

去噪扩散概率模型

摘要

该论文提出使用去噪扩散概率模型 (Denoising Diffusion Probabilistic Models, DDPM) 进行高质量的图像合成，该一类生成模型的灵感来自非平衡热力学，通过模拟随机漫步的过程来生成数据，核心思想是通过逐步引入噪声，将简单分布演变为复杂分布，从而生成高质量的样本。最佳结果是通过在加权变分界上进行训练获得的，该变分界是根据扩散概率模型和与 Langevin 动力学匹配的去噪分数之间的新联系设计的，并且模型允许自然地渐进有损解压缩方案，可以解释为自回归解码的扩展。通过复现 DDPM，学习包括处理概率分布、采样和训练深度生成模型相关知识，有助于在科学应用中应用，并可以验证原始论文中的结果，确保其可重现性。我们在此基础上进一步扩展和改进该模型，添加了新的框架设计和替代的损失函数，使模型达到了更好的生成结果。

关键词：Diffusion Model；分数匹配；马尔可夫链

1 引言

人工智能 (AI) 生成领域近年来取得了显著的进展，涵盖了多个子领域，包括生成对抗网络 [5, 7, 15, 20, 25, 26, 31, 37] (Generative Adversarial Networks, GANs)、自动编码器 [12, 21–23, 27, 32, 36] (AutoEncoders, AEs)、变分自编码器 [2, 6, 8, 14, 33, 41] (Variational Autoencoders, VAEs) 等，这些技术的不断发展推动了计算机视觉、自然语言处理、图像处理等领域的前沿研究，为各种应用场景带来了创新和突破。

Diffusion Models [10] (扩散模型，即 DDPM) 是生成模型领域的一类模型，其核心思想是通过模拟随机漫步的过程，将简单的分布演变为复杂的分布，从而生成高质量的样本，这一类模型在生成对抗网络和变分自编码器等生成模型之外提供了一种不同的建模方式。在深度学习和生成模型领域，研究者们一直在寻求更加稳定、高效、高分辨率的生成模型，Diffusion Models 提供了一种全新的生成思路，其与传统生成模型不同的采样方式和模型结构使得它在一些任务上具有独特的优势。

DDPM 模型采用了与传统生成模型不同的生成机制，通过模拟漫步过程生成数据，这种方法的创新性使得它能够应对一些传统生成模型难以解决的问题，如高分辨率图像的生成，并且它建立在马尔可夫链 [1] 数学思想的基础上，通过梯度信息进行训练，具有较为坚实的理论基础，这一点使得模型具有强大的可解释性，有助于深入研究其性能和特性。因为 Diffusion Models 通常能够生成高质量、高分辨率的样本，这使得其在图像生成等任务上具有潜在的优势，这对于一些对样本质量要求较高的应用场景（如医学图像生成 [39]、艺术创作 [19] 等）具有相当强烈的吸引力。

2 相关工作

各种深度生成模型最近在各种数据模式下展示了高质量的样本。生成对抗网络 (GANs)、自回归模型 [18] 和变分自编码器 (VAEs) 已经合成了引人注目的图像和音频样本，并且目前在基于能量和分数匹配的模型方面取得了显著进展，已经产生了与 GANs 相当的图像。

生成对抗网络 (GANs) 作为一种强大的生成模型，通过训练生成器和判别器的对抗过程，使得生成器能够逐渐生成逼真的样本，GANs 在图像合成领域取得了巨大成功，能够生成细致逼真的图像，甚至在艺术创作和风格转换 [17] 等方面也展现了惊人的效果，然而，GANs 仍然面临一些问题，如训练的不稳定性和模式崩溃等，这使得研究者们寻求其他生成模型的解决方案。

自回归模型是另一类生成模型，它通过对数据进行建模，从而能够生成符合数据分布的样本，这类模型的代表包括 PixelRNN 和 PixelCNN 等，它们通过考虑样本的每个元素的条件概率来生成图像。自回归模型在某些任务上取得了很好的效果，特别是对于具有明显结构的数据，如自然语言处理中的文本生成 [29]，然而，由于自回归模型的串行生成过程，其在生成大型高分辨率图像时往往面临计算效率和模型复杂度的挑战。

变分自编码器 (VAEs) 结合了自编码器和变分推断的思想，通过学习数据的潜在分布，实现了更高效的生成过程。VAEs 在图像生成和数据降维等任务上取得了显著的成果，通过引入潜在变量，VAEs 能够在生成样本的同时学习数据的表征，具有一定的生成样本多样性。然而，一些研究者 [28] 指出，VAEs 在生成过程中可能出现模糊的样本，这成为其改进的研究方向之一。

最近，基于能量 [3, 13, 30, 35, 38, 40] 和分数匹配 [4, 11, 16, 24] 的模型吸引了越来越多的关注，生成结果图 1 所示。这类模型通过考虑能量函数和分数函数，实现了更加稳定的训练过程，产生了与 GANs 相当水平的图像，由于这些模型具有更好的训练稳定性，一些研究者认为它们可能成为未来生成模型的主要方向之一。

扩散模型是一种在生成模型领域中备受关注的模型，其独特的训练机制和能够生成高质量样本的特性引起了广泛关注。在该论文中，研究者们提出并证明了扩散模型实际上能够生成高质量的样本，并且在某些情况下表现优于其他类型的生成模型，本文将深入探讨扩散模型的定义、训练效率、样本生成能力以及与其他生成模型的比较。

扩散模型的定义相对简单，其核心思想是通过引入噪声并通过逐步去噪的方式来学习数据的概率分布。这种逐步去噪的过程使得扩散模型能够更好地捕捉数据的复杂结构。相比于其他生成模型，扩散模型的训练效率较高，这一特性使得它在大规模数据集上能够更快地收敛，为生成高质量样本提供了有力支持。关于样本生成能力，该论文强调了扩散模型实际上能够生成高质量的样本，并通过实验证据证明了这一点，这使得扩散模型在实际应用中具有广泛的潜力，尤其在需要高质量样本的任务中。另外，扩散模型的某种参数化揭示了在训练过程中多个噪声水平上的去噪分数匹配和采样过程中的等价性，这一发现可能为理解扩散模型的训练机制提供了新的角度，同时也为改进和拓展该模型打开了新的研究方向。与其他基于似然的模型相比，扩散模型在对数似然方面表现优于基于能量的模型和分数匹配产生的采样。



图 1. 基于能量和分数匹配的模型在 CelebA-HQ 数据集生成的样本

3 本文方法概述

3.1 总体框架

扩散模型是一类用于建模数据生成过程的概率模型，其主要思想是通过一系列逐步传播的扩散步骤来生成数据，其中每个步骤都引入一些随机性，这一过程的逆过程可以被用于估计模型参数或执行生成任务，扩散模型总体框架如图 2 所示。模型分为两个过程，称之为前向过程和后向过程，前向过程为马尔科夫链过程，通过一系列扩散步骤，逐步对输入图像 x_0 添加高斯噪音 ϵ ，将初始样本逐渐传播到更复杂的分布，最终生成 x_T 。每个扩散步骤都引入一些随机性，使得数据在空间中进行逐步的、随机的漫步。后向过程为马尔科夫链逆过程，根据给定生成的数据样本，逐步逆向扩散，还原到初始样本的分布，利用逆向扩散过程中的信息，使用深度来估计模型的参数，使得逆向生成的过程更好地逼近实际数据生成过程，这个过程需要考虑扩散步骤中引入的随机性。

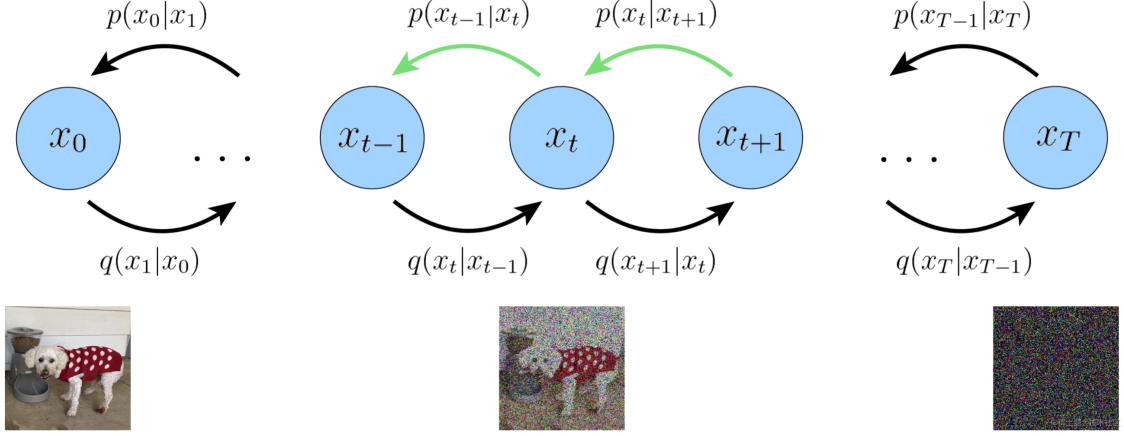


图 2. 扩散模型总体框架

3.2 重参数技巧

高斯分布的重参数技巧 [34] 是一种用于通过参数化的采样过程来近似采样高斯分布的方法，以便在梯度下降中对分布参数进行有效的反向传播。这个技巧在训练深度生成模型，如变分自编码器（VAE）等中得到了广泛的应用。假设我们有一个标准正态分布 $\mathcal{N}(0, 1)$ ，想要生成一个均值为 μ 、标准差为 σ 的一般正态分布 $\mathcal{N}(\mu, \sigma^2)$ 的样本。标准正态分布的采样可以通过使用随机噪声 ϵ 与均值为零、标准差为一的正态分布进行缩放和平移来实现，即：

$$z = \mu + \sigma \cdot \epsilon \quad (1)$$

其中， ϵ 是从标准正态分布中采样的随机噪声。这个过程可以用数学公式表示为：

$$z = g(\mu, \sigma, \epsilon) = \mu + \sigma \cdot \epsilon \quad (2)$$

这里， g 表示重参数化函数。使用重参数技巧后，可以将采样过程推导为对分布参数的微分，这对于梯度下降和反向传播非常有用。具体而言，对于生成样本 z 关于分布参数 μ 和 σ 的梯度可以通过链式法则计算：

$$\nabla_{\mu, \sigma} z = \nabla_{\mu, \sigma} g(\mu, \sigma, \epsilon) = \nabla_{\mu, \sigma} (\mu + \sigma \cdot \epsilon) \quad (3)$$

在训练生成模型时，可以通过计算这些梯度来更新模型参数，从而使生成的样本更好地逼近目标分布。

3.3 马尔科夫链

马尔科夫链是一种随机过程，具有“无记忆”的性质，即下一步的状态仅依赖于当前的状态，而不依赖于过去的状态。数学上，马尔科夫链的性质可以用以下概率公式表示：

$$P(X_{t+1} = j | X_t = i) = P_{ij} \quad (4)$$

其中， P_{ij} 表示从状态 i 转移到状态 j 的转移概率， $P(X_0 = i)$ 为初始状态为 i 的概率，即在时刻 $t = 0$ 的状态的概率分布。 $P(X_{t+1} = j | X_t = i)$ 为在给定当前状态 i 的条件下，下

一时刻状态为 j 的概率。马尔科夫链的关键性质是马尔科夫性，即在给定当前状态的情况下，未来的状态只依赖于当前状态，与过去的状态无关。在矩阵形式中，可以用状态转移矩阵 P 来表示：

$$P = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1n} \\ P_{21} & P_{22} & \cdots & P_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{n1} & P_{n2} & \cdots & P_{nn} \end{bmatrix} \quad (5)$$

其中， P_{ij} 表示第 i 个状态到第 j 个状态的转移概率。这样，给定初始分布和状态转移矩阵，就可以通过迭代计算得到马尔科夫链在任意时刻的状态分布。

3.4 前向过程

扩散模型的前向过程涉及数据生成的逐步传播过程。对于初始样本 x_0 通过一系列逐步的扩散步骤，将初始样本逐渐传播到更复杂的分布。每个扩散步骤都引入一些随机性，通常使用一个噪声项表示。假设在第 t 步，引入的随机噪声为 $\epsilon_t \sim \mathcal{N}(0, \mathbf{I})$ 。逐步扩散的数学表达可以表示为：

$$x_t = \sqrt{\alpha_t} x_{t-1} + \sqrt{1 - \alpha_t} \cdot \epsilon_t \quad (6)$$

其中， α 是控制扩散的参数。经过一定数量的扩散步骤后，生成最终的数据样本 x_T 。这可以通过对逐步扩散公式的迭代得到：

$$x_T = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t} \cdot \bar{\epsilon}_t \quad (7)$$

其中， $\bar{\alpha}_t = \prod_{i=1}^T \alpha_i$, $\bar{\epsilon}_t \sim \mathcal{N}(0, \mathbf{I})$ 。

3.5 后向过程

扩散模型的后向过程涉及从生成的数据样本逆向推导出初始样本的过程。这一逆向推导涉及到对逐步扩散的逆过程，需要考虑噪声的逆过程以及参数的逆过程。逆向扩散的核心是逆向地考虑扩散过程，从最终的数据样本 x_T 逐步逆向到初始样本 x_0 。如果我们能够逐步得到逆转后的分布 $q(x_{t-1}|x_t)$ ，就可以从完全的标准高斯分布 $x_T \sim \mathcal{N}(0, \mathbf{I})$ 还原出原图分布 x_0 。我们无法简单推断 $q(x_{t-1}|x_t)$ ，因此我们使用深度学习模型去预测这样的一个逆向的分布 p_θ ：

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, x_0), \sum_\theta(x_t, t)) \quad (8)$$

虽然无法得到逆转后的分布 $q(x_{t-1}|x_t)$ ，但是如果知道 x_0 ，可以通过贝叶斯公式得出：

$$q(x_{t-1}|x_t, x_0) = \mathcal{N}(x_{t-1}; \tilde{\mu}_\theta(x_t, x_0), \tilde{\beta}_t \mathbf{I}) \quad (9)$$

我们可以得到 (9) 中的方差和均值为：

$$\tilde{\beta}_t = \frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t} \cdot \beta_t \quad (10)$$

$$\tilde{\mu}_t(x_t, x_0) = \frac{\sqrt{\alpha_t}(1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t}x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1 - \bar{\alpha}_t}x_0 \quad (11)$$

如果我们知道基于高斯数据分布的 x_{t-1} 的均值和方差，根据重参数技巧，我们可以得出 x_{t-1} 数据表示，如此依次迭代，可以得出最终的 x_0 。

并且，根据公式 (7)，可以得到：

$$\tilde{\mu}_t = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{1 - \bar{\alpha}_t}}\bar{\epsilon}_t) \quad (12)$$

其中高斯分布 $\bar{\epsilon}'_t$ 为深度模型所预测的噪声，因此通过对真实噪音 $\bar{\epsilon}_t$ 来更新深度模型参数。

$$Loss = E_{x_0, \bar{\epsilon}_t}(||\bar{\epsilon}_t - \bar{\epsilon}'_t||^2) \quad (13)$$

$$\bar{\epsilon}'_t = \Phi(x_t) \quad (14)$$

其中 $\Phi(\cdot)$ 为深度模型表现形式。

4 复现细节

4.1 与已有开源代码对比

在复现扩散模型的过程中，我们着眼于提高模型性能、迁移 Pytorch 平台以及提高代码的可维护性。原文代码基于 TensorFlow 框架，使用 Cloud TPU 服务器，具有很大的迁移使用难度，为了克服这些限制，进行了一系列的修改和迁移工作，以确保代码的稳健性和灵活性，使之能够匹配 GPU 服务器和 Pytorch 平台，此外，部分生成结果基于开源框架 stable-diffusion-webui [9] 进行修改。

在复现扩散模型的过程中，我们通过深入研究原文生成模型构建原理，决定进行模型架构的修改，以进一步改进模型的性能和适应性。我们对扩散模型的编码器和解码器部分进行了修改，通过引入新的神经网络层、增加模型的深度或宽度，以及改变激活函数等方式来优化编码器和解码器的架构，这一调整旨在增加模型对输入数据的表达能力，从而提高生成样本的质量。扩散模型的核心在于通过逐步迭代的扩散步骤来生成样本，我们对原文代码中的扩散步骤的参数设置、噪声引入方式等进行了调整，以达到更好的生成效果，这包括对扩散步骤的数量、噪声水平、扩散过程的可解释性等方面的改进。扩散模型通常包含一个潜在空间，其中表示了生成样本的特征，通过减少潜在空间的维度，改变潜在变量之间的关系，来调整模型的表示能力和生成样本的多样性。

在进行这些模型架构的修改时，应该特别注重训练参数对模型的训练过程的影响。这包括对优化器的选择、学习率的调整以及对损失函数的改进等方面的考虑，通过重新设计损失函数，来提高模型的收敛速度和生成样本的质量，这种改变涉及到对模型的数学推导和实验结果的重新验证，确保修改后的模型在性能上有所提升。我们确保模型在新的架构下能够更快、更稳定地收敛，同时保持对训练数据的良好拟合。值得注意的是，对于模型架构的修改并非一成不变的过程。我们可能进行了多次实验和调整，以找到最优的模型配置，这一过程是迭代的，需要不断地观察模型在训练和测试集上的表现，并根据反馈进行有针对性的改进。

4.2 实验环境搭建

在进行扩散模型的复现代码过程中，我们注重搭建一个稳定而高效的实验环境，以确保代码的顺利运行和实验结果的可重现性。我们深度模型平台为适用于 CUDA 的 Python3.9.1，以充分利用 GPU 加速。Pytorch 版本为 1.13.1，Cuda 版本为 11.6.0，使用版本控制工具（Git）来跟踪和管理代码的不同版本，以便在需要时回滚到先前的稳定版本，着重创建一个清晰、高效、可重现的实验环境，以便能够准确地复现扩散模型的代码并进行后续的实验和分析。

4.3 创新点

4.3.1 新的模型框架

我们设计了一种新型的模型框架，如图 3 所示。新型模型框架在设计上经过思考和创新，与原论文中的深度模型有了很大的区别。模型中选择将生成的目标定位在输入图像上，与原论文中基于高斯分布的噪音 $\bar{\epsilon}_t$ 的生成目标有了本质的不同，这一设计决策带来了许多优势，使得模型在去噪任务中展现出更为出色的性能。通过将生成目标直接定位在输入图像上，避免了随机噪音带来的不确定性和不稳定性，在传统的深度模型中，噪音的引入可能会导致生成结果的随机性，使得模型的训练和生成过程变得难以控制。相比之下，我们以输入图像为目标，减少了生成过程中的随机性，使得模型的输出更加可控和稳定。由于模型生成目标为输入图像，因此在去噪任务中具有更好的重建质量，原文在处理噪声时可能会受到噪音分布的影响，导致生成的图像质量不稳定，而我们直接以清晰的输入图像为目标，更加专注于学习图像的内在特征和结构，从而在去噪过程中实现更高质量的重建。此外，新模型还表现出更强的鲁棒性，通过将目标定位在输入图像上，使模型更加适应真实场景中的各种噪声和复杂结构，这种鲁棒性使得在处理各种实际应用场景中都能够表现出色，包括但不限于图像去噪、图像修复等任务。

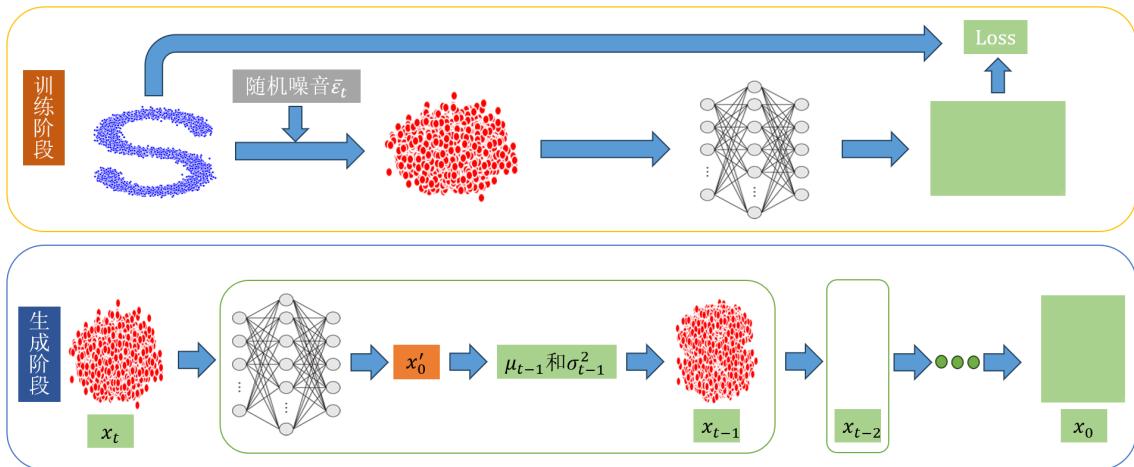


图 3. 扩散模型框架

4.3.2 新的损失函数

根据公式 (11)，我们可以得知 x_{t-1} 的均值和方差与 x_0 、 x_t 有关。在已知 x_t 的情况下，选择深度模型去拟合 x_0 ，同样可以得出基于高斯分布的数据 x_{t-1} 的均值和方差，并且有望消

除随机噪音 $\bar{\epsilon}_t$ 的影响。在深度模型的训练过程中，利用已知的 x_t 信息，通过模型学习 x_0 的分布特征。这使得能够在一定程度上还原数据生成的真实分布，进而推断 x_{t-1} 的均值和方差，通过这种方式，可以在去噪任务中获得更加准确和可靠的数据还原结果，消除了随机噪音的影响，提高了对数据生成过程的理解，从而在去噪任务中取得更好的效果。

因此深度模型损失函数可以设计为：

$$Loss = E_{x_0}(\|x_0 - \Phi(x_t)\|^2) \quad (15)$$

我们可以看到深度模型目标为原有数据，和之前的生成模型（如 Gans）目标相同，因此之前的生成模型可以在原有的基础上加上扩散过程进行生成，以达到更好的效果。

4.4 其它探索

4.4.1 探索 1：直接生成 x_{t-1}

根据公式 (6)，可以得到：

$$x_{t-1} = \frac{x_t}{\sqrt{\alpha_t}} - \frac{\sqrt{1 - \alpha_t}}{\sqrt{\alpha_t}} \epsilon_t \quad (16)$$

已知 ϵ_t 和深度模型拟合的 $\bar{\epsilon}_t'$ 都属于标准高斯分布噪音，是否可以将 ϵ_t 用 $\bar{\epsilon}_t'$ 代替，直接生成 x_{t-1} ，而不是通过 x_{t-1} 的均值和方差生成 x_{t-1} ？根据这一思路，实验结果如图 4 所示。

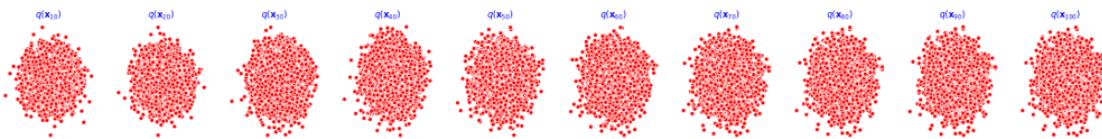


图 4. 探索 1 实验结果。原数据为呈现“S”型的图片，每 10 次迭代输出生成结果。

可以看到，我们无法根据上述思路生成原图片，进行合理性推测，认为 ϵ_t 虽然和 $\bar{\epsilon}_t'$ 都属于标准高斯分布，但是数据表现形式不同，无法进行替代。

4.4.2 探索 2：不使用深度模型生成 x_{t-1}

根据重参数技巧，我们可以通过知道高斯分布 x_{t-1} 的均值和方差生成 x_{t-1} 。根据公式 (12)，我们知道均值与 x_t 、 $\bar{\epsilon}_t$ 有关，那么如果 x_t 、 $\bar{\epsilon}_t$ 已知，能否不使用深度模型，根据 x_t 、 $\bar{\epsilon}_t$ 得出 x_{t-1} 的均值和方差，从而迭代生成最终的 x_0 呢？根据这一思路，实验结果如图 5 所示。

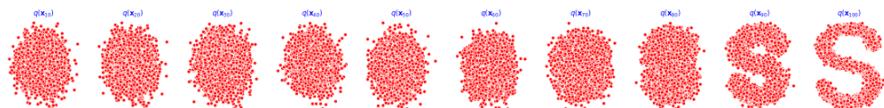


图 5. 探索 2 实验结果

可以看到，即使没有深度模型，我们仍然可以生成原图片，但是现实中 $\bar{\epsilon}_t$ 是未知的，我们需要运用深度模型来逼近 $\bar{\epsilon}_t$ ，或者采用其它方法。

4.4.3 探索 3：使用随机 x_t

根据探索 2 我们知道，根据 x_t 、 $\bar{\epsilon}_t$ 我们可以生成 x_0 ，那么如果我们就不采用根据 x_0 添加噪音生成的 x_t ，而是随机生成的杂乱无序的 x_t ，我们是否可以生成原来的 x_0 呢？根据这一思路，实验结果如图 6 所示。

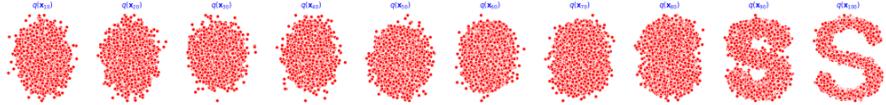


图 6. 探索 3 实验结果

由此我们可以得出结论，扩散模型生成结果和 $\bar{\epsilon}_t$ 有关，我们通过调控 $\bar{\epsilon}_t$ 来控制模型生成结果和方向。

5 实验结果分析

表 1 中展示了在 CIFAR10 数据集上的 Inception 分数 (IS)、FID 分数和负对数似然 (NLL) 等评价指标。原始 DDPM 在 FID 得分上表现为 3.17，在样本质量方面相较于文献中的许多模型具有更好的性能，而我们改进后的模型则取得了更为显著的提升。

值得注意的是，原始 DDPM 的 FID 分数是基于训练集计算的，这是一种标准做法。在相对于测试集计算时，FID 分数为 5.24。尽管相对于训练集来说有所下降，但仍然优于文献中许多训练集 FID 分数的表现。原文中的研究发现，在真实变分界上训练模型相较于在简化目标上训练，产生了更好的码长，然而，后者在样本质量上表现更为出色。CIFAR10 数据集生成的结果如图 7 所示，而 LSUN 数据集生成的结果则展示在图 8 中。

表 1. CIFAR10 实验结果

	IS	FID	NLL
Gated PixelCNN	4.60	65.93	3.03
Sparse Transformer			2.80
PixelIQN	5.29	49.46	
EBM	6.78	38.2	
NCSNv2		31.75	
NCSN	8.87	25.32	
SNGAN	8.22	21.70	
SNGAN-DDLS	9.09	15.42	
StyleGAN2+ADA(v1)	9.74	3.26	
DDPM	9.46	3.17	≤ 3.70
Ours	9.61	3.11	



图 7. CIFAR10 数据集生成结果, 左 FID=10.81, 中 FID=5.61, 右 FID=3.11

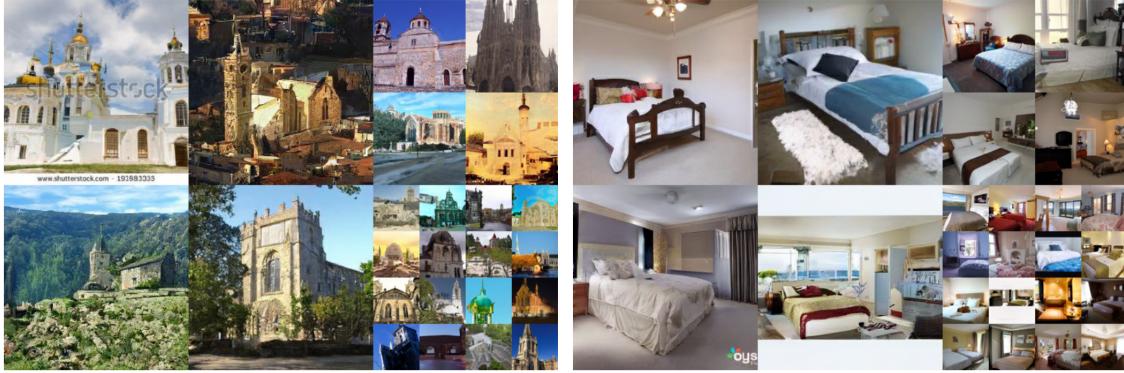


图 8. LSUN 数据集生成结果, 左 FID=7.89, 右 FID=4.90

我们进行了一个渐进的无条件生成实验, 该实验通过逐步解压随机比特来模拟逆向过程, 即 $x_t \rightarrow x_0$ 。简而言之, 我们预测了后向过程的结果。图 9展示了在逆向过程中生成的 x_0 样本的质量。在这一实验中, 观察到在较大比例尺上首先出现图像的整体特征, 而细节逐渐呈现。相比之下, 当进行 $x_0 \rightarrow x_t$ 的正向过程时, 当 t 较小时, 除了细微的细节外, 几乎所有细节都被保留下来; 而当 t 增大时, 只有大规模的特征被保留。这一实验结果揭示了模型在逆向生成过程中对图像特征的分层提取能力。这种逐步解压的生成方式为模型的应用提供了一种有趣的途径, 尤其在需要逐渐生成细节或从粗到精的场景中, 该方法可能具有一定的实际应用前景。



图 9. $x_t \rightarrow x_0$ 逐步生成结果, 每隔 10 步输出一次结果

此外我们在相同参数设置下对比原模型和改进模型生成结果, 采用模型生成结果和原图像的 MSE 损失作为对比指标, 令 T 为不同迭代生成次数, 对比结果如表 2所示。

表 2. 相同参数下 MSE 对比结果

T	原模型生成结果	改进模型生成结果
100	21.47	20.08
200	19.23	19.11
300	19.21	18.83
400	18.89	17.32
500	18.33	17.12
600	18.31	16.92
700	18.32	16.94

可以看到，随着迭代次数 T 的增加，模型生成效果在不断提高最终达到饱和，相同迭代次数下，我们改进后的模型生成结果优于原模型生成结果。

6 总结与展望

原文中使用了扩散模型以提供高质量的图像样本，并且发现了扩散模型与变分推理之间的联系，适用于训练马尔可夫链、去噪分数匹配、退火 Langevin 动力学，以及扩展的基于能量的模型，自回归模型和渐进有损压缩。

在原文的基础上，我们设计了新的模型和损失函数，排除了随机噪音对生成结果的影响。由于新模型和损失函数似乎对图像数据有很好的归纳偏差，我们期待着进一步研究它们在其他数据模式中的效用，以及作为其他类型生成模型和机器学习系统中的组件的潜在价值。我们的设计在模型框架和生成目标的选择上与原文有所不同。与原文中生成基于高斯分布噪音 $\bar{\epsilon}_t$ 的目标不同，我们的模型的生成目标为输入图像，摒弃了随机噪音引入的不确定性和不稳定性，这一选择使得我们的模型在去噪任务中表现出更好的重建质量和鲁棒性。

我们的实验结果表明，在 CIFAR10 数据集上，我们的改进模型在 IS 分数、FID 分数和负对数似然（NLL）方面均获得了提升。特别是在 FID 分数上，相较于原始 DDPM 的 3.17 分，我们改进后的模型取得了更佳的表现。我们还观察到，相对于训练集计算的 FID 分数，我们的模型在测试集上也取得了更好的分数，这显示出了我们模型的泛化能力。此外，我们运行了一个渐进的无条件生成实验，通过逐步解压随机比特模拟逆向过程。这一实验结果显示，在逆向生成过程中，模型展示了对图像特征的逐层提取能力，即大比例尺的整体特征先出现，细节随后呈现。这为模型在逐渐生成细节或从粗到精的场景中的应用提供了一种可能的途径。

我们的工作在原文的基础上进行了有益的拓展和改进，使得扩散模型在新的模型设计和损失函数选择上更具实用性和性能优势。另外由于新模型和损失函数似乎对图像数据有很好的归纳偏差，期待研究它们在其他数据模式中的效用，以及作为其他类型的生成模型和机器学习系统中的组件。

参考文献

- [1] Org. Cambridge. Ebooks. Online. Book. Author@Bdfd. *Markov Chains*. Markov Chains.
- [2] Merlijn Blaauw and Jordi Bonada. Modeling and transforming speech using variational autoencoders. In *Interspeech*, 2016.
- [3] F. T. Calkins, R. C. Smith, and A. B. Flatau. Energy-based hysteresis model for magnetostrictive transducers. *IEEE Transactions on Magnetics*, 36(2):429–439, 2002.
- [4] Hyunsoek Choi and Miyoung Shin. Learning radial basis function model with matching score quality for person authentication in multimodal biometrics. In *Asian Conference on Intelligent Information and Database Systems*, 2009.
- [5] Kihwan Choi, Joon Seok Lim, and Sung Won Kim. Real-time image reconstruction for low-dose ct using deep convolutional generative adversarial networks (gans). In *Physics of Medical Imaging*, 2018.
- [6] Carl Doersch. Tutorial on variational autoencoders. 2016.
- [7] Arsham Ghahramani, Fiona M Watt, and Nicholas M Luscombe. Generative adversarial networks uncover epidermal regulators and predict single cell perturbations. 2018.
- [8] Laurent Girin, Simon Leglaive, Xiaoyu Bie, Julien Diard, Thomas Hueber, and Xavier Alameda-Pineda. Dynamical variational autoencoders: A comprehensive review. *Foundations and trends in machine learning*, page 15, 2022.
- [9] GitHub. GitHub Topics - Stable Diffusion, 2023. Accessed on 2023-11-20.
- [10] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. 2020.
- [11] Aapo Hyvonen and Peter Dayan. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(4):695–709, 2005.
- [12] Alexander G. Ororbia II, C. Lee Giles, and David Reitter. Online semi-supervised learning with deep hybrid boltzmann machines and denoising autoencoders. *Computer Science*, 2015.
- [13] H Isack and Y Boykov. Energy based multi-model fitting and matching for 3d reconstruction. In *IEEE*, 2014.
- [14] Unnat Jain, Ziyu Zhang, and Alexander Schwing. Creativity: Generating diverse questions using variational autoencoders. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

- [15] Jin Young Kim, Seok Jun Bu, and Sung Bae Cho. Zero-day malware detection using transferred generative adversarial networks based on deep autoencoders. *Information ences*, pages 83–102, 2018.
- [16] Urs Köster and Aapo Hyvärinen. A two-layer ica-like model estimated by score matching. In *Artificial Neural Networks - ICANN 2007, 17th International Conference, Porto, Portugal, September 9-13, 2007, Proceedings, Part II*, 2007.
- [17] Oran Lang, Yossi Gandelsman, Michal Yarom, Yoav Wald, Gal Elidan, Avinatan Hassidim, William T. Freeman, Phillip Isola, Amir Globerson, and Michal Irani. Explaining in style: Training a gan to explain a classifier in stylespace, 2021.
- [18] Wong Wai Keung Li. On a mixture autoregressive model. *Journal of the Royal Statistical Society*, 62(455):95–115, 2010.
- [19] L. A. Lievrouw and J. T. Pope. Contemporary art as aesthetic innovation applying the diffusion model in the art world. *Science Communication*, 15(4):373–395, 2020.
- [20] Ping Lu, Matt Morris, Seth Brazell, Cody Comiskey, and Yuan Xiao. Using generative adversarial networks to improve deep-learning fault interpretation networks. *Leading edge*, 2018.
- [21] Alireza Makhzani and Brendan Frey. A winner-take-all method for training sparse convolutional autoencoders. *Eprint Arxiv*, 2014.
- [22] Diego Marcheggiani and Ivan Titov. Discrete-state variational autoencoders for joint discovery and factorization of relations. *Transactions of the Association for Computational Linguistics*, 4(2):231–244, 2016.
- [23] Vahid Mirjalili, Sebastian Raschka, Anoop Namboodiri, and Arun Ross. Semi-adversarial networks: Convolutional autoencoders for imparting privacy to face images. *IEEE*, pages 82–89, 2018.
- [24] Suzanne O’Keefe. Job creation in california’s enterprise zones: a comparison using a propensity score matching model. *Journal of Urban Economics*, 2004.
- [25] Noseong Park, Mahmoud Mohammadi, Kshitij Gorde, Sushil Jajodia, Hongkyu Park, and Youngmin Kim. Data synthesis based on generative adversarial networks. *Proceedings of the VLDB Endowment*, 2018.
- [26] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *Computer ence*, 2015.
- [27] Adam Roberts, Jesse Engel, and Douglas Eck. Hierarchical variational autoencoders for music. 2017.

- [28] Anurag Sarkar and Seth Cooper. Dungeon and platformer level blending and generation using conditional vaes. 2021.
- [29] Sandeep Subramanian, Sai Rajeswar Mudumba, Alessandro Sordoni, Adam Trischler, Aaron C. Courville, and Chris Pal. Towards text generation with adversarially learned neural outlines. In *Neural Information Processing Systems*, 2018.
- [30] G. Swoboda and Q. Yang. An energy-based damage model of geomaterials—ii. deduction of damage evolution laws. *International Journal of Solids and Structures*, 36(12):1735–1755, 1999.
- [31] Subarna Tripathi, Zachary C Lipton, and Truong Q Nguyen. Correction by projection: Denoising images with generative adversarial networks. 2018.
- [32] Meng Wang, Youbin Chen, and Xingjun Wang. Recognition of handwritten characters in chinese legal amounts by stacked autoencoders. In *International Conference on Pattern Recognition*, 2014.
- [33] Gregory P Way and Casey S Greene. Extracting a biologically relevant latent space from cancer transcriptomes with variational autoencoders. pages 80–91, 2018.
- [34] James T. Wilson, Riccardo Moriconi, Frank Hutter, and Marc Peter Deisenroth. The reparameterization trick for acquisition functions. 2017.
- [35] Jianwen Xie, Yang Lu, Song Chun Zhu, and Ying Nian Wu. A theory of generative convnet. *JMLR.org*, 2016.
- [36] Junhai Zhai, Sufang Zhang, Junfen Chen, and Qiang He. Autoencoder and its various variants. In *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2018.
- [37] Ying Zhan, Dan Hu, Yuntao Wang, and Xianchuan Yu. Semisupervised hyperspectral image classification based on generative adversarial networks. *IEEE Geoscience and Remote Sensing Letters*, 15(2):1–5, 2018.
- [38] Zhang, Yahui, Moumni, Ziad, Zhu, Jihong, Van Herpen, Alain, and Weihong. Energy-based fatigue model for shape memory alloys including thermomechanical coupling. *Smart Materials and Structures*, 2016.
- [39] Meng Meng Zhang, Ling Ma, Zhi Hui Yang, Yang Yang, and Hui Hui Bai. A medical image segmentation method fusing anisotropic diffusion model. *Advanced Materials Research*, 268–270:1121–1126, 2011.
- [40] Yuhong Zhang and Wei Li. An energy-based stochastic model for wireless sensor networks. *Wireless Sensor Network*, 3(9):322–328, 2011.

- [41] Tiancheng Zhao, Ran Zhao, and Maxine Eskenazi. Learning discourse-level diversity for neural dialog models using conditional variational autoencoders. 2017.