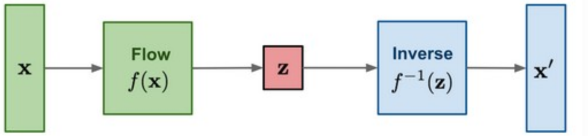
Diffusion Models

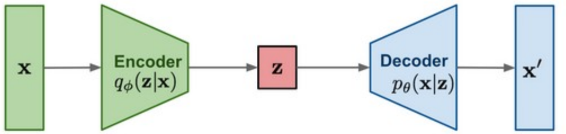
Narada Warakagoda

Generative modelling approaches with Deep Learning

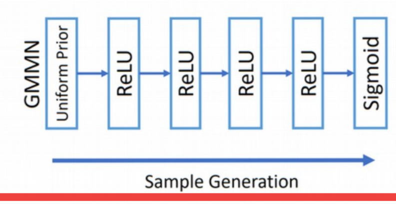
- Normalizing Flows



- Variational Auto Encoders (VAE)

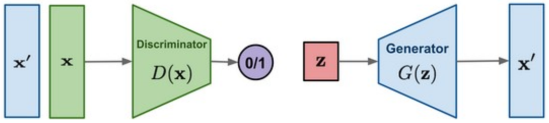


- Moment Matching Networks (MMN)

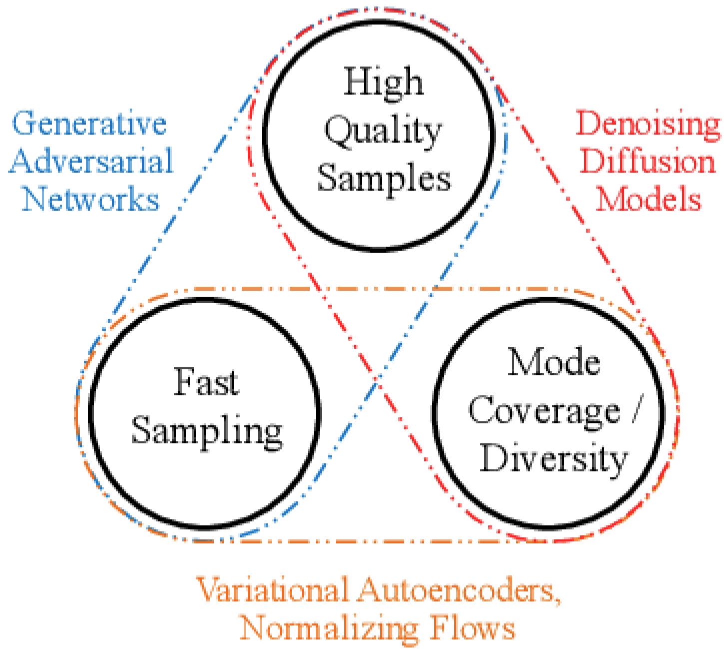


- Diffusion models

- Generative Adversarial Networks (GAN)

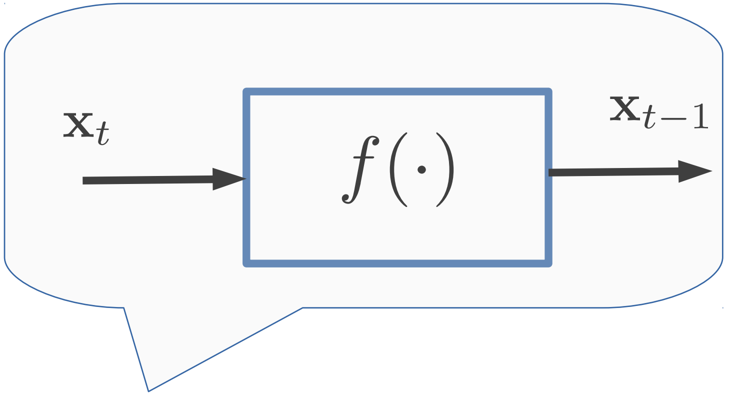


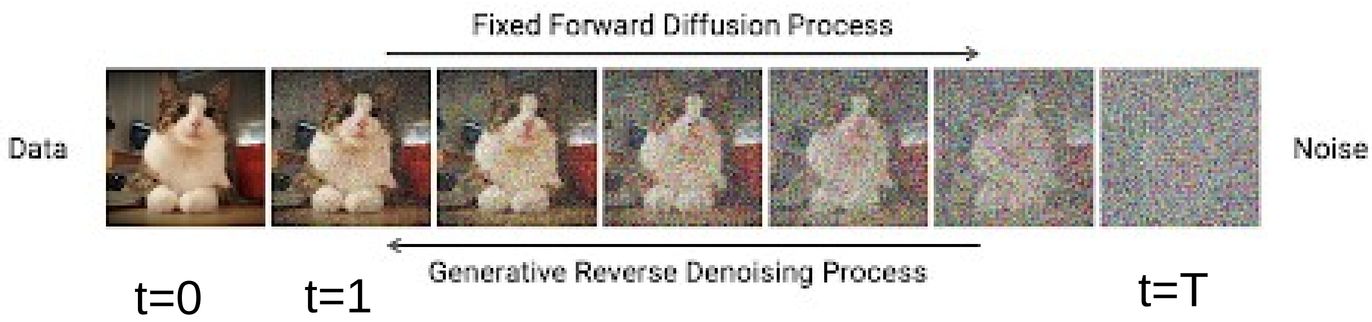
Generative modeling trilemma



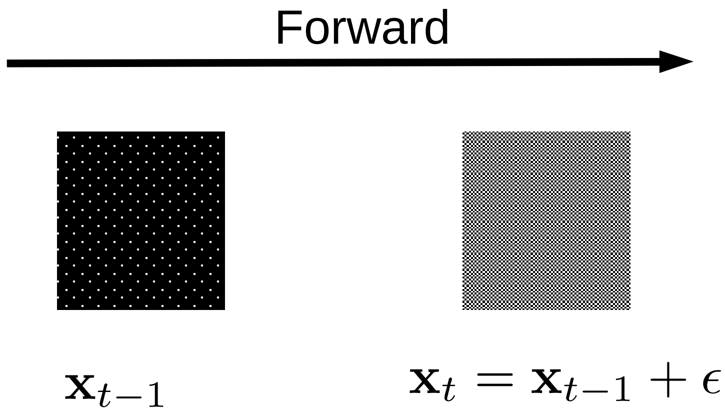
Denoising Diffusion Models (DDM) principle

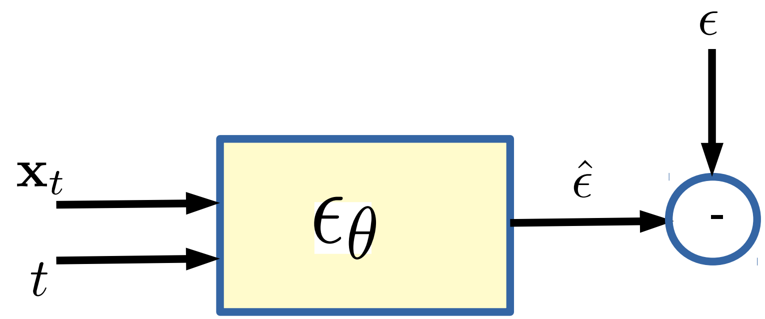
- Forward diffusion process
  - Add noise to the image at each step
  - Stop when the result is only noise
- Reverse diffusion process
  - Learn a mapping (neural net) from the noisy image to the previous image at each step
- Generation
  - Apply the mapping recursively starting from noise





DDM Algorithm





Algorithm 1 Training

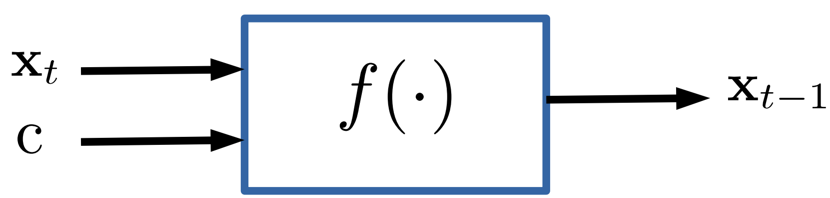
1: **repeat**
2: $\quad \mathbf{x}_0 \sim q(\mathbf{x}_0)$
3: $\quad t \sim \mathrm{Uniform}(\{1, \dots, T\})$
4: $\quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
5: $\quad$ Take gradient descent step on
$$\nabla_\theta \left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\underbrace{\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}}_{\mathbf{x}_t}, t) \right\|^2$$
6: **until** converged

Algorithm 2 Sampling

1: $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
2: **for** $t = T, \dots, 1$ **do**
3: $\quad \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$
4: $\quad \mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \underbrace{\frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)}_{\hat{\epsilon}} \right) + \sigma_t \mathbf{z}$
5: **end for**
6: **return** $\mathbf{x}_0$

Text-to-image generation

- Conditional generation
  - Condition $(\mathbf{c})$ = text string **(eg:A corgi playing a flame throwing trumpet)**



Famous diffusion systems

- GLIDE
- DALL-E2
- IMAGEN