Writeup of Deep Unsupervised Learning using Nonequilibrium Thermodynamics

Daniel Lomholt Christensen (tqv636), Simon Hilding-Nørkjær (xsj415),
Eric Steen Ravn Bowman (ztp370), Chamilla Terp (bmx626)

March 2025

Abstract

Diffusion models offer a framework that balances flexibility and tractability by systematically introducing noise into data and learning how to reverse the process. By gradually adding noise, the model transforms complex data distributions into analytically tractable forms, allowing for efficient training and sampling. The reverse process then reconstructs data by learning an optimal mapping from noise to the original distribution. This approach enables high-quality data synthesis and sampling from complex distributions. Their ability to maintain stability during training while still capturing intricate data structures has led to widespread adoption in image generation, speech processing, and scientific simulations. In this work, we introduce diffusion models, review their essential concepts, and highlight their application in two-dimensional image data.

1 Introduction

In statistics and machine learning, diffusion models are a class of models that progressively apply noise to data and then learn how to reverse the process [1]. They utilize a Monte Carlo Markov Chain (MCMC) approach to apply noise iteratively - an idea used in nonequilibrium statistical physics. A neural network is then trained to optimize parameters that minimize the difference between the original input data and the self-generated data. A visualization of the process can be seen in Figure 1.

Diffusion models enable the design of probabilistic models that are both *tractable* and *flexible*. In this context, *tractable* refers to a model that can be analytically evaluated and is easy to fit, while *flexible* denotes the ability to adapt to the structure of any given data. Historically, achieving both of these objectives simultaneously has been challenging [4].

In this work, we introduce the model, review its essential concepts, and highlight its applications to two-dimensional image data.

2 Algorithm

The diffusion model algorithm consists of four key components: the forward process, the reverse process, the model probability evaluation and the neural network training to maximize the lower bound on the likelihood. In the **forward process**, input data is gradually transformed into an analytically tractable distribution by iteratively applying a diffusion kernel (i.e., tractable noise) with a carefully chosen diffusion rate. This process effectively destroys the original structure of the data, eventually reaching a null-information distribution. The **reverse process** then reconstructs a distribution that resembles the original data by learning the optimal parameters through training on the diffusion kernels. The goal is to recover the data distribution from its

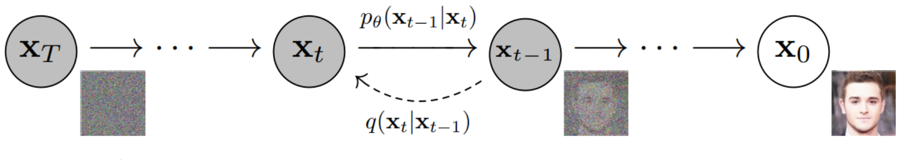

Figure 1: Visualization of the MCMC used for diffusion models. The data goes from noisy picture (left) to the original picture (right). From Ho, Jain, and Abbeel [2].

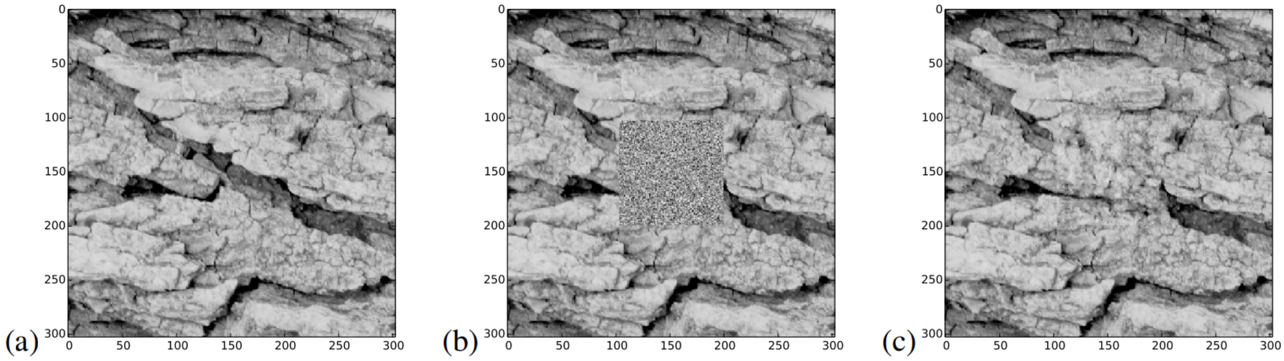

Figure 2: (a) Sample bark image [3]. (b) Central 100x100 pixel region replaced by Gaussian noise. (c) Central region replaced by sampling from the model trained on similar bark images. The sampling was conditioned on the rest of the image, and was therefore able to reconstruct some of the long range order [4].

completely randomized state. The **model probability** of correctly reconstructing the original data is computed from the reverse process evaluated at $t = 0$:

$$p(\mathbf{x}^{(t=0)}) = \int d\mathbf{x}^{(t=1\cdots T)} \underbrace{q(\mathbf{x}^{(t=1\cdots T)}|\mathbf{x}^{(t=0)})}_{\text{Forward trajectory sample}} \cdot \underbrace{p(\mathbf{x}^{(t=T)}) \prod_{\tau=1}^{T} \frac{p(\mathbf{x}^{(t=\tau-1)}|\mathbf{x}^{(t=0)})}{q(\mathbf{x}^{(t=\tau)}|\mathbf{x}^{(t=\tau-1)})}}_{\text{Relative probability of the forward and reverse trajectories}} . \tag{1}$$

Here $q(\mathbf{x}^{(t)})$ is the data distribution and $p(\mathbf{x}^{(t)})$ is the probability the generative model assigns to the data. From this probability the likelihood function can be computed as the test statistic:

$$L = \int d\mathbf{x}^{(t=0)} q(\mathbf{x}^{(t=0)}) \ln \left[p(\mathbf{x}^{(t=0)}) \right] \geq K. \tag{2}$$

To ensure stability during training, a lower bound K on the likelihood is established and maximized based on the Kullback–Leibler (KL) divergence, D_{KL}, which quantifies the statistical distance between the forward and reverse process:

$$\begin{aligned} K = -\sum_{t=2}^{T} \int d\mathbf{x}^{(0)} d\mathbf{x}^{(1)} q(\mathbf{x}^{(0)}|\mathbf{x}^{(0)}) \cdot \overbrace{D_{KL}\left(q(\mathbf{x}^{(t-1)}|\mathbf{x}^{(t)}, \mathbf{x}^{(0)} || p(\mathbf{x}^{(t-1)}|\mathbf{x}^{(t)} \right)}^{\text{KL-divergens: Relative entropy between q and p}} \\ + \underbrace{H_q(\mathbf{X}^{(T)}|\mathbf{X}^{(0)})}_{\substack{\text{Final entropy of the} \\ \text{forward process} \\ \text{given data}}} - \underbrace{H_q(\mathbf{X}^{(1)}|\mathbf{X}^{(0)})}_{\substack{\text{Initial entropy} \\ \text{given data}}} - \underbrace{H_p(\mathbf{X}^{(T)})}_{\substack{\text{Final entropy of} \\ \text{reverse process}}} . \end{aligned} \tag{3}$$

The lower bound K is further constrained by entropy principles from statistical physics: entropy must increase during the forward diffusion process (H_q) and decrease during the reverse process (H_p). These constraints can be computed analytically.

3 Application

The model described above has been tested on data with both one-dimensional and two-dimensional structures. For briefness, we focus here on the *two-dimensional* case.

An array of images are taken through the forward process, where Gaussian noise is applied to them. A convolutional neural network (CNN) is then trained to reverse this process. It does this by fitting its neurons, to maximize K. Finally, novel images resembling the original samples, is generated by inputting an arbitrary noisy image.

A result from this implementation on a specific image can be seen in Figure 2.

4 Conclusions

Diffusion models provide a powerful and theoretically grounded approach to generative modeling. By systematically introducing and reversing noise, they offer a balance between tractability and flexibility. Their application to image data showcases their potential for high-quality sample generation.

References

[1] Minshuo Chen et al. *An Overview of Diffusion Models: Applications, Guided Generation, Statistical Rates and Optimization*. 2024. arXiv: 2404.07771 [cs.LG]. URL: https://arxiv.org/abs/2404.07771.

[2] Jonathan Ho, Ajay Jain, and Pieter Abbeel. *Denoising Diffusion Probabilistic Models*. arXiv:2006.11239 [cs]. Dec. 2020. DOI: 10.48550/arXiv.2006.11239. URL: http://arxiv.org/abs/2006.11239 (visited on 02/27/2025).

[3] Svetlana Lazebnik, Cordelia Schmid, and J. Ponce. "A sparse texture representation using local affine region". In: *IEEE transactions on pattern analysis and machine intelligence* 27 (Sept. 2005), pp. 1265–78. DOI: 10.1109/TPAMI.2005.151.

[4] Jascha Sohl-Dickstein et al. *Deep Unsupervised Learning using Nonequilibrium Thermodynamics*. 2015. arXiv: 1503.03585 [cs.LG]. URL: https://arxiv.org/abs/1503.03585.