

论文阅读报告:

Denoising Diffusion Probabilistic Models

刘昭阳 25215133

中山大学数学学院 (珠海)

2025 年 12 月 1 日

摘要

本报告对 Jonathan Ho, Ajay Jain 和 Pieter Abbeel 撰写的论文《Denoising Diffusion Probabilistic Models》(去噪扩散概率模型)进行了全面的分析。我们探讨了扩散模型的理论基础、论文中提出的具体实现细节以及所取得的成果。此外,本报告还讨论了该工作的主要贡献、从研究中获得的个人见解以及对未来研究的潜在衍生想法。本报告旨在深入理解扩散模型如何作为一类强大的生成模型出现,并挑战 GAN 的主导地位。

目录

1	引言	3
2	背景与相关工作	3
3	方法论	4
3.1	前向扩散过程	4
3.2	逆向去噪过程	4
3.3	目标函数	5
3.4	模型架构	6
3.5	实现细节	6
3.6	训练与采样算法	6
4	实验与结果	7
4.1	样本质量	7
4.2	与其他模型的比较	7
4.3	渐进式编码与插值	8
4.4	消融实验	8

目录	2
5 主要贡献	8
6 个人启发与见解	9
6.1 迭代优化的力量	9
6.2 信噪比作为课程学习	9
6.3 生成建模作为逆物理过程	9
7 衍生想法与未来方向	9
7.1 加速采样	9
7.2 条件生成与控制	10
7.3 其他领域的扩散	10
7.4 潜在扩散	10
8 理论分析	10
8.1 变分下界的推导	11
8.2 与朗之万动力学的联系	11
9 结论	12

1 引言

生成建模是机器学习中的一项基本任务，旨在从一组观测样本中学习潜在的数据分布 $p_{data}(x)$ 。在过去十年中，深度生成模型在合成高保真图像、音频和文本方面取得了显著的成功。最著名的生成模型家族包括生成对抗网络 (GAN)、变分自编码器 (VAE) 和基于流的模型 (Flow-based models)。

由 Goodfellow 等人引入的 GAN 采用生成器和判别器之间的极小极大博弈。虽然它们能够生成高质量的样本，但由于模式崩溃和训练不稳定性，它们通常难以训练。另一方面，VAE 优化对数似然的变分下界。它们在理论上是有根据的，并且训练稳定，但与 GAN 相比，通常会产生模糊的样本。基于流的模型允许精确的似然计算和有效的推理，但通常需要专门的架构，并且计算成本高昂。

最近，一类被称为去噪扩散概率模型 (DDPM) 的新型生成模型已成为一种强大的替代方案。受非平衡热力学的启发，扩散模型定义了一个前向过程，该过程逐渐向数据添加噪声，直到其变成纯高斯噪声；以及一个逆向过程，该过程学习逐步去噪数据以恢复原始样本。

Jonathan Ho, Ajay Jain 和 Pieter Abbeel (2020) 的论文《Denoising Diffusion Probabilistic Models》证明，扩散模型可以生成与 GAN 质量相当甚至更高的质量图像，同时保持基于似然的模型的优点，如稳定的训练和良好的模式覆盖。本报告深入探讨了这篇开创性论文的细节，分析了其方法论、结果以及对生成式 AI 领域的影响。

2 背景与相关工作

使用扩散过程进行生成建模的概念最早由 Sohl-Dickstein 等人在 2015 年提出。他们提出通过迭代的前向扩散过程破坏数据分布中的结构，然后学习逆向扩散过程来恢复结构。然而，这项早期工作在当时并没有达到与 GAN 相当的样本质量。

Ho 等人 (2020) 在此基础上，建立了扩散概率模型与朗之万动力学去噪分数匹配 (Song & Ermon, 2019) 之间的联系。他们表明，扩散模型的训练目标等价于加权变分下界，可以简化为去噪目标。

与 VAE 不同，VAE 的潜在代码通常比数据低维，而扩散模型在整个过程中保持相同的维度。扩散模型中的潜在变量是一系列噪声图像，而不是压缩表示。这使得模型能够捕捉细粒度的细节并生成高分辨率图像。

DDPM 的成功引发了扩散模型研究的热潮，导致了在采样速度 (DDIM)、条件生成 (分类器引导、无分类器引导) 和文本到图像生成 (GLIDE, DALL-E 2, Stable Diffusion) 方面的改进。因此，理解 DDPM 的核心原理对于掌握生成式 AI 的最新技术至关重要。

3 方法论

DDPM 的核心思想是使用形式为 $p_\theta(x_0) = \int p_\theta(x_{0:T}) dx_{1:T}$ 的潜在变量模型来建模数据分布 $p_\theta(x_0)$ ，其中 x_1, \dots, x_T 是与数据 $x_0 \sim q(x_0)$ 具有相同维度的潜在变量。联合分布 $p_\theta(x_{0:T})$ 定义为具有学习到的高斯转移的马尔可夫链，起始于 $p(x_T) = \mathcal{N}(x_T; \mathbf{0}, \mathbf{I})$ 。

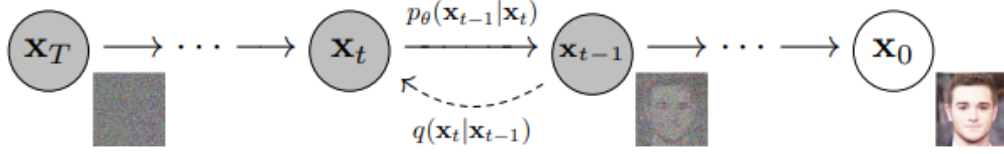


图 1: 去噪扩散概率模型 (DDPM) 的概率图模型。前向过程 q (从右到左) 逐渐添加高斯噪声，将数据 x_0 转换为噪声 x_T 。逆向过程 p_θ (从左到右) 学习逐步去噪以恢复数据。

3.1 前向扩散过程

前向过程 (记为 q) 是一个固定的马尔可夫链，它根据方差调度 β_1, \dots, β_T 逐渐向数据添加高斯噪声。给定数据点 x_0 ，前向过程定义为：

$$q(x_{1:T}|x_0) = \prod_{t=1}^T q(x_t|x_{t-1}) \quad (1)$$

其中转移核为：

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

这里， $\beta_t \in (0, 1)$ 控制每一步添加的噪声量。当 $T \rightarrow \infty$ 时， x_T 接近各向同性高斯分布。

该过程的一个关键性质是我们可以以封闭形式在任意时间步 t 对 x_t 进行采样。令 $\alpha_t = 1 - \beta_t$ 且 $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$ 。则：

$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}x_0, (1 - \bar{\alpha}_t)\mathbf{I}) \quad (3)$$

这允许我们将 x_t 表示为 x_0 和噪声变量 $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ 的线性组合：

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon \quad (4)$$

此属性对于高效训练至关重要，因为它避免了为了获得 x_t 而迭代采样整个链的需要。

3.2 逆向去噪过程

逆向过程定义为具有学习到的高斯转移的马尔可夫链，起始于 $p(x_T) = \mathcal{N}(x_T; \mathbf{0}, \mathbf{I})$ ：

$$p_\theta(x_{0:T}) = p(x_T) \prod_{t=1}^T p_\theta(x_{t-1}|x_t) \quad (5)$$

其中转移核为:

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)) \quad (6)$$

目标是学习参数 θ , 使得逆向过程能够逆转前向过程, 有效地对 x_t 进行去噪以恢复 x_{t-1} 。

作者选择将方差 $\Sigma_\theta(x_t, t)$ 固定为未经训练的时间相关常数, 具体为 $\sigma_t^2 \mathbf{I}$ 。讨论了两种选择: $\sigma_t^2 = \beta_t$ 和 $\sigma_t^2 = \tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t} \beta_t$ 。论文发现这两种选择产生的结果相似, 为了简单起见, 他们在主要实验中使用了 $\sigma_t^2 = \beta_t$ 。

均值 $\mu_\theta(x_t, t)$ 是主要的可学习组件。为了推导 μ_θ 的参数化, 我们观察后验 $q(x_{t-1}|x_t, x_0)$, 当以 x_0 为条件时, 它是易于处理的:

$$q(x_{t-1}|x_t, x_0) = \mathcal{N}(x_{t-1}; \tilde{\mu}_t(x_t, x_0), \tilde{\beta}_t \mathbf{I}) \quad (7)$$

其中

$$\tilde{\mu}_t(x_t, x_0) = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}x_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t \quad (8)$$

使用重参数化 $x_0 = \frac{1}{\sqrt{\bar{\alpha}_t}}(x_t - \sqrt{1-\bar{\alpha}_t}\epsilon)$, 我们可以用 x_t 和 ϵ 重写 $\tilde{\mu}_t$:

$$\tilde{\mu}_t(x_t, x_0) = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} \epsilon \right) \quad (9)$$

这表明我们应该参数化 $\mu_\theta(x_t, t)$ 来预测 $\tilde{\mu}_t$ 。由于 x_t 作为输入可用, 我们可以参数化模型来预测噪声 ϵ 。令 $\epsilon_\theta(x_t, t)$ 为旨在从 x_t 预测 ϵ 的函数逼近器。则:

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} \epsilon_\theta(x_t, t) \right) \quad (10)$$

这个公式将扩散模型与去噪分数匹配联系起来。

3.3 目标函数

训练是通过优化负对数似然的变分下界 (ELBO) 来进行的:

$$\mathbb{E}[-\log p_\theta(x_0)] \leq \mathbb{E}_q \left[-\log p(x_T) - \sum_{t \geq 1} \log \frac{p_\theta(x_{t-1}|x_t)}{q(x_t|x_{t-1})} \right] = L \quad (11)$$

损失 L 可以重写为 KL 散度项的总和:

$$L = L_T + L_{T-1} + \dots + L_0 \quad (12)$$

其中 L_T 是常数 (因为 $q(x_T|x_0)$ 固定为高斯), 其他项涉及高斯之间的 KL 散度。具体来说, 第 t 步的项是:

$$L_{t-1} = \mathbb{E}_q [D_{KL}(q(x_{t-1}|x_t, x_0) || p_\theta(x_{t-1}|x_t))] \quad (13)$$

代入均值的参数化, 作者表明最小化此 KL 散度等价于最小化真实噪声 ϵ 和预测噪声 $\epsilon_\theta(x_t, t)$ 之间的均方误差。Ho 等人提出的简化目标是:

$$L_{simple}(\theta) = \mathbb{E}_{t, x_0, \epsilon} [\|\epsilon - \epsilon_\theta(\sqrt{\alpha_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, t)\|^2] \quad (14)$$

其中 t 在 1 到 T 之间均匀分布。这种简化的损失丢弃了从 ELBO 推导出的加权系数，作者发现这对样本质量有益。它有效地降低了小 t 处的损失权重，使模型能够专注于较大 t 处更困难的去噪任务。

3.4 模型架构

函数 $\epsilon_\theta(x_t, t)$ 使用 U-Net 架构实现，类似于语义分割和非对齐图像到图像转换中使用的架构。该架构的主要特征包括：

- 带有 ResNet 块的 U-Net 主干。
- 在 16×16 分辨率级别的自注意力机制。
- 时间步 t 的正弦位置嵌入，添加到每个残差块中。
- 整个网络中使用组归一化 (Group normalization)。

模型将噪声图像 x_t 和时间嵌入作为输入，并输出与 x_t 形状相同的噪声预测。

3.5 实现细节

作者提供了他们在 CIFAR-10 上实验的具体超参数，这对于复现结果至关重要：

- **时间步 (T):** 扩散过程使用 $T = 1000$ 步。
- **噪声调度:** 方差调度 β_t 从 $\beta_1 = 10^{-4}$ 线性增加到 $\beta_T = 0.02$ 。这个特定的调度确保了逆向过程在每一步都有可控的方差，同时确保 x_T 足够接近标准高斯分布。
- **优化器:** 模型使用 Adam 优化器训练，学习率为 2×10^{-4} 。不使用权重衰减。
- **批量大小:** 训练使用的批量大小为 128。
- **训练步数:** 模型训练约 800k 步。
- **Dropout:** 在残差块上应用比率为 0.1 的 Dropout 以防止过拟合，特别是在像 CIFAR-10 这样的小型数据集上。
- **EMA:** 模型参数的指数移动平均 (EMA) 用于采样，衰减率为 0.9999。这是 GAN 和其他生成模型中常用的技术，用于稳定生成结果。

U-Net 架构基于 Wide ResNet。它由四个特征图分辨率 (32×32 到 4×4) 组成。每个分辨率级别有两个残差块。通道倍增器为 (1, 2, 2, 2)。时间嵌入是正弦嵌入（类似于 Transformer 位置嵌入），后跟一个 MLP，然后添加到每个残差块中。

3.6 训练与采样算法

Algorithm 1 训练

```

1: repeat
2:    $x_0 \sim q(x_0)$ 
3:    $t \sim \text{Uniform}(\{1, \dots, T\})$ 
4:    $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ 
5:   对  $\nabla_{\theta} \|\epsilon - \epsilon_{\theta}(\sqrt{\alpha_t}x_0 + \sqrt{1 - \alpha_t}\epsilon, t)\|^2$  执行梯度下降步骤
6: until converged

```

Algorithm 2 采样

```

1:  $x_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ 
2: for  $t = T, \dots, 1$  do
3:    $z \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$  如果  $t > 1$ , 否则  $z = \mathbf{0}$ 
4:    $x_{t-1} = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{1 - \alpha_t}{\sqrt{1 - \alpha_t}}\epsilon_{\theta}(x_t, t)) + \sigma_t z$ 
5: end for
6: return  $x_0$ 

```

4 实验与结果

作者在几个标准图像生成基准上评估了他们的模型，包括 CIFAR-10 (32×32)、LSUN Bedroom (256×256) 和 LSUN Church (256×256)。

4.1 样本质量

用于评估的主要指标是 Fréchet Inception Distance (FID)，它测量 Inception-V3 网络特征空间中生成图像分布与真实图像分布之间的距离。较低的 FID 分数表示更好的样本质量。他们还报告了 CIFAR-10 的 Inception Score (IS)。

在 CIFAR-10 上，提出的 DDPM 实现了 3.17 的 FID，这与当时最先进的 GAN（例如 StyleGAN2-ADA）具有竞争力，并且明显优于大多数基于似然的模型。Inception Score 为 9.46。

对于 LSUN 数据集，该模型生成了高质量、多样化的样本。作者指出，样本展示了细粒度的细节和连贯的全局结构，例如卧室中逼真的纹理和教堂中的建筑细节。

4.2 与其他模型的比较

论文提供了与其他生成模型的详细比较：

- **与 GAN 相比：**虽然 GAN 通常获得稍好的 FID 分数，但 DDPM 生成的样本具有相当的感知质量，且没有对抗性训练相关的训练不稳定和模式崩溃问题。
- **与 VAE 和 Flow 相比：**DDPM 在样本质量 (FID) 方面明显优于 VAE 和基于流

的模型。虽然 VAE 和 Flow 在采样方面更快，但 DDPM 证明了迭代优化可以带来卓越的视觉保真度。

4.3 渐进式编码与插值

作者还分析了模型学习到的潜在表示。由于前向过程破坏信息，逆向过程可以看作是逐步添加信息。

- **源到目的地插值:** 通过固定源噪声 x_T 并在两个随机噪声向量之间进行插值，模型生成了两幅图像之间的平滑过渡。这证实了模型已经学习到了有意义的潜在空间。
- **从粗到细的生成:** 作者可视化了生成过程，显示模型首先建立全局结构（低频分量），然后随着 $t \rightarrow 0$ 填充细节（高频分量）。
- **无损压缩:** 论文还讨论了扩散模型用于无损压缩的潜力。通过离散化数据并使用变分下界，他们表明扩散模型可以实现具有竞争力的比特率。

4.4 消融实验

作者进行了消融实验以证明其设计选择的合理性：

- **预测 ϵ vs. $\tilde{\mu}$:** 他们发现预测噪声 ϵ 比直接预测均值 $\tilde{\mu}$ 效果更好。
- **固定方差 vs. 学习方差:** 在他们的初步实验中，将方差 σ_t^2 固定为 β_t 效果良好。后来的工作（如 Improved DDPM）表明，学习方差可以进一步提高对数似然。
- **损失加权:** 简化损失 L_{simple} （ ϵ 上的未加权 MSE）在样本质量方面优于理论推导的加权损失，尽管它优化的目标与精确对数似然不同。

5 主要贡献

论文《Denoising Diffusion Probabilistic Models》对生成深度学习领域做出了几个关键贡献：

1. **利用扩散模型进行高质量合成:** 最重要的贡献是证明了扩散模型可以生成与 GAN 具有竞争力的高质量样本。在这项工作之前，扩散模型在理论上很有趣，但在视觉保真度方面实际上不如人意。
2. **与分数匹配的联系:** 作者建立了扩散概率模型与朗之万动力学去噪分数匹配之间清晰的理论联系。通过表明 ELBO 可以简化为加权去噪分数匹配目标，他们统一了两条不同的研究路线。

3. **简化的训练目标:** 简化损失函数 L_{simple} 的提出是一个关键的工程突破。通过丢弃从 ELBO 推导出的复杂加权项，他们发现了一个既易于实现又对样本质量更有效的训练目标。
4. **有效的架构设计:** 采用带有自注意力和时间嵌入的 U-Net 架构为扩散模型提供了强大的主干，这已成为后续研究的标准。
5. **渐进式编码视角:** 论文提供了自回归解码的新视角，表明扩散模型可以解释为广义自回归模型，其中预测的“顺序”不是空间的，而是时间的（沿着扩散步骤）。

6 个人启发与见解

阅读这篇论文为生成建模的本质提供了几个深刻的见解：

6.1 迭代优化的力量

与试图在单次（或几次）传递中生成复杂图像的 GAN 不同，扩散模型依赖于长串小的迭代更新。这模仿了人类通常创作艺术的方式——从粗略的草图开始，逐渐完善细节。这种“迭代优化”似乎是建模复杂分布的一种更稳健的方法，因为模型只需要在每一步解决一个简单的去噪任务，而不是一次性学习从噪声到数据的整个映射。

6.2 信噪比作为课程学习

扩散过程自然地为模型创建了一个课程。在高噪声水平（大 t ），模型学习粗略的全局结构（低频信息）。随着噪声水平降低（小 t ），模型专注于细节（高频信息）。这种隐式课程学习可能是扩散模型能够很好地扩展到高分辨率和多样化数据集的关键原因。

6.3 生成建模作为逆物理过程

与非平衡热力学的类比令人着迷。前向过程是一个增加熵的物理过程（扩散）。生成过程是这个物理过程的时间反转，有效地减少熵以从混乱中创造结构。这表明物理定律可以为设计机器学习模型提供强大的归纳偏置。

7 衍生想法与未来方向

基于论文中提出的概念，出现了几个未来探索的想法：

7.1 加速采样

DDPM 的主要缺点是采样速度慢，需要 T （例如 1000）次前向传递。

- **想法:** 我们能否学习一个“学生”模型，将多个扩散步骤蒸馏成单个步骤？这将结合扩散模型的高质量和 GAN 的速度。
- **想法:** 我们能否使用非马尔可夫逆向过程来跳过步骤，正如后来在 DDIM 中探索的那样？

7.2 条件生成与控制

该论文侧重于无条件生成。

- **想法:** 通过在类标签或文本嵌入上调节 U-Net（例如，通过交叉注意力），我们可以控制生成过程。这是现代文本到图像模型的基础。
- **想法:** 我们可以使用预训练分类器的梯度（分类器引导）来引导采样过程，以在生成的图像中强制执行特定属性。

7.3 其他领域的扩散

虽然论文侧重于图像，但迭代去噪的原理是通用的。

- **音频:** 波形可以建模为连续信号。扩散可用于高保真语音合成或音乐生成。
- **3D 点云:** 向 3D 坐标添加噪声并学习去噪可能是生成 3D 形状的有效方法。
- **离散数据 (文本):** 由于高斯噪声的连续性，将扩散应用于离散数据具有挑战性。开发离散扩散过程（例如，使用转移矩阵而不是高斯噪声）可能会解锁强大的文本生成模型。

7.4 潜在扩散

在像素空间中运行扩散过程计算成本高昂。

- **想法:** 我们可以首先使用 VAE 或 VQ-GAN 将图像压缩到低维潜在空间，然后在潜在空间中执行扩散。这将大大降低计算成本并允许更高分辨率的生成。

8 理论分析

为了充分理解 DDPM 的贡献，深入研究证明目标函数合理性的数学推导是有益的。

8.1 变分下界的推导

目标是最大化数据的对数似然 $\log p_\theta(x_0)$ 。由于边缘似然难以处理，我们最大化证据下界 (ELBO)：

$$\log p_\theta(x_0) = \log \int p_\theta(x_{0:T}) dx_{1:T} \quad (15)$$

$$= \log \int \frac{p_\theta(x_{0:T})}{q(x_{1:T}|x_0)} q(x_{1:T}|x_0) dx_{1:T} \quad (16)$$

$$\geq \mathbb{E}_{q(x_{1:T}|x_0)} \left[\log \frac{p_\theta(x_{0:T})}{q(x_{1:T}|x_0)} \right] \quad (17)$$

使用马尔可夫性质展开联合分布：

$$\text{ELBO} = \mathbb{E}_q \left[\log \frac{p(x_T) \prod_{t=1}^T p_\theta(x_{t-1}|x_t)}{\prod_{t=1}^T q(x_t|x_{t-1})} \right] \quad (18)$$

$$= \mathbb{E}_q \left[\log p(x_T) + \sum_{t=1}^T \log \frac{p_\theta(x_{t-1}|x_t)}{q(x_t|x_{t-1})} \right] \quad (19)$$

利用性质 $q(x_t|x_{t-1}) = q(x_t|x_{t-1}, x_0)$ 和贝叶斯规则 $q(x_t|x_{t-1}, x_0) = \frac{q(x_{t-1}|x_t, x_0)q(x_t|x_0)}{q(x_{t-1}|x_0)}$ ，我们可以重写这些项以对 KL 散度进行分组。

$$L = \mathbb{E}_q \left[D_{KL}(q(x_T|x_0)||p(x_T)) + \sum_{t=2}^T D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t)) - \log p_\theta(x_0|x_1) \right] \quad (20)$$

这种分解强调了训练目标包括将前向过程的后验 $q(x_{t-1}|x_t, x_0)$ 与逆向过程参数化 $p_\theta(x_{t-1}|x_t)$ 进行匹配。

8.2 与朗之万动力学的联系

朗之万动力学提供了一种使用分布 $p(x)$ 的分数函数 $\nabla_x \log p(x)$ 从中采样的方法。更新规则为：

$$x_{t-1} = x_t + \frac{\epsilon}{2} \nabla_x \log p(x_t) + \sqrt{\epsilon} z_t \quad (21)$$

将其与 DDPM 中的逆向过程更新进行比较：

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_\theta(x_t, t) \right) + \sigma_t z \quad (22)$$

我们可以看到 $\epsilon_\theta(x_t, t)$ 起着与分数函数 $-\nabla_{x_t} \log p(x_t)$ 成比例的作用。具体来说， $\epsilon_\theta(x_t, t) \approx -\sqrt{1 - \bar{\alpha}_t} \nabla_{x_t} \log p(x_t)$ 。这证实了学习预测噪声等价于学习不同噪声水平下数据分布的分数。

9 结论

总之，《Denoising Diffusion Probabilistic Models》是一篇具有里程碑意义的论文，它重新定义了生成建模的格局。通过成功地将概率扩散过程的理论优雅性与深度学习的实践力量相结合，Ho 等人提供了一个用于高质量图像合成的强大框架。

这篇论文的贡献不仅仅在于取得了良好的结果；它提供了一种将生成视为迭代去噪过程的新思维方式。这种观点已被证明是非常富有成效的，催生了大量的后续工作，这些工作在扩散模型的速度、质量和可控性方面进行了改进。

展望未来，扩散模型有望在 AI 创造力的进步中发挥核心作用，从生成逼真的图像到合成复杂的科学数据。从这篇论文中获得的见解无疑将在未来几年继续塑造该领域。