1. 扩散模型本质：参数化的马尔可夫链（可训练过程·），使用变分推断进行训练（加权变分下界上进行训练），这个链的转移被学习来逆转一个扩散过程，该扩散过程是一个逐步向数据中添加噪声的马尔可夫链，方向与采样方向相反，直到信号被破坏
2. 扩散模型前后向过程：扩散过程由少量高斯噪声组成，采样也为高斯分布，这让它的参数化很容易（网络结构很简洁）
3. 扩散模型缺点：无法生成高质量图像
4. DDPM核心贡献：证明验证扩散模型可以生成高质量的图像，展示了扩散模型的某种特定参数化揭示了它与训练过程中的多噪声水平去噪得分匹配以及采样过程中的退火朗之万动力学之间的等效性，很大一部分码长被用来刻画人眼几乎看不见的细节，模型的采样过程可以被视为一种渐进式解码，这类似于自回归模型的解码方式，但更加通用
5. 核心公式：

为原始数据

为最终要学习的数据分布。我们的目标是让模型能够生成看起来像是来自原始数据的新样本。这个分布是由模型参数 θ 决定的

对所有隐变量进行积分得到最终分布

这是一个联合分布，它描述了**逆向过程**，也就是从纯噪声逐步去噪恢复到原始图像的整个过程

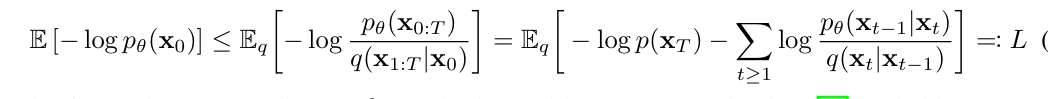
1. 逆向过程：一个马尔可夫链，从开始

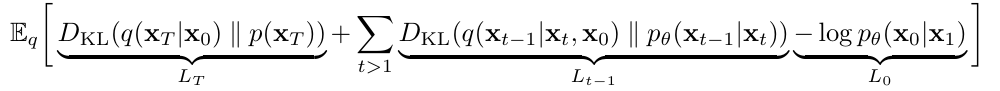
马尔可夫链：随机过程，未来状态只和现在状态有关

可学习：全过程都是可学习的高斯分布构成，神经网络的目的就是学习高斯分布参数的过程

7.前向过程：固定的马尔可夫链，向原始数据加高斯噪声（由预先制定的方差表决定，噪声由小到大）一个重要的特性是，由于前向过程是固定的高斯分布，我们可以直接计算任何时间步 t 的状态 xt​，而不需要一步一步地进行。这使得训练变得更加高效

8.训练过程：优化变分下界，训练模型的目的是让逆向过程尽可能地接近前向过程的逆过程，采用随机梯度下降法（因为变分下界 L 包含了所有时间步 t 的项，如果每次都计算所有项会非常耗时）





这个公式将复杂的对数似然项转换成了三个部分，其中最主要的部分是KL散度,这使得训练变得更加稳定。因为KL散度比较的两个分布都是高斯分布,它们的KL散度可以直接用公式计算出来。这避免了使用不稳定的蒙特卡洛方法，确保了训练的稳定性和高效性

9.如何通过独特的参数化方法来简化训练和得到更好地结果

本文建立了扩散模型和去噪得分匹配之间一个新的明确联系，这为扩散模型带来了一个简化的加权变分下界目标。

的处理：固定化，不采用训练方式，成了常数，简化了训练

的优化：

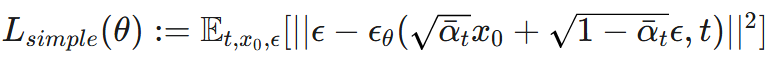
设置为与时间相关的、未经训练的常数

：对它进行特定的参数化，与其直接让神经网络去预测复杂的均值，不如让它去预测加在图像上的噪声 ϵ。这种预测噪声的参数化方式，是DDPM成功的关键之一，作者说类似于去噪得分匹配，将的训练目标与去噪得分匹配联系起来，进一步验证了这种“预测噪声”方法的合理性（它类似于朗之万动力学，它通过一个“梯度”来引导样本向高概率区域移动。作者发现，他们的采样过程与此非常相似，他们预测的噪声就扮演了“梯度”的角色）

预测 ϵ 的参数化方式，既类似于朗之万动力学，又将扩散模型的变分下界简化为一个类似于去噪得分匹配的目标

1. 理论上: 它与朗之万动力学有深层联系，使得采样过程有坚实的理论基础。
2. 训练上: 它将复杂的训练目标简化为一个易于优化的目标。
3. 结果上: 最终的实验结果证明了这种方法的有效性。

10.简化训练目标



这就是作者最终使用的简化训练目标。这个公式非常简洁：它计算了模型预测的噪声与真实噪声 ϵ 之间的均方误差（MSE），并希望这个误差尽可能小。他们选择的 βt​ 使得在 t 较小时（即图像只添加了很少噪声），损失项的数值本身就很小，从而自然地降低了这些项的权重。这些项训练网络去噪只含微量噪声的数据，因此降低它们的权重是有益的，这样网络就可以专注于在更大的 t 项上更困难的去噪任务

11. 数据缩放、逆向过程解码器和

作者在训练前对图像数据进行了预处理，将其像素值从 [0, 255] 的整数范围缩放到 [-1, 1] 的浮点数范围，这有助于神经网络的稳定训练

虽然最后一步有理论上的噪声，但在实际生成最终图像时，作者选择直接使用均值作为最终的像素值，以获得更清晰的结果

12.实验过程

架构选择：U-Net

为什么选择这么小的 βt​ 值：当噪声很小时，逆向和前向过程的数学形式会非常相似，这有助于模型更容易地学习

参数处理：模型的去噪网络是共享参数的，即同一个网络被用于所有的去噪步骤。

时间嵌入：为了让网络知道它当前处于哪个时间步（比如是第1步去噪还是第500步去噪），作者借鉴了Transformer模型中的正弦位置嵌入技术将时间信息编码后作为输入的一部分

13.相关工作：

q 没有参数: 这是一个关键区别。流模型和VAE通常需要训练一个“编码器”来学习前向过程，而扩散模型的前向过程是固定的

互信息几乎为零: 这意味着当图像完全变成噪声时，它与原始图像之间几乎没有任何信息关联。这使得模型的起点（纯噪声）非常明确，有助于稳定训练

通过独特的参数化，他们不仅将扩散模型与去噪得分匹配联系起来，也与一种名为退火朗之万动力学的采样方法建立了联系

直接评估对数似然: 这意味着我们可以用一个明确的数学指标来衡量模型的拟合能力

明确训练采样器: 扩散模型的训练目标直接优化了它的采样过程，而其他一些模型则是在训练后额外添加采样器

14.结论

与多个领域产生了联系，扩散模型在图像领域表现出色，未来可以应用于其他类型的数据（如音频、文本）或作为更复杂系统的一部分