

# Diffusion Models

Plote

Diffusion Models is a probabilistic generative model based on gradually corrupting data into noise and generating data from noise by learning the inverse denoising process. It is widely used for high-quality sample generation and is trained by optimizing the variational lower bound.

September 25, 2024

# Contents

1. INTRODUCTION .....	3
2. FOUNDATIONS OF DIFFUSION MODELS .....	4
2.1. Denoising Diffusion Probabilistic Models (DDPMs) .....	4
2.1.1. Forward Diffusion Process: .....	4
2.1.2. Reverse Generation Formula .....	5

# 1. INTRODUCTION

扩散模型已经成为最先进的深度生成模型家族。它们打破了生成对抗网络(GANs)在具有挑战性的图像合成任务中的长期统治地位，并且在各种领域也显示出潜力，包括计算机视觉，自然语言处理，时间数据建模，多模态建模，鲁棒机器学习，到计算化学和医学图像重建等领域的跨学科应用。

## 2. FOUNDATIONS OF DIFFUSION MODELS

Diffusion models are a family of probabilistic generative models that progressively destruct data by injecting noise, then learn to reverse this process for sample generation.

目前对扩散模型的研究主要基于三种主要的公式:

- denoising diffusion probabilistic models (DDPMs)
- score-based generative models (SGMs)
- stochastic differential equations(Score SDEs)

我们将对这三个公式进行独立的介绍, 同时讨论它们之间的联系。

### 2.1. Denoising Diffusion Probabilistic Models (DDPMs)

#### 2.1.1. Forward Diffusion Process:

正向扩散过程实际上是在原始图像上逐步添加高斯噪声。这个过程可以视为对图像进行“模糊化”, 每一步都会增加一定程度的噪声, 最终使得图像变得不可辨认, 接近于标准高斯噪声。这个过程可以看作是一个马尔可夫链

随着时间步  $t$  的增加, 图像中的信息逐渐被噪声覆盖, 直到最终形成一个几乎完全随机的噪声图像。这个过程是逐步的, 每一步都可以被看作是对图像进行微小的扰动。

在每一步  $t$ , 我们用一个高斯分布  $q(x_t | x_{t-1})$  来描述上一步  $x_{t-1}$  到当前步骤  $x_t$  的转换, 我们最终可以得到接近高斯分布的噪声

均值:  $\sqrt{1 - \beta_t}x_{t-1}$  表示前一个状态  $x_{t-1}$  经过缩放, 控制当前状态的中心位置

方差:  $\beta_t I$  是一个标量乘以单位矩阵  $I$ , 表示添加噪声的强度, 随着  $t$  增加,  $\beta_t$  通常增大, 导致噪声逐步增大

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

这里的  $\beta_t$  在正向扩散中控制噪声强度的超参数, 通常是自己设置的超参数类型:

一般有 3 种方式来设置:

- 线性调度: 可以在一定范围内线性增加

$$\beta_t = \text{Max\_beta} \cdot \frac{t}{T} \quad (2)$$

这里的 "MAX\_beta" 是最大噪声强度,  $T$  是总的时间步数

- 余弦调度: 使用余弦函数来调整噪声强度。这种方式可以提供更平滑的变化:

$$\beta_t = \frac{1 - \cos(\frac{t}{T} \cdot \pi)}{2} \quad (3)$$

逐步将数据添加噪声, 相当于逐步丢失信息, 最终得到一个噪声状态。而模型也通过多次小幅添加噪声, 更细致地学习每一步的变化

### 2.1.2. Reverse Generation Formula

反向生成可以视为对正向扩散过程的回溯。

反向生成过程开始于一个随机噪声样本，通过一系列逐步的去噪操作，逐渐生成清晰的样本。这是通过学习数据的逆扩散过程来实现的。

$$p_{\theta}(x_{t-1} | x_t) = \mathcal{N}(x_{t-1}; \mu_{\theta}(x_t, t), \Sigma_{\theta}(x_t, t)) \quad (4)$$

从标准高斯分布生成初始化  $x_T$ , 作为反向生成的起点

逐步去噪音: 在每一个时间步  $t$  中, 使用神经网络预测前一个状态  $x_{t-1}$ :

$$x_{t-1} = \mu_{\theta}(x_t, t) + \sigma_{\theta}(x_t, t) \cdot \epsilon \quad (5)$$

这里  $\epsilon$  是从标准高斯分布中采样的噪声，用于引入随机性。

这里的  $\mu_{\theta}$  和  $\sigma_{\theta}$  是通过训练神经网络学习得到的，目标是最小化重构损失，使得模型能够准确预测每一步的去噪声过程。

重复进行去噪, 从  $t = T$  到  $t = 1$ , 逐步生成最终样本  $x_0$ , 最终得到的是与训练数据分布相似的有效样本