

Diffusion Models in Depth: From a Theoretical and a Practical Perspective

Gregory Sedykh

June 28th 2024



**UNIVERSITÉ
DE GENÈVE**

FACULTY OF SCIENCE
Department of Informatics

Bachelor thesis for the degree of Bachelor of Science in Computer Science
Supervised by Prof. Stéphane Marchand-Maillet

Abstract

Contents

1	Introduction	2
2	Denoising Diffusion Probabilistic Models	3
2.1	Forward Process	3
2.2	Reverse Process	3
2.3	Comparaison with other diffusion models	3
3	Improvements upon DDPMs	4
4	Results	5
5	Conclusion	6

1 Introduction

Diffusion models have gained widespread popularity since 2020, as models such as DALL-E, Stable Diffusion and Midjourney have proven to be capable of generating high-quality images given a text prompt. Furthermore, OpenAI's recent announcement of Sora has shown that diffusion models have now also become highly capable of generating minute long high-definition videos from a text prompt. [1]

These models date back to 2015, where the idea of a diffusion model appeared, based on diffusion processes used in thermodynamics. [5]

Denosing Diffusion Probabilistic Models (DDPMs) were a development of the original diffusion probabilistic model introduced in 2015. [3]

Subsequently, OpenAI improved upon the original DDPMs which did not have ideal log likelihoods [3] while also using less forward passes and therefore speeding up the sampling process. [4]

The most recent progress done by OpenAI has allowed their diffusion models to obtain better metrics and better sample quality than Generative Adversarial Networks (GANs) which were previously considered the state-of-the-art in image generation. [2]

The fairly recent apparition of diffusion models means not only that there is still a lot to be discovered about them, but also that progress is being made rapidly.

The theory behind diffusion models was mainly founded when Ho et al. [3] introduced their DDPMs in 2020, but many improvements have been made upon their work since then.

To understand what these models are and how they work, it is crucial to understand how DDPMs were developed, what choices were made when developing them and why these choices were made, as well as why changes were made to the original model and how they were made.

This report aims to provide an overview of the theory behind diffusion models, as well as a practical guide on how to implement a simple diffusion model using PyTorch, in order to compare what the theory shows us and what the practical implementation gives us.

2 Denoising Diffusion Probabilistic Models

A **diffusion model** is a generative model that consists of two Markov chains, one forward and one reverse.

Given an input (e.g. an image), the forward process will destroy the image by gradually adding Gaussian noise at each step of the process. [3]

The reverse process' objective is to "reverse" the forward process, starting from the noisy image and step-by-step, estimating the noise that was added to the image at each step and removing it until the first step is reached, where we should obtain the original image. [3]

2.1 Forward Process

The forward process is a Markov process that starts from the original image x_0 and adds Gaussian noise during T steps, where each step $i \in [1, T]$ has a size $\beta_i \in \{\beta_1, \dots, \beta_T\}$, which results in a more noisy image x_i . [3]

Formally, we obtain:

$$q(x_1, \dots, x_T | x_0) = \prod_{t=1}^T q(x_t | x_{t-1}) \quad (1)$$

$$q(x_t | x_{t-1}) = \mathcal{N}(x_t; \sqrt{1 - \beta_t} x_{t-1}, \beta_t I) \quad (2)$$

Equation 2 gives us a single step forward: given the previous image x_{t-1} , we add noise from a Gaussian distribution with mean $\sqrt{1 - \beta_t} x_{t-1}$ and variance $\beta_t I$ to obtain the next image x_t .

Equation 1 gives us the full forward process from the original image x_0 to the final image x_T .

It is important to note that as $T \rightarrow \infty$ and with a correct choice of β_t , x_T will become a sample of an isotropic Gaussian distribution ($\mathcal{N}(0, I)$). [4] [5].

This is important for the reverse process, as it will allow us to take a sample $x_T \sim \mathcal{N}(0, I)$ and reverse the forward process to obtain the original image x_0 (however this cannot be done so simply, as seen in section 3) [4].

2.2 Reverse Process

2.3 Comparaison with other diffusion models

3 Improvements upon DDPMs

4 Results

5 Conclusion

References

- [1] Tim Brooks et al. “Video generation models as world simulators”. In: (2024). URL: <https://openai.com/research/video-generation-models-as-world-simulators>.
- [2] Prafulla Dhariwal and Alex Nichol. *Diffusion Models Beat GANs on Image Synthesis*. 2021. arXiv: [2105.05233 \[cs.LG\]](#).
- [3] Jonathan Ho, Ajay Jain, and Pieter Abbeel. *Denoising Diffusion Probabilistic Models*. 2020. arXiv: [2006.11239 \[cs.LG\]](#).
- [4] Alex Nichol and Prafulla Dhariwal. *Improved Denoising Diffusion Probabilistic Models*. 2021. arXiv: [2102.09672 \[cs.LG\]](#).
- [5] Jascha Sohl-Dickstein et al. *Deep Unsupervised Learning using Nonequilibrium Thermodynamics*. 2015. arXiv: [1503.03585 \[cs.LG\]](#).