Diffusion Models: DALL-E

Deep Learning and Neural Networks: Advanced Topics

Fabio Brau

March 1, 2023

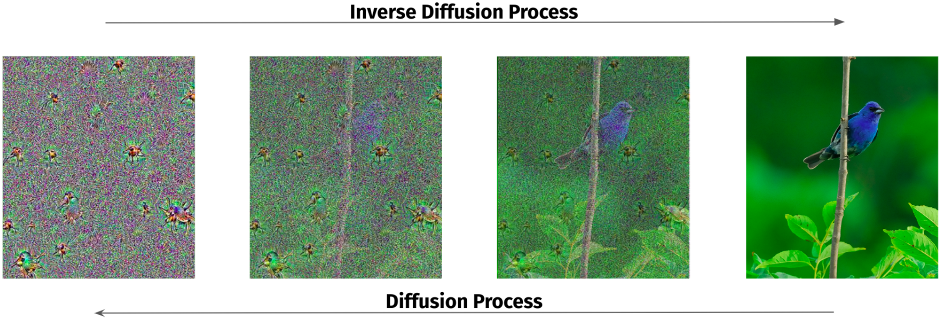
Scuola Superiore Sant'Anna, Pisa.

Introduction

Diffusion Models

Overview

Diffusion models are generative models that aim at denoising data



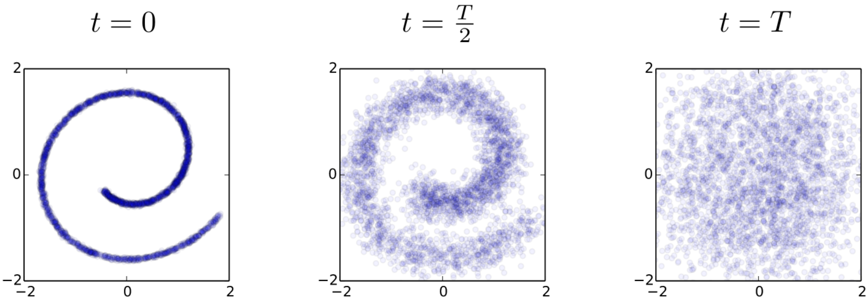
Timeline

2015) *...Non-equilibrium Thermodynamics*. Sohl-Dickstein et al. ICML

2020) *Denoising Diffusion Probabilistic Models*. Ho et al. NeurIPS.

2021) *Score-Based Generative Modeling Through SDE*. Song et al. ICLR.

Deep Unsupervised Learning using Non-Equilibrium Thermodynamics



Diffusion process as a **Markov Chain** with **Continuous State Space** and **Discrete Time**.[1]

[1] Sohl-Dickstein et al., "Deep Unsupervised Learning using Nonequilibrium Thermodynamics".

Reminder: Markov Chains with Discrete Time

Informal Definition

A sequence of random variables $\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(t)}, \cdots$, such that:

- $\mathbf{x}^{(t)} \in S$, where S **State Space**
- The future $\mathbf{x}^{(t+1)}$ depends on the present $\mathbf{x}^{(t)}$ but not on the past $\mathbf{x}^{(t-1)}$

Discrete State Space S

Continuous State Space S

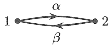
Reminder: MCDT with Discrete State Space

Definition

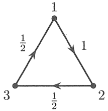
A sequence of random variables $\{\mathbf{x}^{(t)}\}_{t\in\mathcal{T}} \subseteq S$

- Discrete Time Property
  $\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(t)}, \cdots$
- Markov Property
  $\mathbb{P}\left(\mathbf{x}^{(t+1)} \in A \,|\, \mathbf{x}^{(0)}, \ldots, \mathbf{x}^{(t)}\right) = \mathbb{P}\left(\mathbf{x}^{(t+1)} \in A \,|\, \mathbf{x}^{(t)}\right)$

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$



$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix}$$



Two MCs with a discrete state space of respectively 2 and 3 states each.

Reminder: DTMC with Continuous State Space

Let assume $\mathbf{x}, \mathbf{y} \in S$ where S continuous state space (e.g. $S = \mathbb{R}^d$)˙**Joint Distribution** $p(\mathbf{x}, \mathbf{y})$

$$\mathbb{P}\left(\mathbf{x} \in A \,|\, \mathbf{y} \in B\right) = \int_A \int_B p\left(\mathbf{x}, \mathbf{y}\right)\, d\mathbf{x}\, d\mathbf{y}$$

Transactional Kernel $p(\mathbf{x} \,|\, \mathbf{y})$

$$p\left(\mathbf{x}, \mathbf{y}\right) = p(\mathbf{x} \,|\, \mathbf{y})\, p\left(\mathbf{x}\right)$$

Marginal Distribution $p(\mathbf{x})$

$$p\left(\mathbf{x}\right) = \int_S p\left(\mathbf{x}, \mathbf{y}\right)\, d\mathbf{y} = \int_S p(\mathbf{x} \,|\, \mathbf{y})\, p\left(\mathbf{x}\right)\, d\mathbf{y}$$

Definition

A sequence of random variables $\left\{\mathbf{x}^{(t)}\right\}_{t\in\mathcal{T}} \subseteq S$, such that the future $\mathbf{x}^{(t+1)}$ depends on the present $\mathbf{x}^{(t)}$ but not on the past $\mathbf{x}^{(t-1)}$.

- Discrete Time Property
  $\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(t)}, \cdots$
- Markov Property
  $\mathbb{P}\left(\mathbf{x}^{(t+1)} \in A \mid \mathbf{x}^{(0)}, \ldots, \mathbf{x}^{(t)}\right) = \mathbb{P}\left(\mathbf{x}^{(t+1)} \in A \mid \mathbf{x}^{(t)}\right)$

Discrete State Space S

Continuous State Space S

Broader Impacts

CLIP Model

“We also found discrepancies across gender and race for people categorized into the ‘crime’ and ‘non-human’ categories...”[2]

[2] Radford et al., “Learning Transferable Visual Models From Natural Language Supervision”.

Thanks for the attention

Fabio Brau

Scuola Superiore Sant'Anna, Pisa

fabio.brau@santannapisa.it

retis.santannapisa.it/~f.brau

linkedin.com/in/fabio-brau

TELECOMMUNICATIONS, COMPUTER ENGINEERING, AND PHOTONICS INSTITUTE

Sant'Anna School of Advanced Studies – Pisa

ECCELLENZA MIUR 2018-2022

ROBOTICS & AI

Sant'Anna Scuola Universitaria Superiore Pisa

Retis Real-Time Systems Laboratory