Deep Generative Models

Chapter 6: Generation by Diffusion Process

Ali Bereyhi

ali.bereyhi@utoronto.ca

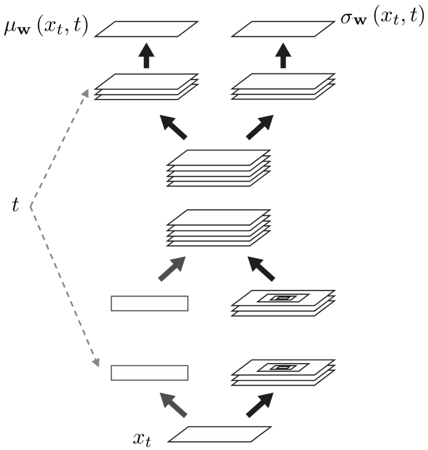
Department of Electrical and Computer Engineering

University of Toronto

Summer 2025

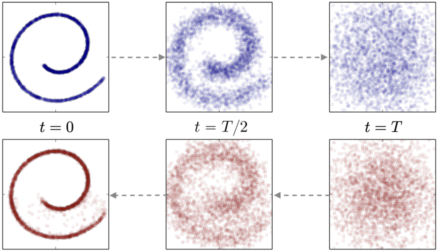
DPM: *Sohl-Dickstein et al.*

Sohl-Dickstein et al. implemented the first DPM in their work

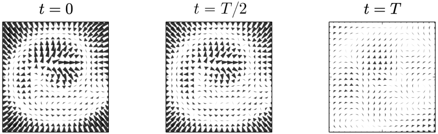


DPM: *Sohl-Dickstein et al.*

They tried it on the Swiss roll dataset to demonstrate the diffusion process



Interestingly, they could see the drift behavior



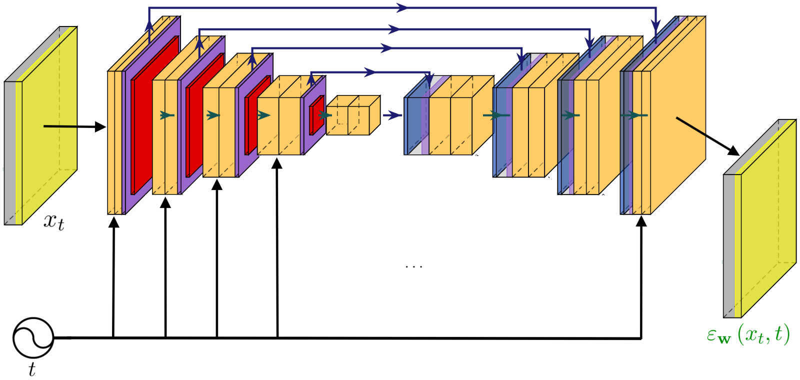
DPM: *Sohl-Dickstein et al.*

They also tried couple of classical datasets including MNIST and CIFAR

But, it was not the beating the benchmark at the time, i.e., 2015

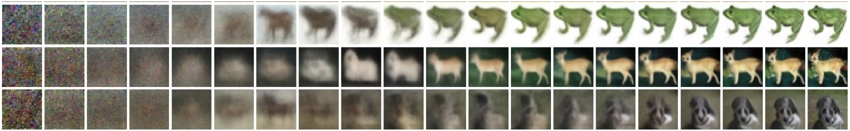
DDPM: *Ho et al.*

Ho et al. proposed DDPM and implemented it by embedding time in U-Net



DDPM: *Ho et al.*

They could demonstrate efficient reverse diffusion



and very inspiring high-quality samples

Improved DDPM: *Nichol and Dhariwal*

Nichol and Dhariwal kept U-Net, but improved training procedure by

- *learning the variances for the reverse diffusion, and*
- *tuning thetime scheduling for the noising process, i.e.,* α_t

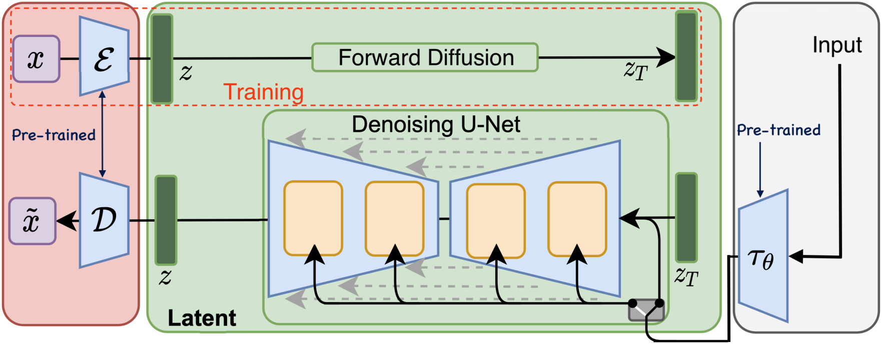
which ended up better sampling quality

Model	ImageNet	CIFAR
Glow (Kingma & Dhariwal, 2018)	3.81	3.35
Flow++ (Ho et al., 2019)	3.69	3.08
PixelCNN (van den Oord et al., 2016c)	3.57	3.14
SPN (Menick & Kalchbrenner, 2018)	3.52	-
NVAE (Vahdat & Kautz, 2020)	-	2.91
Very Deep VAE (Child, 2020)	3.52	2.87
PixelSNAIL (Chen et al., 2018)	3.52	2.85
Image Transformer (Parmar et al., 2018)	3.48	2.90
Sparse Transformer (Child et al., 2019)	3.44	**2.80**
Routing Transformer (Roy et al., 2020)	**3.43**	-
DDPM (Ho et al., 2020)	3.77	3.70
DDPM (cont flow) (Song et al., 2020b)	-	2.99
Improved DDPM (ours)	**3.53**	**2.94**

Stable Diffusion: *First Large-Scale LDM*

Stable Diffusion employs a VAE to get into the latent space

↳ *It uses diffusion to generate the latent representation of date*



This is a latent diffusion model (LDM)

DALL-E: *Another LDM*

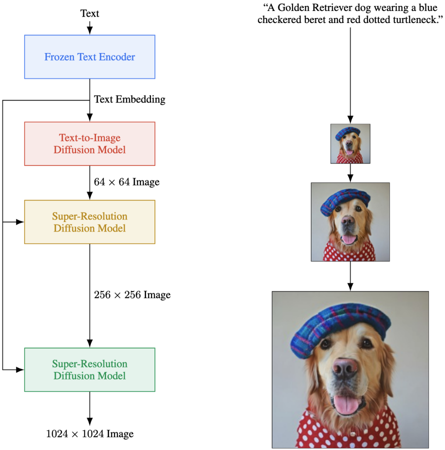
OpenAI also developed an LDM which uses diffusion for image generation

January 5, 2021 Milestone

DALL·E: Creating images from text

Google Imagen: *Diffusion in Image Space*

Google developed Imagen which uses diffusion in the image space



Final Notes

Probabilistic diffusion learns the reverse diffusion process

- *DPMs learn the reverse diffusion by ELBO maximization*
- *DDPMs are special case of DPMs which learn to estimate noising process*
- *Sampling diffusion models can take too long*
  - ↳ *We need to move enough number of time samples, e.g.,* $T = 1000$
  - ↳ *We can reduce sampling time by tricks like DDIM, e.g., down to* $T = 50$

Diffusion models scale slower than other models with data complexity

- **+** *We can learn to sample complex data more efficiently*
  - ↳ *We only need to learn how to denoise through time*
  - ↳ *Data complexity does not involve directly in the process*
- **−** *They are overkill for simple data*
  - ↳ *We can learn to sample MNIST much easier via GANs or VAEs*
  - ↳ *For simple data, we better stick to easier models*