

实验流程

实验一：样本质量评估实验

- 实验数据集：
 - CIFAR10 (32×32)
 - CelebA-HQ (256×256)
 - LSUN (256×256)
- 实验设置：
 - 前向过程总步数设为 $T = 1000$ 。
 - 前向过程方差 β_t 采用线性增加方式：从 $\beta_1 = 10^{-4}$ 到 $\beta_T = 0.02$ 。
 - 模型采用U-Net结构，带有自注意力（self-attention）机制。
 - 所有实验使用 EMA（指数滑动平均）方式更新模型参数。

这个地方把参数做成表格，然后放几个图片介绍数据集

CIFAR-10 是机器学习领域中经典的小型图像分类数据集之一，包含 60,000 张分辨率为 32×32 像素的彩色图像，共 10 个类别

CelebA-HQ 全称是 CelebFaces Attributes High-Quality Dataset，由 NVIDIA 基于原始 CelebA 数据集制作。

包含了 30,000 张 高质量的、分辨率为 1024×1024 像素的高清人脸图像。

图像来自名人面部照片，具有较高的多样性，包括了丰富的面部表情、姿势、配饰、背景等。分辨率极高，细节丰富，适合训练生成高质量、高分辨率人脸图像的生成模型。

LSUN 全称 Large-scale Scene Understanding Dataset，由普林斯顿大学提出。

包含了大量高质量、现实场景的图像，主要用于场景理解任务。

数据集中含多个场景类别，比如卧室（Bedroom）、教堂（Church）、厨房（Kitchen）、猫（Cat）等。

每个类别的图片数量巨大，例如：

LSUN Bedroom 类别超过 300 万张图像。

LSUN Church 类别也超过 12 万张图像。

论文采用了生成模型领域广泛使用的几个经典评价标准：

1. Inception Score (IS):

衡量生成图像的多样性与真实性，值越高代表生成图像质量越好。

2. FID (Fréchet Inception Distance):

用于衡量生成数据与真实数据的距离，越小表示生成质量越高。

3. 负对数似然 (Negative Log Likelihood, NLL) :

衡量模型对真实数据分布的拟合能力，单位为 bits/dim，越低表示模型拟合效果越好。

(1) 样本生成质量

在无条件CIFAR10数据集上：

模型	Inception Score↑	FID↓	NLL↓ (bits/dim)
Gated PixelCNN	4.60	65.93	3.03
PixelIQN	5.29	49.46	-
EBM	6.78	38.2	-
NCSNv2	-	31.75	-
SNGAN	8.22	21.7	-
StyleGAN2+ADA	9.74	3.26	-
本论文提出扩散模型(Lsimple)	9.46	3.17	≤3.75

- 论文提出的模型取得了非常优秀的FID分数（3.17），超过了大部分的基线模型，包括GAN类和Autoregressive类模型，表现卓越。

(2) 负对数似然(NLL)

- 扩散模型在负对数似然方面表现（≤3.75 bits/dim）优于传统的能量模型（EBM和NCSN），但不如Autoregressive模型（如PixelCNN等）。
- 作者指出扩散模型有较强的压缩能力（更多的编码能力被用于描述难以察觉的细节），从而对样本质量有更好的控制。

实验二

论文在理论部分（第3节）提出了一种特殊的逆过程参数化方式，即网络预测前向过程中的噪声 ϵ ，并提出了一个更简单的训练目标函数 L_{simple} 。作者希望明确：

- **逆过程参数化方式**（网络预测 $\hat{\mu}$ 还是预测噪声 ϵ ）对模型性能有多大影响？
- 采用**简化目标函数**（ L_{simple} ）相较于原始目标函数（Variational Bound L ）是否能真正提升生成质量？
- 逆过程中的**方差**（ Σ ）是学习获得更好效果，还是固定方差能得到更稳定、更好的结果？

🔥 实验设计与流程：

本实验围绕以下几个因素进行对比：

1. **预测目标**（逆过程参数化）：
 - 预测前向过程后验均值 $\tilde{\mu}$ （传统方法）
 - 预测添加到图像中的噪声 ϵ （作者提出的方法）
2. **训练目标函数**：
 - 标准的变分界目标函数（记作 L ）
 - 简化后的目标函数（记作 L_{simple} ）
3. **逆过程方差设置** (Σ)：
 - 方差可学习（learned diagonal Σ ）
 - 方差固定（fixed isotropic Σ ）

因此，本实验综合考虑了三个变量，共设置了几个不同的实验条件进行组合验证：

参数化方式	目标函数	方差类型
预测 $\tilde{\mu}$	L	方差可学习
预测 $\tilde{\mu}$	L	方差固定
预测 ϵ (作者提出)	L	方差可学习
预测 ϵ (作者提出)	L	方差固定
预测 ϵ (作者提出)	Lsimple	方差固定

实验的评估指标为：

- **FID（越低越好）** 和 **Inception Score (IS, 越高越好)**。

作者在 CIFAR-10 数据集上做了上述的对比实验，具体结果如下：

参数化方式	目标函数	方差设置	IS (↑)	FID (↓)	稳定性
预测 $\tilde{\mu}$	L	学习的方差 Σ	7.28 ± 0.10	23.69	较差
预测 $\tilde{\mu}$	L	固定方差 Σ	8.06 ± 0.09	13.22	一般
预测 ϵ	L	学习的方差 Σ	不稳定	不稳定	很差（难训练）
预测 ϵ	L	固定方差 Σ	7.67 ± 0.13	13.51	较好
预测 ϵ	Lsimple	固定方差 Σ	9.46 ± 0.11	3.17	最好（稳定高效）

从上述结果可以得出：

- 预测噪声 ϵ 比预测 $\tilde{\mu}$ 更适合扩散模型（FID明显更低，IS更高）。
- 学习方差 Σ 很不稳定，甚至可能无法成功训练。
- 固定方差 Σ 可以提高模型的稳定性与生成质量。
- 使用作者提出的简化训练目标函数 L_{simple} ，不仅训练更简单，而且模型表现显著提高，达到最佳效果。

通过 4.2 节的消融实验，作者明确了最佳的扩散模型训练方案：

“逆过程预测噪声 ϵ + 固定方差 Σ + 简化目标函数 L_{simple} ”

4.3

(3) 模型参数化方式与训练目标的影响

作者还进行了对比实验，分析逆过程参数化方式对模型性能的影响：

- 如果网络直接预测前向过程的均值 $\tilde{\mu}$ ：
 - 训练不够稳定，样本质量一般。
- 如果网络预测噪声 ϵ （作者提出的方法）：
 - 使用简化目标函数 L_{simple} 时，FID 和 IS 达到最佳效果。

实验结果验证了作者提出的预测噪声 ϵ 和简化目标函数 L_{simple} 的有效性。

(4) 压缩性能

扩散模型具备逐步的“损失压缩”的特性：

- 将扩散过程看作逐步压缩数据的过程。
- 逐步去除噪声的过程类似逐步解码（逐步还原数据）。
- 作者给出了 CIFAR10 数据集的压缩比特率-失真曲线，展示扩散模型的压缩能力。

实验结果显示：

- 扩散模型大部分比特用于编码人眼难以察觉的细节，这体现出极强的压缩能力。

(5) 图像插值能力

扩散模型能够在图像的潜空间进行插值，实现平滑过渡和属性的渐进变化：

- 在 CelebA-HQ 数据集上的插值实验显示，插值效果非常自然和平滑，能够很好地表达图像间的过渡特征。