifiers as a basis for trustworthy image classification." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.

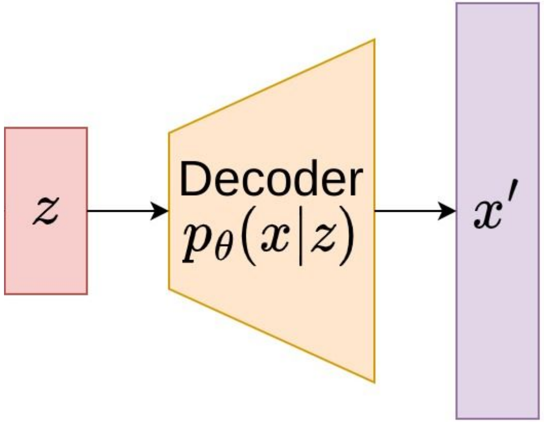
Requirements

- Handle high dimensional data
- Fast, efficient sampling
- High sample quality
- Diverse samples
- [Optional] Density evaluation
- Low dimensional latent

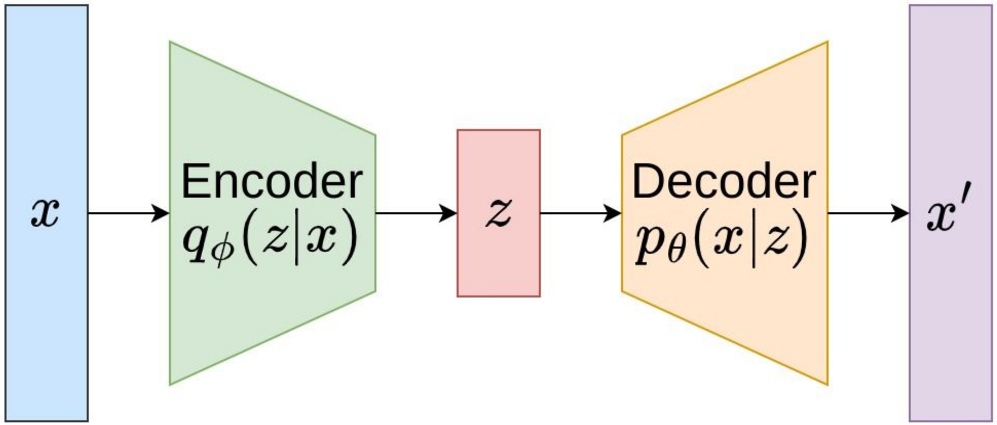
Variational autoencoder

$$\theta^* = \arg\max_\theta \sum_{i=1}^{n} \log p_\theta(\mathbf{x}^{(i)})$$

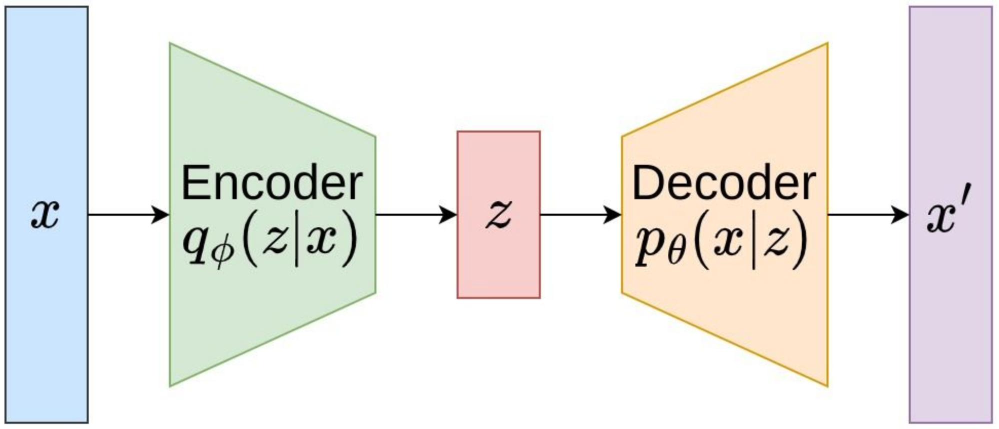
$$p_\theta(\mathbf{x}^{(i)}) = \int p_\theta(\mathbf{x}^{(i)}|\mathbf{z})p_\theta(\mathbf{z})d\mathbf{z}$$



Variational autoencoder



Variational autoencoder

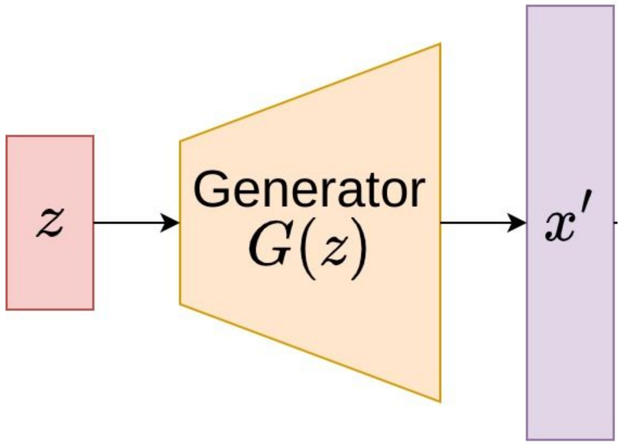


$$-L_{\text{VAE}} = \log p_\theta(\mathbf{x}) - D_{\text{KL}}(q_\phi(\mathbf{z}|\mathbf{x}) \| p_\theta(\mathbf{z}|\mathbf{x})) \leq \log p_\theta(\mathbf{x})$$

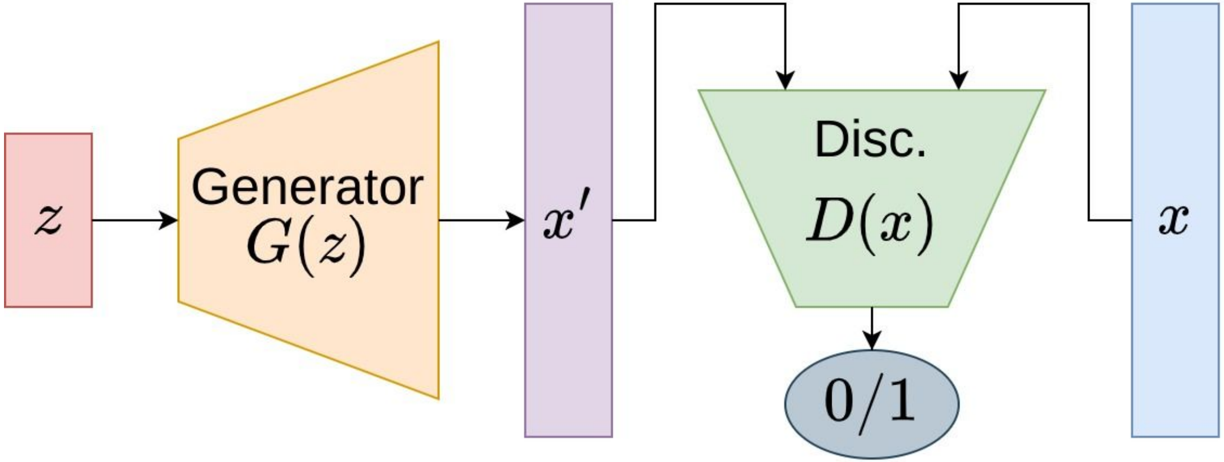
UNIVERSITY OF CAMBRIDGE

Kingma, Diederik P., and Max Welling. "Auto-encoding variational bayes." *arXiv preprint arXiv:1312.6114* (2013).

Generative adversarial network



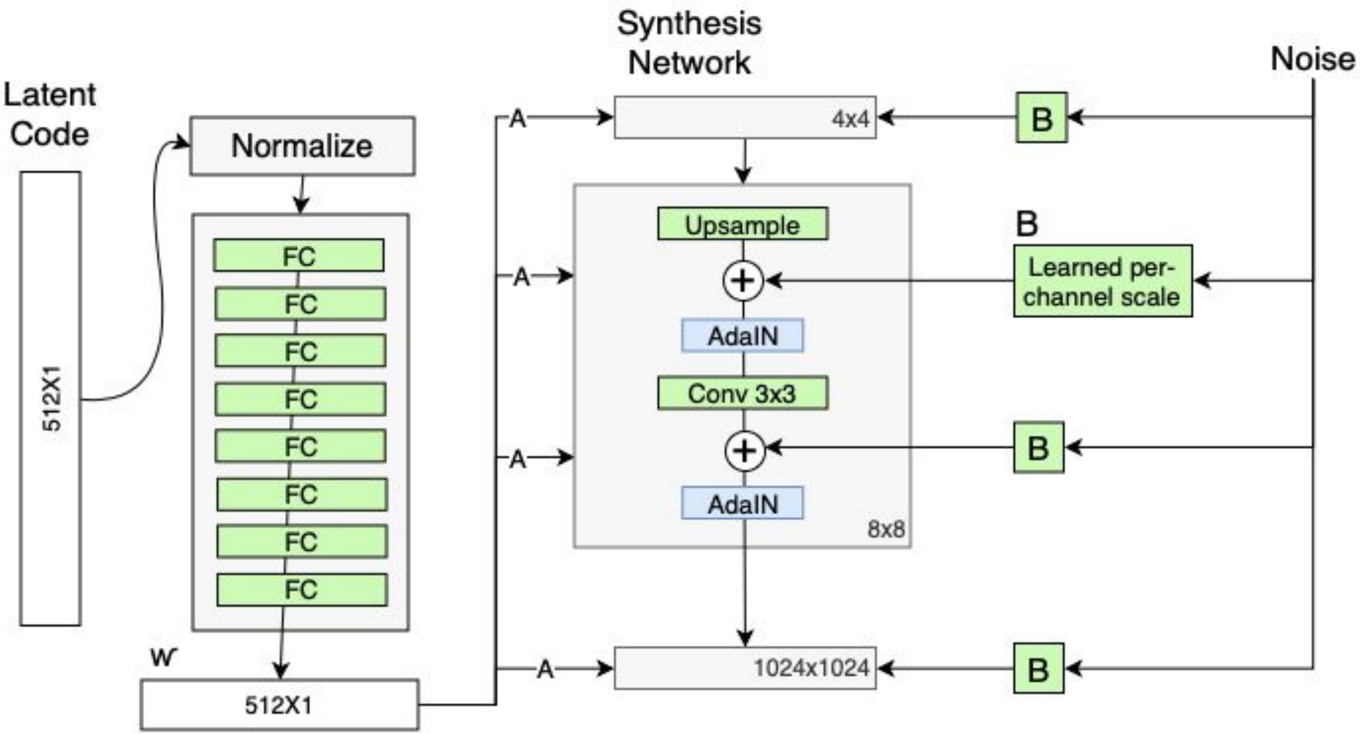
Generative adversarial network



$$\min_G \max_D V(D, G) = \mathbb{E}_{\boldsymbol{x} \sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z} \sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))]$$

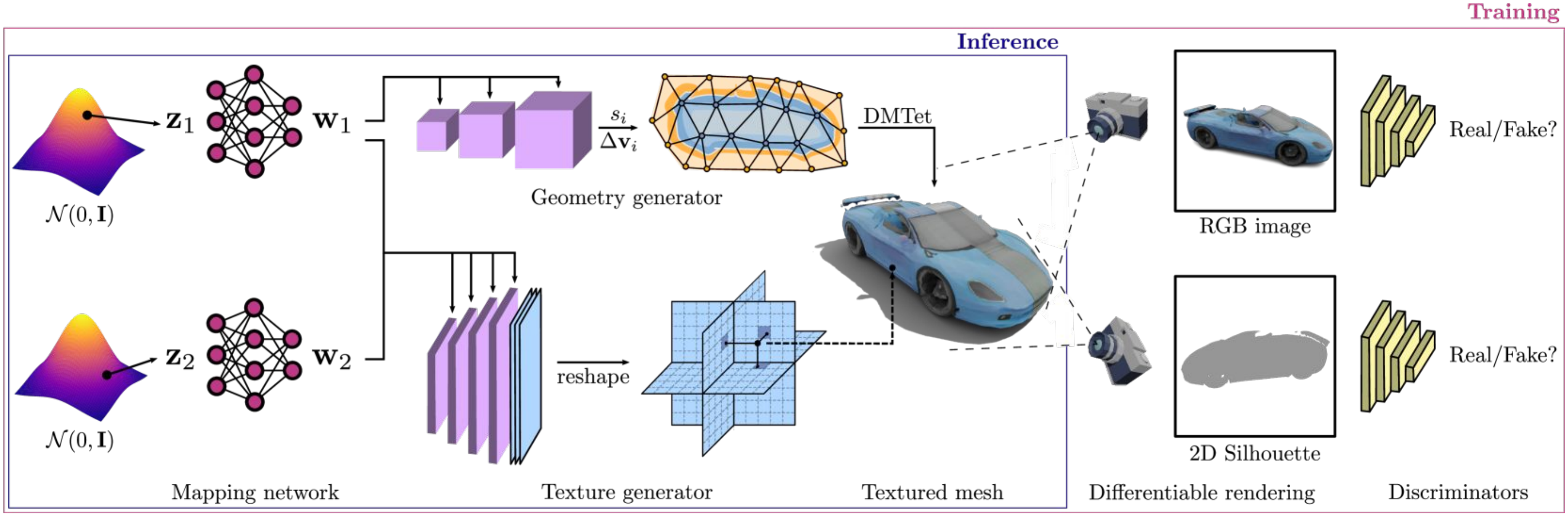
Goodfellow, Ian, et al. "Generative adversarial networks." *Advances in Neural Information Processing Systems (2014)*..

StyeGAN



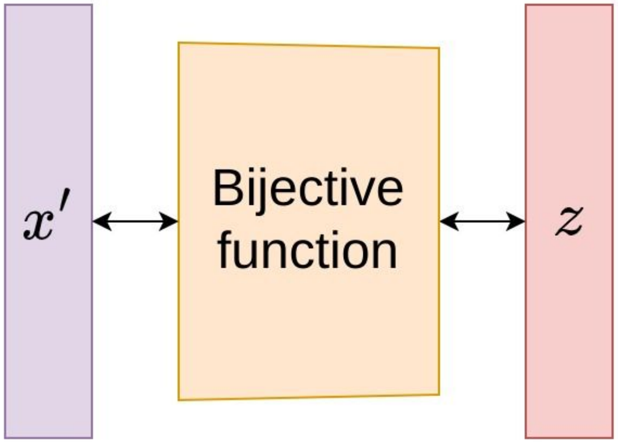
Karras, Tero, Samuli Laine, and Timo Aila. "A style-based generator architecture for generative adversarial networks." *Proceedings of the IEEE/CVF CVPR*. 2019.

GANs in complex systems

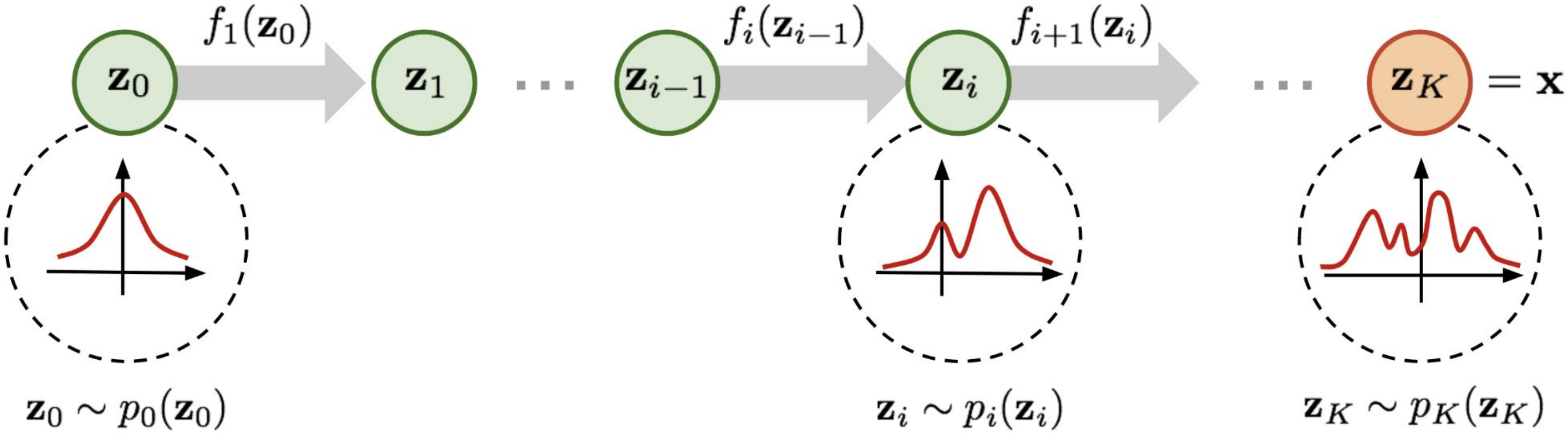


Zhu, Jun-Yan, et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks." *Proceedings of the IEEE international conference on computer vision*. 2017.

Normalizing flows



Normalizing flows

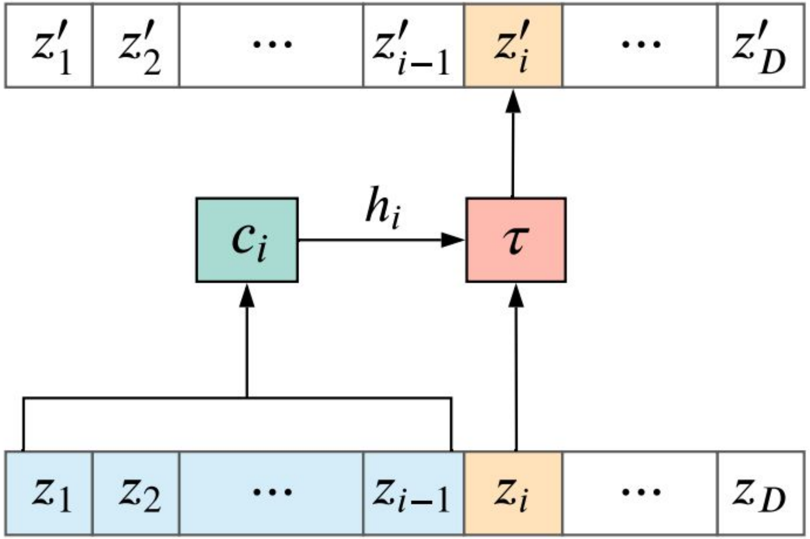


$$p_i(\mathbf{z}_i) = p_{i-1}(f_i^{-1}(\mathbf{z}_i)) \left| \det \frac{df_i^{-1}}{d\mathbf{z}_i} \right|$$

UNIVERSITY OF CAMBRIDGE

Rezende, Danilo, and Shakir Mohamed. "Variational inference with normalizing flows." *International conference on machine learning*. PMLR, 2015.

Normalizing flows - coupling

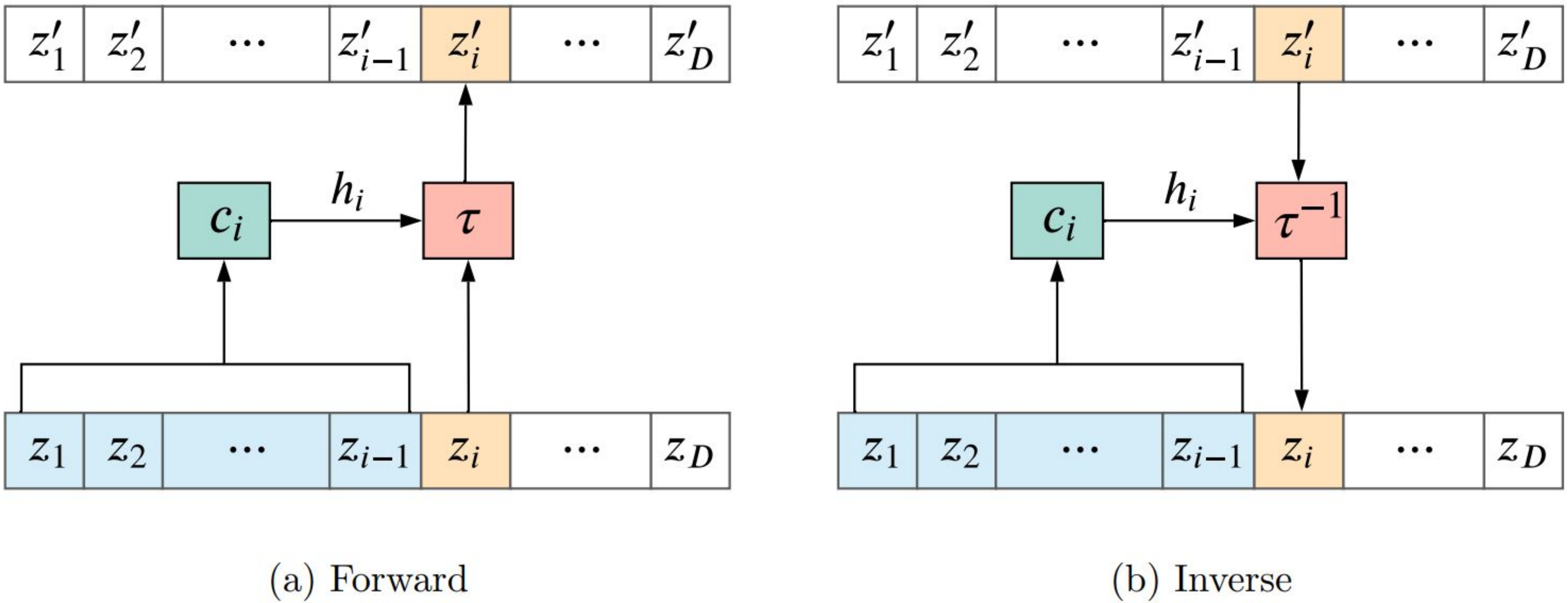


(a) Forward

UNIVERSITY OF CAMBRIDGE

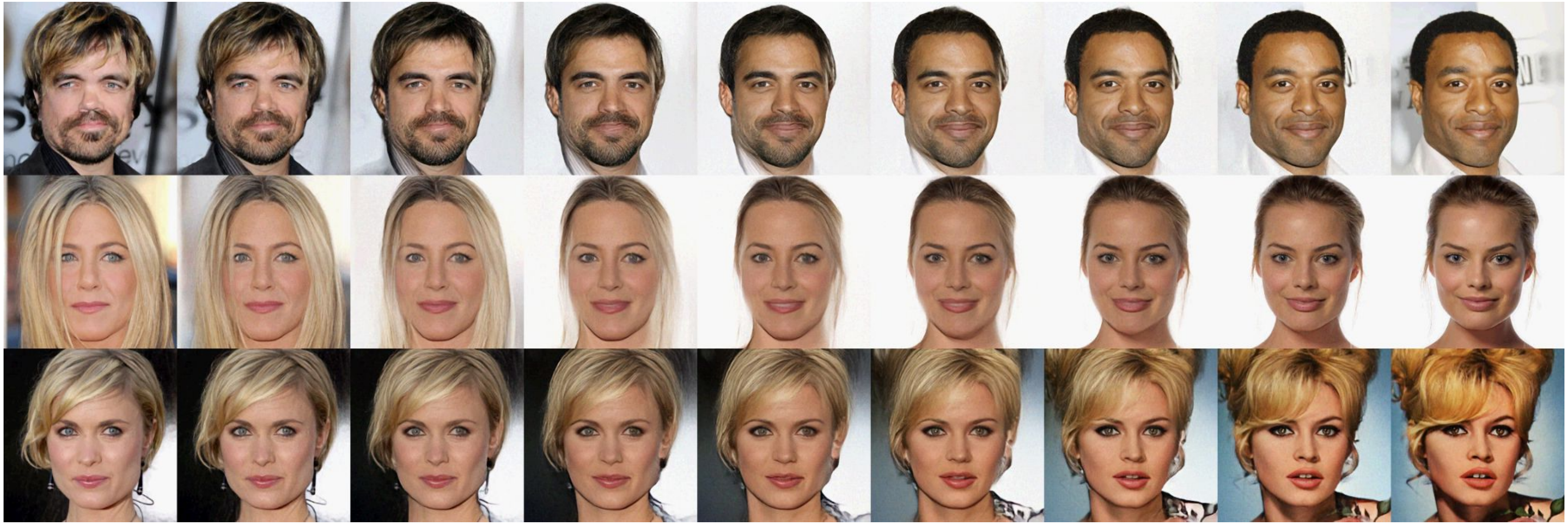
Dinh, Laurent, Jascha Sohl-Dickstein, and Samy Bengio. "Density estimation using real nvp." *arXiv preprint arXiv:1605.08803* (2016).

Normalizing flows - coupling



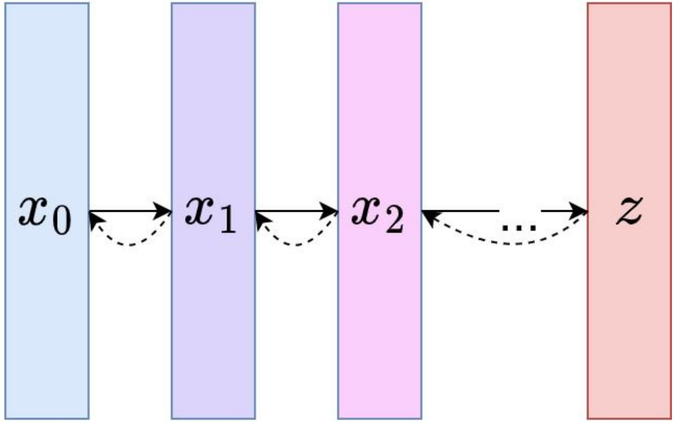
Dinh, Laurent, Jascha Sohl-Dickstein, and Samy Bengio. "Density estimation using real nvp." *arXiv preprint arXiv:1605.08803* (2016).

Flows - latent manipulation



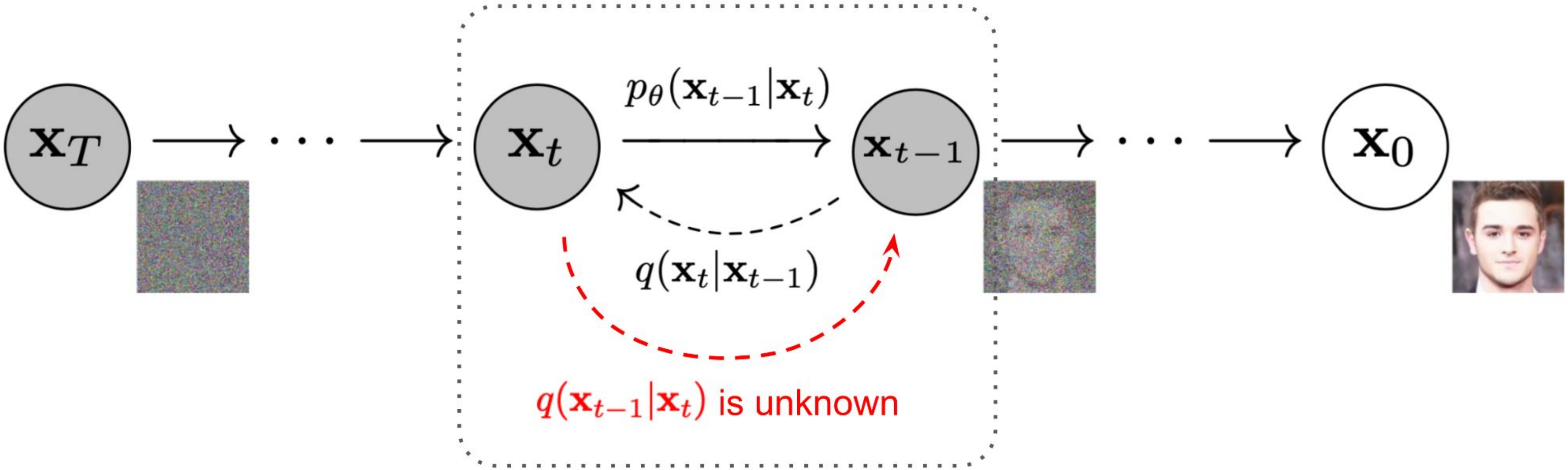
Kingma, Durk P., and Prafulla Dhariwal. "Glow: Generative flow with invertible 1x1 convolutions." *Advances in neural information processing systems* 31 (2018).

Diffusion models



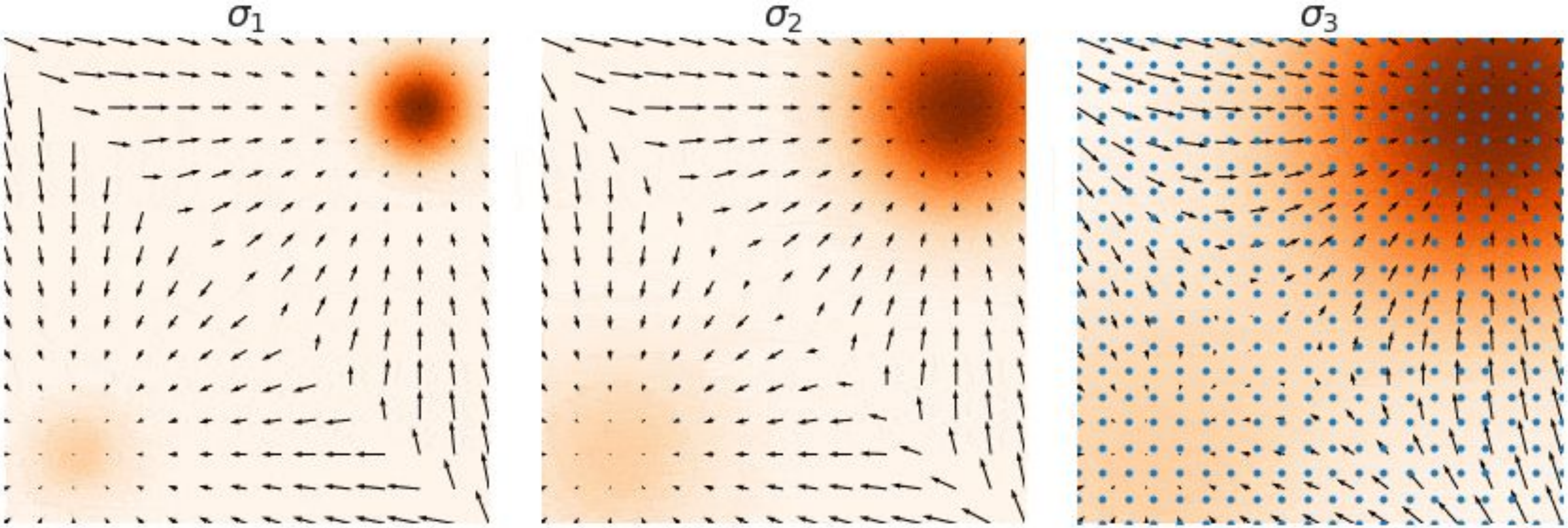
Diffusion models

Use variational lower bound



Sohl-Dickstein, Jascha, et al. "Deep unsupervised learning using nonequilibrium thermodynamics." *International Conference on Machine Learning*. PMLR, 2015.

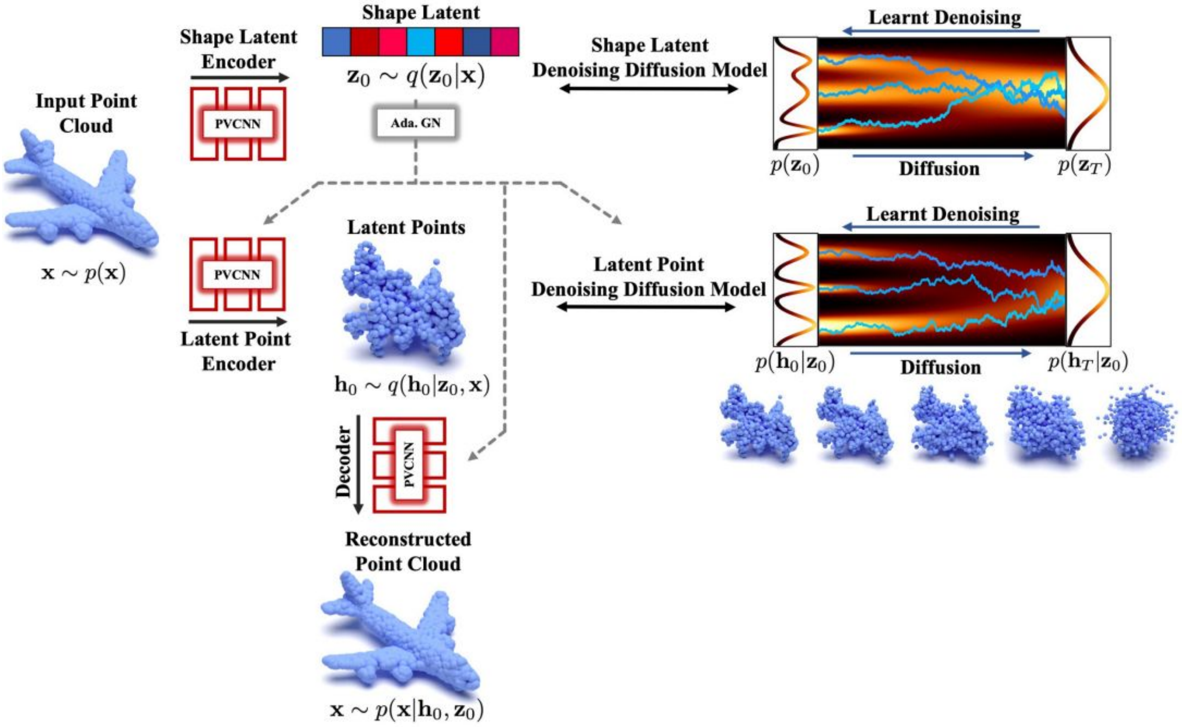
Score matching



$$\mathbf{x}_{i+1} \leftarrow \mathbf{x}_i + \epsilon \nabla_{\mathbf{x}} \log p(\mathbf{x}) + \sqrt{2\epsilon}\, \mathbf{z}_i, \quad i = 0, 1, \cdots, K$$

Song, Yang, et al. "Score-based generative modeling through stochastic differential equations." *arXiv preprint arXiv:2011.13456* (2020).

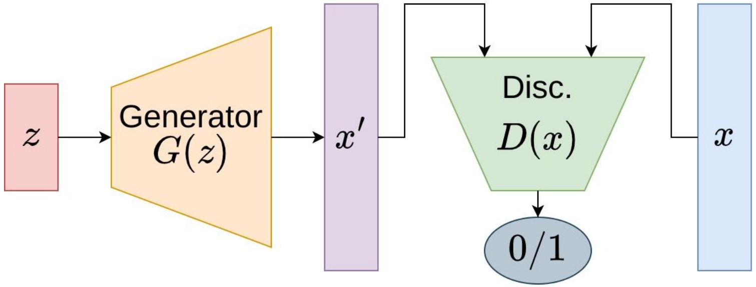
Hybrid models



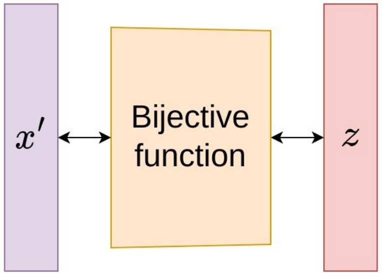
UNIVERSITY OF CAMBRIDGE

Zeng, Xiaohui, et al. "LION: Latent point diffusion models for 3D shape generation." *arXiv preprint arXiv:2210.06978* (2022).

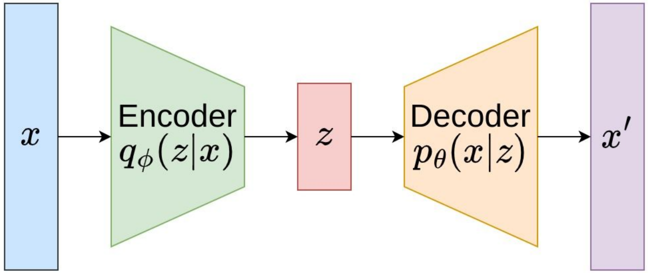
Evaluation with likelihoods



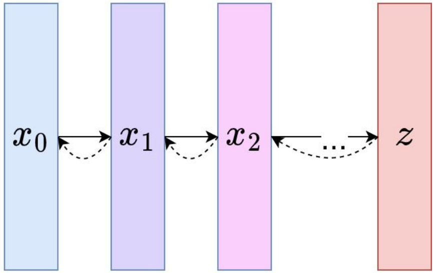
Generative adversarial network



Normalizing flow

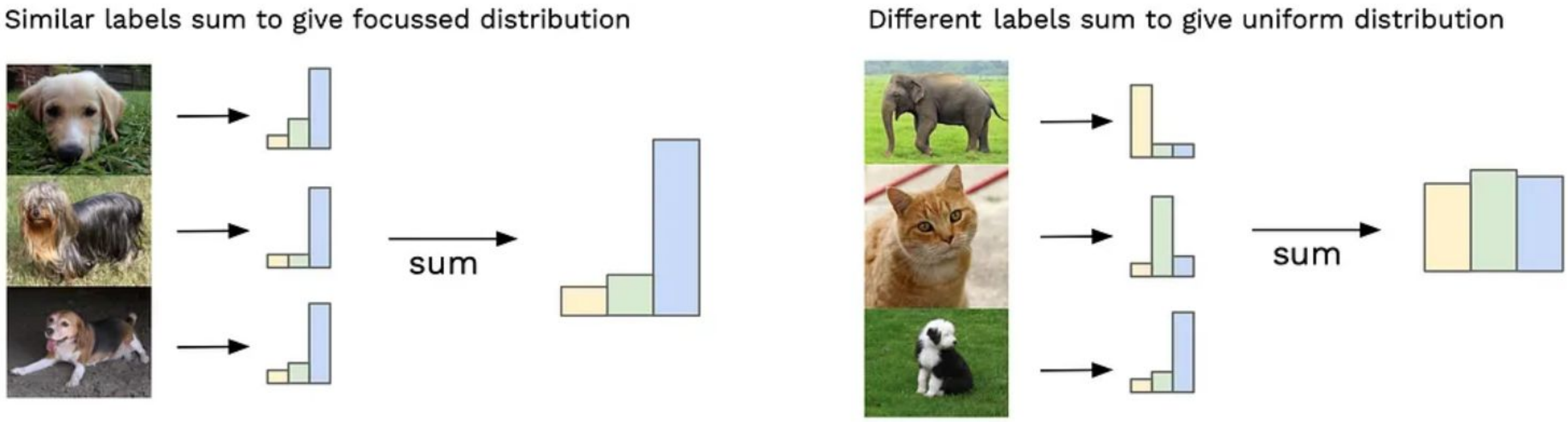


Variational autoencoder



Diffusion method

Inception score



Salimans, Tim, et al. "Improved techniques for training gans." *Advances in neural information processing systems* 29 (2016).

Frechet Inception distance

- Squared Wasserstein distance between generated and reference distributions
- Use intermediate activations of pre-trained classifier
- Assume distributions are multivariate normals

$$\text{FID} = ||\mu - \mu_w||_2^2 + \text{tr}(\Sigma + \Sigma_w - 2(\Sigma^{1/2}\Sigma_w\Sigma^{1/2})^{1/2}).$$

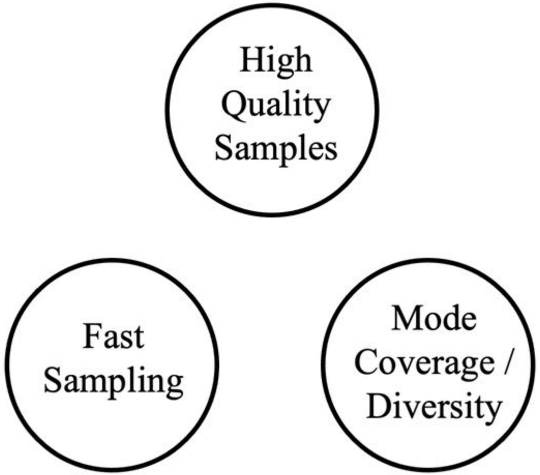
Heusel, Martin, et al. "Gans trained by a two time-scale update rule converge to a local nash equilibrium." *Advances in neural information processing systems* 30 (2017).

Requirements

- Handle high dimensional data
- Fast, efficient sampling
- High sample quality
- Diverse samples
- [Optional] Density evaluation
- [Optional] Low dimensional latent

Summary

The Generative Learning Trilemma



UNIVERSITY OF CAMBRIDGE