

BELM: 用于扩散模型精确反演的双向显式线性多步采样器

摘要

扩散模型采样的反演，其目的是找到样本对应的初始噪声，在各种任务中起着至关重要的作用。最近，人们提出了几种启发式精确反演采样器，以不需要训练的方式解决不精确反演问题。然而，这些启发式采样器的理论性质仍然未知，它们往往表现出平庸的采样质量。复现论文引入了精确反演采样器的一个通用公式，双向显式线性多步 (BELM) 采样器，并且通过 LTE 最小化方法提出了最优 BELM (O-BELM) 采样器。本工作聚焦于复现论文所提出的 O-BELM 采样器，并通过在 COCO 数据集上进行重建实验验证该采样器的重建有效性，在预训练的扩散模型中验证是否在实现高质量采样的同时具有精确的反演特性。同时在 SD1.5 和 SD2.0 两个预训练模型上验证 O-BELM 采样器在图像编辑中的潜力。最后我们针对 O-BELM 采样器在图像编辑中的不足创新地提出了改进策略。

关键词：扩散模型；精确反演；采样器

1 引言

新兴的扩散模型 (diffusion models, DMs) [12, 29, 32, 33]，通过学习反向扩散过程从初始噪声中生成数据分布样本，已被证明是一种有效的数据分布建模技术，特别是在生成高质量图像方面 [3, 7, 13, 26–28]。扩散模型中的扩散过程及其采样过程可以描述为正向和相应的后向随机微分方程 (SDE) [1, 33]。此外，采样过程也可以表示为确定性扩散常微分方程 (ODE) [30, 33]，在一些论文中也称为概率流 ODE (PF-ODE)。值得注意的是，后向 SDE 和扩散 ODE 具有相同的边际分布 [33]。

扩散采样的反演旨在阐明样本与初始噪声之间的对应关系，在扩散模型的各种任务中起着至关重要的作用。扩散反演有多种下游应用，包括图像编辑 [10, 34]、图像插值 [30]、绘制 [5] 和超分辨率 [38]。一些研究 [5, 17, 18] 试图在基于 SDE 的扩散采样的背景下解决反演任务。然而，由于 SDE 固有的随机性，这些工作未能实现数学上精确的反演。

相反，扩散 ODE 自然地给出了样本和噪声之间的对应关系。通过考虑扩散 ODE 的一阶显式欧拉离散化，得到了著名的 DDIM [30] 及其反演。然而，正如 [10] 的工作所指出的那样，由于 DDIM 与其反演之间的原理图不匹配，DDIM 反演引入了不一致问题（见图 1）。使用 DDIM 反演从 \mathbf{x}_0 编码到 \mathbf{x}_T ，然后使用 DDIM 解码，通常会导致原始样本的不精确重建（见图 2）。null-text 反演 [24] 的工作引入了迭代优化的密集训练，但仍然无法实现数学上精确的反演。

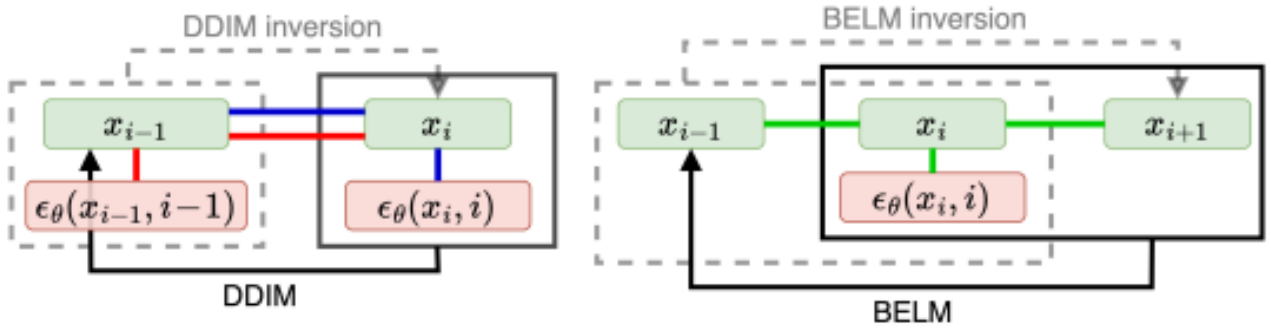


图 1. DDIM(左)和 BELM(右)的示意图。DDIM 使用 \mathbf{x}_i 和 $\epsilon_\theta(\mathbf{x}_i, i)$ 基于 $\mathbf{x}_i, \mathbf{x}_{i-1}$ 和 $\epsilon_\theta(\mathbf{x}_i, i)$ 之间的线性关系（由蓝线表示）来计算 \mathbf{x}_{i-1} 。然而，DDIM 反演使用 \mathbf{x}_{i-1} 和 $\epsilon_\theta(\mathbf{x}_{i-1}, i-1)$ 基于不同的线性关系（由红线表示）来计算 \mathbf{x}_i 。这种不匹配导致 DDIM 的反演不精确。相比之下，BELM 试图建立 $\mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{x}_{i+1}$ 和 $\epsilon_\theta(\mathbf{x}_i, i)$ 之间的线性关系（由绿线表示）。BELM 及其反演是从这个单一关系中推导出来的，这有助于实现精确的反演。具体来说，BELM 使用 $\mathbf{x}_i, \mathbf{x}_{i+1}$ 和 $\epsilon_\theta(\mathbf{x}_i, i)$ 的线性组合来计算 \mathbf{x}_{i-1} ，而 BELM 反演使用 $\mathbf{x}_{i-1}, \mathbf{x}_i$ 和 $\epsilon_\theta(\mathbf{x}_i, i)$ 的线性组合来计算 \mathbf{x}_{i+1} 。双向显式约束意味着这种线性关系不包括双向端点的导数，即 $\epsilon_\theta(\mathbf{x}_{i-1}, i-1)$ 和 $\epsilon_\theta(\mathbf{x}_{i+1}, i+1)$ 。

最近，已经提出了几种启发式精确反演采样器，以不需要训练的方式解决这种不精确反演问题 [36, 39]。这些采样器使数学上精确的反演不需要额外的训练，因此与预训练模型兼容。受正态流中的仿射耦合层的启发 [8, 9]，EDICT [36] 直观地引入了辅助扩散状态，并对原始和辅助扩散状态进行交替的混合更新。后来，BDIA [39] 采用对称双向积分结构，直观地实现精确反演。然而，这些启发式精确反演采样器由于其直观的公式设计，往往会损害采样质量。它们还可能引入不必要的额外计算开销或非鲁棒超参数。

我们首先通过论文给出的一般精确反演采样器一般公式，并结合 LTE 最小化方法来设计 BELM 内最优情况的公式，得到最优双向显式线性多步（O-BELM）采样器。而后，我们在 COCO 数据集上进行了图像重建实验，复现的 O-BELM 采样器实现了和论文一致的精确反演，即使其能够精确地恢复复杂的图像特征。此外，我们还在 CIFAR10、CelebA-HQ 数据集上预训练的扩散模型以及 SD1.5 和 SD2.0 这类大规模预训练的扩散模型中进行无条件条件和条件图像生成实验，复现的 O-BELM 采样器得到和论文中一样的采样质量。

最后在复现图像编辑下游任务中，我们发现直接通过修改文本并利用 O-BELM 采样器编辑后的图片容易保持不变或者生成不了想要的内容，这与论文中的效果有所出入，因此本工作也根据此问题进行了创新改进。

本文剩余的章节将按以下逻辑展开，在第二章介绍与本文相关的工作，第三章详细介绍本文的方法，第四章介绍复现的细节以及本文创新点，最后的总结本文的结论和展望。

2 相关工作

2.1 扩散模型和扩散 SDEs

假设我们有一个 d 维随机变量 $\mathbf{x}(0) \in \mathbb{R}^d$ ，它遵循一个未知的目标分布 $q_0(x_0)$ 。扩散模型 (DMs) 定义了一个从 $\mathbf{x}(0)$ 开始的前向过程 $\{\mathbf{x}(t)\}_{t \in [0, T]}$ ，其中 $T > 0$ ，使得在给定 $\mathbf{x}(0)$ 的条



图 2. 50 步下 DDIM 和精确反演采样器的图像重建结果和 MSE 误差。红色矩形表示 DDIM 重建图像中不一致的部分。

件下， $\mathbf{x}(t)$ 的分布满足

$$q_{t|0}(\mathbf{x}(t)|\mathbf{x}(0)) = \mathcal{N}(\mathbf{x}(t); \alpha(t)\mathbf{x}(0), \sigma^2(t)\mathbf{I}), \quad (1)$$

其中 $\alpha(\cdot), \sigma(\cdot) \in \mathcal{C}([0, T], \mathbb{R}^+)$ 具有有界导数，为简单起见，我们将它们记为 α_t 和 σ_t 。 α_t 和 σ_t 的选择被称为扩散模型的噪声调度。根据 [16, 19, 21]，在对 $\alpha(\cdot)$ 和 $\sigma(\cdot)$ 进行一些假设的情况下，前向过程可以建模为一个线性 SDE，也称为 Ornstein-Uhlenbeck 过程：

$$d\mathbf{x}(t) = f(t)\mathbf{x}(t)dt + g(t)dB_t, \quad (2)$$

其中 B_t 是标准的 d 维布朗运动 (BM)， $f(t) = \frac{d \log \alpha_t}{dt}$ 且 $g^2(t) = \frac{d \sigma_t^2}{dt} - 2 \frac{d \log \alpha_t}{dt} \sigma_t^2$ 。在一些正则条件下，上述前向 SDE(2) 具有从时间 \tilde{T} 到 0 的反向 SDE，该过程从 $\mathbf{x}(t)$ 开始 [1]：

$$d\mathbf{x}(t) = [f(t)\mathbf{x}(t) - g^2(t)\nabla_{\mathbf{x}(t)} \log q(\mathbf{x}(t), t)] dt + g(t)d\tilde{B}_t, \quad (3)$$

其中 \tilde{B}_t 是反向时间布朗运动, $q(\mathbf{x}(t), t)$ 是前向过程的单时间边际分布。在实践中, 扩散模型 [12, 33] 使用 $\varepsilon_\theta(\mathbf{x}(t), t)$ 来估计 $-\sigma(t)\nabla_{\mathbf{x}(t)} \log q(\mathbf{x}(t), t)$, 并且参数 θ 通过以下目标进行优化:

$$\theta^* = \arg \min_{\theta} \mathbb{E}_t \left\{ \lambda_t \mathbb{E}_{x_0, x_t} [\|s_\theta(x_t, t) - \nabla_{x_t} \log p(x_t, t|x_0, 0)\|^2] \right\}. \quad (4)$$

2.2 扩散 ODE 和 DDIM

注意到反向 SDE(3)有一个相关的概率流 ODE (也称为扩散 ODE), 这是一个确定性过程, 共享相同的单时间边际分布 [33]:

$$d\mathbf{x}(t) = \left[f(t)\mathbf{x}(t) - \frac{1}{2}g^2(t)\nabla_{\mathbf{x}(t)} \log q(\mathbf{x}(t), t) \right] dt. \quad (5)$$

将 $f(t)$ 和 $g(t)$ 代入方程(5), 我们得到以下一阶形式:

$$d\left(\frac{\mathbf{x}(t)}{\alpha_t}\right) = \varepsilon_\theta(\mathbf{x}(t), t) d\left(\frac{\sigma_t}{\alpha_t}\right). \quad (6)$$

著名的 DDIM 采样器 [30] 可以通过将显式欧拉方法应用于方程(6) 得到:

$$\mathbf{x}_{i-1} = \frac{\alpha_{i-1}}{\alpha_i} \mathbf{x}_i + \left(\sigma_{i-1} - \frac{\alpha_{i-1}}{\alpha_i} \sigma_i \right) \varepsilon_\theta(\mathbf{x}_i, i). \quad (7)$$

DDIM 的反转是通过在方程 (6) 的反向应用显式欧拉方法得到的:

$$\mathbf{x}_i = \frac{\alpha_i}{\alpha_{i-1}} \mathbf{x}_{i-1} + \left(\sigma_i - \frac{\alpha_i}{\alpha_{i-1}} \sigma_{i-1} \right) \varepsilon_\theta(\mathbf{x}_{i-1}, i-1). \quad (8)$$

2.3 扩散模型的直观精确反演采样器

在实践中, 我们观察到 DDIM 反演 (8) 存在不一致性问题。考虑一个样本 \mathbf{x}_0 ; 使用 DDIM 反演, 我们得到相应的噪声 \mathbf{x}_T , 然后使用 DDIM 重建一个 \mathbf{x}_0^* 。重建的 \mathbf{x}_0^* 会与原始样本 \mathbf{x}_0 表现出显著的不一致性。最近, 两种精确反演采样器 EDICT 和 BDIA 被启发式地提出, 以无训练的方式解决这种不一致性问题。

EDICT 采样器 受归一化流中仿射耦合层的启发 [8, 9], 最近的工作 [36] 提出了 EDICT 来实现精确的扩散反演。其基本思想是引入一个辅助扩散状态 \mathbf{y}_t 与 \mathbf{x}_t 耦合。记 $a_i = \frac{\alpha_{i-1}}{\alpha_i}$ 和 $b_i = \sigma_{i-1} - \frac{\alpha_{i-1}}{\alpha_i} \sigma_i$, EDICT 的公式如下:

$$\begin{cases} \mathbf{x}_i^{inter} = a_i \mathbf{x}_i + b_i \varepsilon_\theta(\mathbf{y}_i, i), & \mathbf{y}_i^{inter} = a_i \mathbf{y}_i + b_i \varepsilon_\theta(\mathbf{x}(t)^{inter}, i), \\ \mathbf{x}_{i-1} = p \mathbf{x}_i^{inter} + (1-p) \mathbf{y}_i^{inter}, & \mathbf{y}_{i-1} = p \mathbf{y}_i^{inter} + (1-p) \mathbf{x}_{i-1}. \end{cases} \quad (9)$$

其中 $p \in (0, 1)$ 是混合系数。

BDIA 采样器 BDIA 采样器 [39] 利用对称的双向积分结构来实现精确反演。BDIA 将 DDIM 的表达式 (7) 重新表述为 $\mathbf{x}_{i-1}^{DDIM} = \mathbf{x}_i^{DDIM} + \Delta(i \rightarrow i-1 | \mathbf{x}_i^{DDIM})$, 并将 DDIM 反演的

表达式 (8) 重新表述为 $\mathbf{x}_i^{\text{DDIM}} = \mathbf{x}_{i-1}^{\text{DDIM}} + \Delta(i-1 \rightarrow i | \mathbf{x}_{i-1}^{\text{DDIM}})$ 。BDIA 直观地利用 $-[(1-\gamma)(\mathbf{x}_{i+1} - \mathbf{x}_i) + \gamma\Delta(i \rightarrow i+1 | \mathbf{x}_i)]$ 来近似从 x_{i+1} 到 x_i 的增量, 并将 $\Delta(i \rightarrow i-1 | \mathbf{x}_i)$ 作为从 x_i 到 x_{i-1} 的增量。因此, BDIA 的更新规则为:

$$\mathbf{x}_{i-1} = \underbrace{\mathbf{x}_{i+1} - [(1-\gamma)(\mathbf{x}_{i+1} - \mathbf{x}_i) + \gamma\Delta(i \rightarrow i+1 | \mathbf{x}_i)]}_{\text{increment}(\mathbf{x}_{i+1} \rightarrow \mathbf{x}_i)} + \underbrace{\Delta(i \rightarrow i-1 | \mathbf{x}_i)}_{\text{increment}(\mathbf{x}_i \rightarrow \mathbf{x}_{i-1})}. \quad (10)$$

然而, 这些启发式采样器的理论性质仍然未知, 并且它们通常会表现出采样质量的下降。据我们所知, 目前还没有系统的方法来推导出一种同时具备精确扩散反演特性并保持高采样质量的扩散采样器。

3 本文方法

3.1 通用双向显式线性多步 (BELM) 采样器

在本节中, 我们首先将扩散采样过程建模为一个适定的初值问题, 以便于后续分析。通过对 DDIM 反演的重新思考, 我们提出了通用的双向显式线性多步 (BELM) 采样器, 并以变步长变公式 (VSVF) 的方式实现。我们进一步说明, EDICT 和 BDIA 实际上是 BELM 框架的特殊实例。

扩散采样问题作为初值问题通过记 $\bar{\mathbf{x}}(t) \equiv \frac{\mathbf{x}(t)}{\alpha_t}$ 、 $\bar{\sigma}(t) \equiv \frac{\sigma_t}{\alpha_t}$ 和 $\bar{\epsilon}_\theta(\bar{\mathbf{x}}(t), \bar{\sigma}_t) \equiv \epsilon_\theta(\mathbf{x}(t), t)$, 扩散模型(6)的确定性采样过程可以看作是一个特殊的反向时间扩散初值问题 (IVP) [35] [4]:

$$d\bar{\mathbf{x}}(t) = \bar{\epsilon}_\theta(\bar{\mathbf{x}}(t), \bar{\sigma}_t) d\bar{\sigma}_t, \quad (11)$$

其中 $\bar{\mathbf{x}}(T) = \mathbf{x}(T)/\alpha_T$ 。在进行任何进一步分析之前, 一个基本问题是给定的扩散初值问题 (11) 是否允许解存在, 如果存在, 这个解是否唯一。首先, 我们需要对扩散采样问题(6)建立一些正则性假设。

假设 1. $\epsilon_\theta(\mathbf{x}, t)$ 关于 t 连续, 并且关于 \mathbf{x} 是 Lipschitz 连续的, Lipschitz 常数为 L_{ϵ_θ} , 这意味着 $\|\epsilon_\theta(\mathbf{x}_1, t) - \epsilon_\theta(\mathbf{x}_2, t)\|_2 \leq L_{\epsilon_\theta} \|\mathbf{x}_1 - \mathbf{x}_2\|_2$ 。

假设 1 是扩散模型文献中关于噪声预测器 $\epsilon_\theta(\mathbf{x}, t)$ 的常见假设 [31]。在假设 1 的条件下, 我们可以通过直接应用初值问题理论中的存在性和唯一性定理 [4] 来确认扩散初值问题 (11) 是适定的。

命题 1. 在假设 1 的条件下, 扩散初值问题 (11) 存在唯一解。

在本文中, $\mathbf{x}(\cdot)$ 表示连续解, \mathbf{x}_i 表示数值近似。

对 DDIM 反演的重新思考如图 1所示, DDIM (7) 及其反演 (8) 是基于不同的线性关系推导的。我们强调, 这种不匹配导致了 DDIM 的不精确反演。基于这一观察, 一个自然的想法是基于与 DDIM 相同的线性关系来构造 DDIM 反演, 以消除这种不匹配。遗憾的是, DDIM 是基于 \mathbf{x}_i 、 \mathbf{x}_{i-1} 和 $\epsilon_\theta(\mathbf{x}_i, i)$ 之间的关系构造的 (利用 \mathbf{x}_i 和 $\epsilon_\theta(\mathbf{x}_i, i)$ 计算 \mathbf{x}_{i-1}), 而 DDIM 反演无法直接利用这一关系来计算 \mathbf{x}_i , 因为在 DDIM 反演的情况下 $\epsilon_\theta(\mathbf{x}_i, i)$ 也是未知的。这种关系对 DDIM 是显式的, 但对 DDIM 反演是隐式的。需要注意的是, 隐式方程必须使用迭代方法 (如牛顿法 [35]) 来求解, 这些方法耗时且可能在扩散模型的背景下引入数值误差 [14, 22]。

为了解决这个问题，我们建立了相邻状态和导数之间的新关系，该关系可以在两个方向上显式计算。随后，我们基于这一单一的线性关系来构造采样器及其反演，以实现精确反演。这是 BELM 采样器的基本概念。

双向显式线性多步 (BELM) 采样器为了建立 \mathbf{x}_i 、 \mathbf{x}_{i-1} 、 $\boldsymbol{\varepsilon}_\theta(\mathbf{x}_i, i)$ 和 $\boldsymbol{\varepsilon}_\theta(\mathbf{x}_{i-1}, i-1)$ 之间的线性关系，并且能够在两个方向上显式计算，我们必须排除 $\boldsymbol{\varepsilon}_\theta(\mathbf{x}_i, i)$ 和 $\boldsymbol{\varepsilon}_\theta(\mathbf{x}_{i-1}, i-1)$ 。然而，这种排除会导致关系缺乏足够的信息。因此，必须考虑更多的状态。这促使我们探索线性多步 (LM) 方法的概念 [4]，作为推导扩散初值问题相邻状态和导数之间线性关系的手段。然而，扩散模型常用的噪声调度会导致 $\{\bar{\sigma}_i\}, i = 1 \dots N$ 的非等距序列。因此，我们不应使用固定步长的经典 LM 方法，而应考虑变步长变公式 (VSVF) 方式 [6]，它使用动态多步公式以适应不同的步长。设 $t_0 < t_1 < \dots t_N = t_0 + T$ 是 $[t_0, t_0 + T]$ 上的网格， $h_i = \bar{\sigma}_i - \bar{\sigma}_{i-1}, i = N \dots 1, h_0 = \bar{\sigma}_0$ 且 $h = \max h_i$ ，关于方程 (11) 的 k 步 VSVF LM 方法将在点 $\bar{\sigma}_{i-1}$ 处计算 \mathbf{x}_{i-1} ，其差分方程为：

$$\bar{\mathbf{x}}_{i-1} = \sum_{j=1}^k a_{i,j} \cdot \bar{\mathbf{x}}_{i-1+j} + \sum_{j=0}^k b_{i,j} \cdot h_{i-1+j} \cdot \bar{\boldsymbol{\varepsilon}}_\theta(\bar{\mathbf{x}}_{i-1+j}, \bar{\sigma}_{i-1+j}), \quad (12)$$

其中更新系数和步长都依赖于 i 。在本文中，除非另有说明，任何对 LM 的引用默认都指 VSVF LM。如果方程 (12) 中对于所有 i 都有 $b_{i,0} = 0$ ，则该方法称为**显式**，因为该公式可以直接计算 $\bar{\mathbf{x}}_{i-1}$ 。显然，LM (12) 有一个反向公式，它也是一个 k 步 LM，如下所示（假设 $a_{i,k} \neq 0$ ）：

$$\bar{\mathbf{x}}_{i-1+k} = \frac{1}{a_{i,k}} \cdot \bar{\mathbf{x}}_{i-1} - \sum_{j=1}^{k-1} \frac{a_{i,j}}{a_{i,k}} \cdot \bar{\mathbf{x}}_{i-1+j} + \sum_{j=0}^k \frac{b_{i,j}}{a_{i,k}} \cdot h_{i-1+j} \cdot \bar{\boldsymbol{\varepsilon}}_\theta(\bar{\mathbf{x}}_{i-1+j}, \bar{\sigma}_{i-1+j}). \quad (13)$$

如果反向 VSVFM 是显式的，即对于所有 i 都有 $b_{i,k} = 0$ ，我们称原始 LM (12) 为**反向显式**。现在，我们可以定义一个 k 步 LM 为**双向显式**，当它既是显式的又是反向显式的。我们将遵守双向显式约束的 LM 采样器称为双向显式线性多步 (BELM) 采样器，其一般形式为：

$$\bar{\mathbf{x}}_{i-1} = \sum_{j=1}^k a_{i,j} \cdot \bar{\mathbf{x}}_{i-1+j} + \sum_{j=1}^{k-1} b_{i,j} \cdot h_{i-1+j} \cdot \bar{\boldsymbol{\varepsilon}}_\theta(\bar{\mathbf{x}}_{i-1+j}, \bar{\sigma}_{i-1+j}). \quad (14)$$

我们强调，这种双向显式约束是数学上精确扩散反演的关键：

命题 2。任何满足 $a_{i,k} \neq 0$ 的 BELM 方法 (14) 都具有精确反演性质。

作为一个实例，在方程 (14) 中设置 $k = 2$ 可以得到 2 步 BELM 扩散采样器：

$$\bar{\mathbf{x}}_{i-1} = a_{i,2} \bar{\mathbf{x}}_{i+1} + a_{i,1} \bar{\mathbf{x}}_i + b_{i,1} h_i \bar{\boldsymbol{\varepsilon}}_\theta(\bar{\mathbf{x}}_i, \bar{\sigma}_i). \quad (15)$$

有关 3 步 BELM 扩散采样器、一般 k 步情况及其最优设计的详细信息，请读者参阅附录 A.4 和 A.5。在本文的主体部分，除非另有说明，我们默认指 2 步情况。

BDIA 和 EDICT 作为 BELM 的特殊情况我们发现，尽管 BDIA 和 EDICT 是从启发式思想发展而来的，但它们都是 BELM 框架中的特殊情况。也就是说，它们的精确反演性质源于它们本质上是 BELM 采样器的实例。

备注 1。EDICT (9) 和 BDIA (10) 都是 BELM 框架中的特殊情况。

表 1. 不同采样器的理论性质比较

	理论性质			
	精确反演	局部误差	零稳定性	全局收敛性
DDIM [30]	×	$\mathcal{O}(\alpha_i h_i^2)$	✓	✓
EDICT [36]	✓	$\mathcal{O}(\sqrt{\alpha_{i-1}} h_i)$	不明确	不明确
BDIA [39]	✓	$\mathcal{O}(\alpha_i (h_i + h_{i+1})^2)$	不明确	不明确
O-BELM (本文)	✓	$\mathcal{O}(\alpha_i (h_i + h_{i+1})^3)$	✓	✓

3.2 Optimal-BELM (O-BELM) 采样器

在本节中，我们系统地研究了 BELM 框架中的局部截断误差 (LTE)，并表明现有的精确反演采样器的启发式设计会导致次优的 LTE。因此，我们引入了最优 BELM (O-BELM)，它通过 LTE 最小化方法开发了更精细的动态公式。我们还进行了额外的分析，以验证 O-BELM 的理论稳定性和全局收敛性。

3.2.1 局部截断误差分析

局部截断误差 (LTE) 量化了在每一步更新中引入的误差。具体来说，它计算数值解与其真实解之间的差异，假设在前一步状态中已知真实解。

定义 1. *BELM* (15) 在每一步 i 处的局部截断误差定义为：

$$\tau_i = \bar{\mathbf{x}}(t_{i-1}) - a_{i,2}\bar{\mathbf{x}}_{i+1} - a_{i,1}\bar{\mathbf{x}}_i - b_{i,1}h_i\bar{\varepsilon}_\theta(\bar{\mathbf{x}}_i, \bar{\sigma}_i). \quad (16)$$

在假设 2 (详见附录 A.3) 下，我们可以利用泰勒展开来研究 BELM (15) 的 LTE，如下所示：

命题 3. 在假设 2 下，*BELM* (15) 的 LTE 具有以下一般形式：

$$\tau_i = c_{i,1}\bar{\mathbf{x}}(t_{i-1}) + c_{i,2}\bar{\varepsilon}_\theta(\bar{\mathbf{x}}(t_{i-1}), \bar{\sigma}_{i-1}) + c_{i,3}\nabla_{\bar{\sigma}_{i-1}}\bar{\varepsilon}_\theta(\bar{\mathbf{x}}(t_{i-1}), \bar{\sigma}_{i-1}) + \mathcal{O}((h_i + h_{i+1})^3), \quad (17)$$

其中 $c_{i,1} = 1 - a_{i,1} - a_{i,2}$, $c_{i,2} = -a_{i,1}h_i - a_{i,2}(h_i + h_{i+1}) - b_{i,1}h_i$, 且 $c_{i,3} = -\frac{a_{i,1}}{2}h_i^2 - \frac{a_{i,2}}{2}(h_i + h_{i+1})^2 - b_{i,1}h_i^2$ 。

在扩散模型的任务中，我们主要关注的是 \mathbf{x}_{i-1} 上的 LTE，而不是 $\bar{\mathbf{x}}_{i-1}$ 。我们将 \mathbf{x}_i 上的 LTE 记为 \mathbf{e}_i 。显然， $\mathbf{e}_i = \alpha_{i-1}\tau_i$ 。我们研究了现有采样器的 LTE，如下所示：

推论 1. 在假设 2 下，*DDIM* 采样器 (7) 的 LTE \mathbf{e}_i 为 $\mathcal{O}(\alpha_{i-1}h_i^2)$ ；*BDIA* 采样器 (10) 的 LTE \mathbf{e}_i 为 $\mathcal{O}(\alpha_{i-1}(h_i + h_{i+1})^2)$ (对于任意固定的 $\gamma \in [0, 1]$)；*EDICT* 采样器 (9) 的 LTE \mathbf{e}_i 为 $\mathcal{O}(\sqrt{\alpha_{i-1}}h_i)$ (对于任意常数 $p \in (0, 1)$)。

3.2.2 通过 LTE 最小化的最优 BELM 采样器

我们随后证明，通过精心设计公司，我们可以在 BELM 框架内实现比现有次优实例更高阶的 LTE。具体来说，我们采用了 LTE 最小化方法，该方法受到著名 LM 方法（如 Adams-Bashforth 方法 [2] 或 Adams-Moulton 方法 [23, 25]）设计的启发。

命题 4. 在假设 2 下, 当公式设计为 $a_{i,1} = \frac{h_{i+1}^2 - h_i^2}{h_{i+1}^2}$ 、 $a_{i,2} = \frac{h_i^2}{h_{i+1}^2}$ 、 $b_{i,1} = -\frac{h_i + h_{i+1}}{h_{i+1}}$ 时, BELM 扩散采样器 (15) 的 LTE τ_i 可以精确到 $\mathcal{O}((h_i + h_{i+1})^3)$ 。

具体来说, 公式设计如下:

$$a_{i,1} = \frac{h_{i+1}^2 - h_i^2}{h_{i+1}^2}, \quad a_{i,2} = \frac{h_i^2}{h_{i+1}^2}, \quad b_{i,1} = -\frac{h_i + h_{i+1}}{h_{i+1}}.$$

当满足这一条件时, 显然, \mathbf{x}_{i-1} 上的局部截断误差 (LTE) \mathbf{e}_i 为 $\mathcal{O}(\alpha_{i-1}(h_i + h_{i+1})^3)$ 。将设计的公式代入 (15), 我们推导出最优 BELM (O-BELM) 采样器:

$$\mathbf{x}_{i-1} = \frac{h_i^2}{h_{i+1}^2} \frac{\alpha_{i-1}}{\alpha_{i+1}} \mathbf{x}_{i+1} + \frac{h_{i+1}^2 - h_i^2}{h_{i+1}^2} \frac{\alpha_{i-1}}{\alpha_i} \mathbf{x}_i - \frac{h_i(h_i + h_{i+1})}{h_{i+1}} \alpha_{i-1} \varepsilon_\theta(\mathbf{x}_i, i). \quad (18)$$

O-BELM 扩散采样器 (18) 的反演公式为:

$$\mathbf{x}_{i+1} = \frac{h_{i+1}^2}{h_i^2} \frac{\alpha_{i+1}}{\alpha_{i-1}} \mathbf{x}_{i-1} + \frac{h_i^2 - h_{i+1}^2}{h_i^2} \frac{\alpha_{i+1}}{\alpha_i} \mathbf{x}_i + \frac{h_{i+1}(h_i + h_{i+1})}{h_i} \alpha_{i+1} \varepsilon_\theta(\mathbf{x}_i, i). \quad (19)$$

3.2.3 对 O-BELM 的进一步理论分析

在这里, 我们进一步证明, O-BELM 不仅在局部精度上表现出色, 还在**稳定性和全局收敛性**方面表现优异。

从 (15) 可以看出, 我们需要初始值才能将方法应用于扩散初值问题。其中, 初始值由初始条件给出, 但其他值需要通过其他方法计算, 例如使用 DDIM。无论如何, 初始值将包含数值误差, 因此必须确保初始值的扰动不会导致后续步骤中的误差爆炸。这一概念在数值分析中被定义为零稳定性。

定义 2: 如果存在常数 K , 使得对于任何两个序列 $\{\bar{\mathbf{x}}_i\}$ 和 $\{\bar{\mathbf{z}}_i\}$, 它们由相同的公式生成但具有不同的初始值 $\bar{\mathbf{x}}_N, \bar{\mathbf{x}}_{N-1}, \dots, \bar{\mathbf{x}}_{N-k+1}$ 和 $\bar{\mathbf{z}}_N, \bar{\mathbf{z}}_{N-1}, \dots, \bar{\mathbf{z}}_{N-k+1}$, 满足

$$\|\bar{\mathbf{x}}_i - \bar{\mathbf{z}}_i\| \leq K \max \{\|\bar{\mathbf{x}}_N - \bar{\mathbf{z}}_N\|, \|\bar{\mathbf{x}}_{N-1} - \bar{\mathbf{z}}_{N-1}\|, \dots, \|\bar{\mathbf{x}}_{N-k+1} - \bar{\mathbf{z}}_{N-k+1}\|\}, \quad (20)$$

对于所有 i , 且当 h 趋近于 0 时, LM(12)被称为**零稳定**。

我们还希望确保随着步长的减小, 方法会逐渐收敛到真实解, 这一概念与全局收敛性一致。

定义 3: 如果对于 (11) 的每个解 $\bar{\mathbf{x}}(t)$, 满足

$$\lim_{h \rightarrow 0} \max_{0 \leq t \leq N} \|\bar{\mathbf{x}}_i - \bar{\mathbf{x}}(t_i)\| = 0, \quad (21)$$

且初始误差 $\sum_{j=N}^{N-1+k} (\|\bar{\mathbf{x}}_j - \bar{\mathbf{x}}(t_j)\| + h_i \|\bar{\varepsilon}_\theta(\bar{\mathbf{x}}_j, \bar{\sigma}_j) - \bar{\varepsilon}_\theta(\bar{\mathbf{x}}(t_j), \bar{\sigma}_j)\|)$ 趋近于零, 则 LM(12) 是**全局收敛**的。

我们确认, 我们的 O-BELM 采样器具有良好的零稳定性以及全局收敛性。

命题 5: O-BELM 采样器 (18) 具有 (a) **零稳定性** 和 (b) **全局收敛性**。

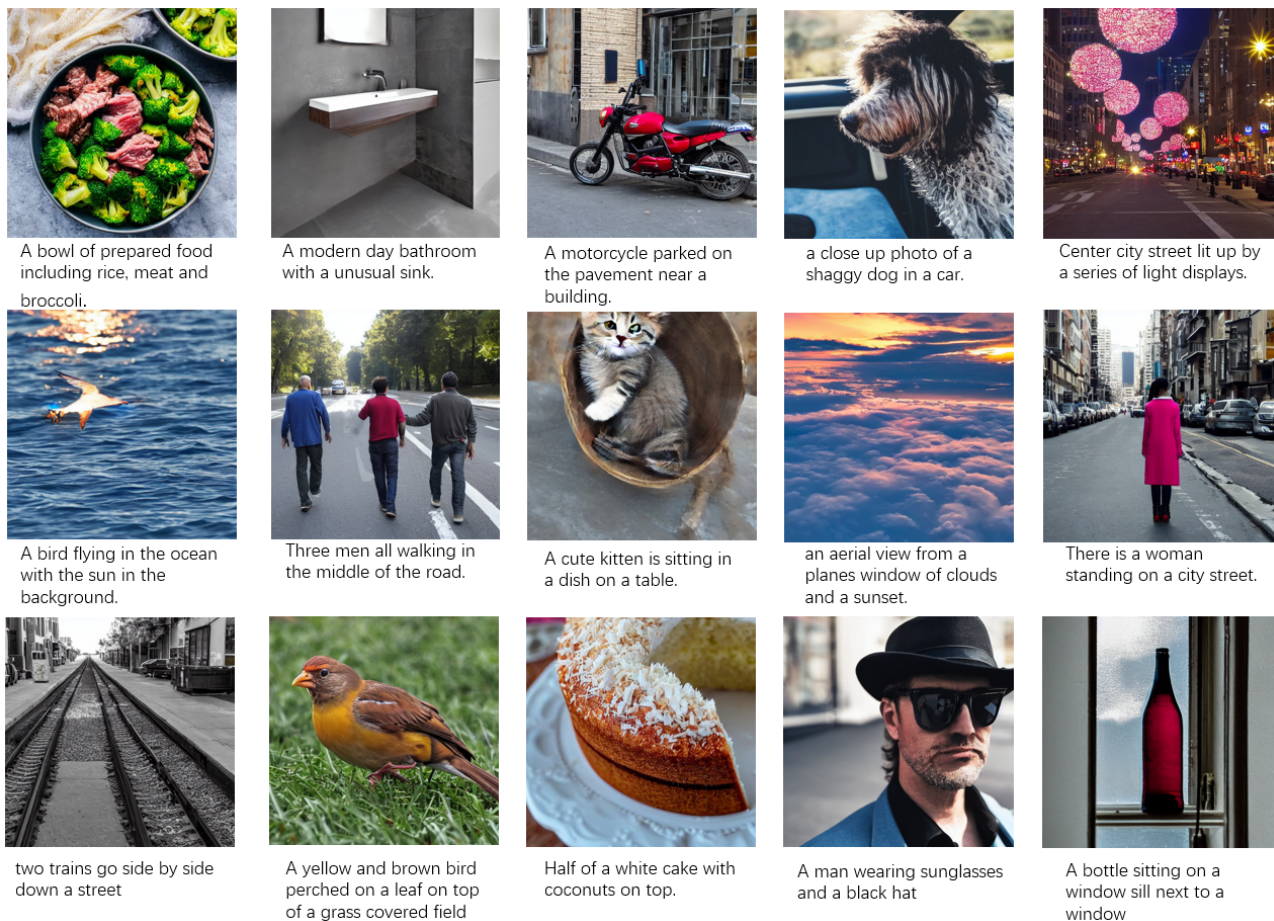


图 3. O-BELM 采样器 100 步下使用 COCO-14 数据集提示在 SD-1.5 生成的图像。

4 复现细节

4.1 与已有开源代码对比

本文所复现的工作是 Wang 等人 [37] 提出的 O-BELM 采样器，参考了论文所提供的官方实现代码。本文在复现完进行实验的过程中，发现官方提供的代码有以下几处缺陷，并进行了解决。

缺陷 1：官方代码在读取原图片时，如果原图的 png 格式，则读取会报错。

解决：读取 png 格式的图片时，官方代码读取方式读取的通道数为 4，需要更改代码将图片读取为 RGB 格式的 3 通道或者在官方代码读取后取前 3 通道，否则没有办法输入进模型。

缺陷 2：在进行条件图像生成实验中，需要根据 COCO-14 验证集的文本提示生成图像，但是官方实现代码并没有提供读取 COCO-14 验证集文本提示的代码。

解决：自行实现了读取 COCO-14 验证集文本提示的代码。

除了修改缺陷外，对于创新方面的代码全由自己实现。

4.2 实验环境搭建

本实验在 Nvidia V100 32G 8 卡服务器上进行，Python 基本环境如下：

Python==3.8 pytorch==2.0.0 cuda==11.8

其余环境依赖请见代码文件中的 requirements.txt 文件

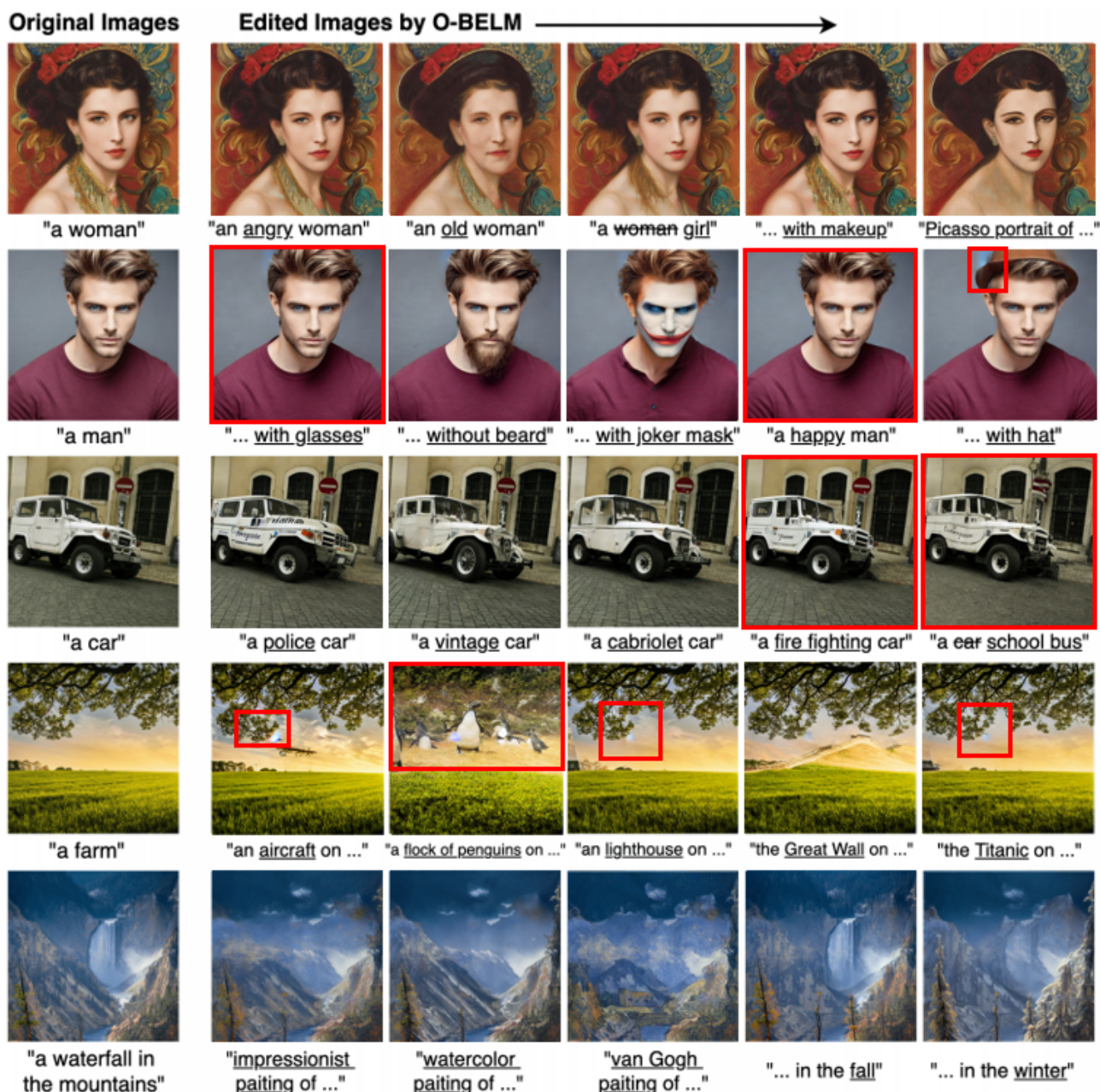


图 4. 在合成图像和真实图像上使用 O-BELM 编辑结果的示例。我们复现了 O-BELM 在一系列任务中的多样化编辑能力，包括人脸修改、内容更改、实体添加和全局样式转换。

4.3 预训练模型、数据集与评价指标

本文复现使用的所有预训练模型都是开源的，可以在网上找到，如下所示：

CIFAR10 generation : ddpm_ema_cifar10

https://github.com/VainF/Diff-Pruning/releases/download/v0.0.1/ddpm_ema_cifar10.zip

CelebA-HQ generation and interpolation : ddpm-ema-celebahq-256

<https://huggingface.co/google/ddpm-ema-celebahq-256>

Text-to-Image generation : stable-diffusion-v1-5, stable-diffusion-2-base

<https://huggingface.co/runwayml/stable-diffusion-v1-5>

<https://huggingface.co/stabilityai/stable-diffusion-2-base>

本文复现使用的数据集是三个公开的数据集，分别是 MS-COCO-2014 验证集 [20] 用于图像反转的重建实验；CIFAR10 和 CelebA-HQ 数据集用于采样质量实验。

本文复现使用的评价指标有两个：对于重建实验，使用均方误差（MSE）作为评价指标，计算原图和重建后图片的差异；对于采样质量实验，使用弗雷歇起始距离（FID）[11] 计算真实数据和生成数据之间的弗雷歇距离，测量生成图像的真实性。



图 5. Direct Inversion 结合 O-BELM 采样器编辑结果的示例。

4.4 创新点

针对直接通过修改文本并利用 O-BELM 采样器编辑后的图片容易保持不变或者生成不了想要的内容这个问题，我们觉得应该是前向去噪过程中文本的控制性不足引起的。据此，我们发现 Direct Inversion [15] 这个工作可以有效提升反转图像编辑的效果，所有我们尝试将 Direct Inversion 中的 DDIM 采样器换成复现的 O-BELM 采样器，发挥这两个方案的优势并验证其编辑效果。

5 实验结果分析

在本章中，我们展示通过实验来验证复现的 O-BELM 采样器的效果，验证其在保持高质量采样能力的同时实现了精确的反演特性，得到的结果与原文的效果相符。同时我们也在图像编辑任务中得到了与原文不符的结果，并分析原因后提出相应的改进意见。

5.1 图像重建结果

在本节中，我们采用和原论文 [37] 中相同的实验设置，使用 MS-COCO-2014 验证集中的 10k 图像 [20] 来证明 O-BELM 的精确扩散反演特性。给定图像，计算反转延迟，并使用 SD-1.5 重建图像。均方误差（MSE）是在归一化为 $[-1, 1]$ 的像素上计算的，并在 10k 幅图像上进行平均。SD 流水线中的自动编码器（AE）重建误差作为下限。像素级的重建误差结果如表2中所示，潜在空间的重建误差结果如表3所示。

表 2. 不同采样器对 COCO-14 MSE 像素级重建损失的比较

	MSE 像素级重建损失					
	DDIM	AE	EDICT	BDIA	O-BELM(原文)	O-BELM(复现)
10 steps	0.026	0.004	0.004	0.004	0.004	0.0039
20 steps	0.016	0.004	0.004	0.004	0.004	0.0039
50 steps	0.008	0.004	0.004	0.004	0.004	0.0038
100 steps	0.007	0.004	0.004	0.004	0.004	0.0038

表 3. 不同采样器对 COCO-14 MSE 潜在空间重建损失的比较

	MSE 潜在空间重建损失					
	DDIM	AE	EDICT	BDIA	O-BELM(原文)	O-BELM(复现)
10 steps	0.414	0.000	0.000	0.000	0.000	0.000
20 steps	0.243	0.000	0.000	0.000	0.000	0.000
50 steps	0.063	0.000	0.000	0.000	0.000	0.000
100 steps	0.041	0.000	0.000	0.000	0.000	0.000

从表2和表3中，我们观察到，无论在像素级还是在潜在空间，无论步长如何，复现的 O-BELM 优化器始终达到最低的 MSE，这表明它们在潜在水平上完全反转，和原文结果一致。

5.2 无条件图像生成结果

在本节中，我们进行了一项无条件的图像生成任务，以验证复现 O-BELM 采样器的高质量采样能力。利用预训练模型，我们在特定数量的步骤中生成 50k 个人工图像，并用真实数据计算相应的弗雷歇初始距离 (FID) 得分。较低的 FID 意味着生成的数据更真实。表5总结了 CIFAR10 和 Celebra-HQ 数据集的计算 FID 评分。可以看出在 CIFAR10 数据集上，复现的 O-BELM 采样器的 FID 评分与原文中的结果相差较大，原因可能是 CIFAR10 数据集分辨率较低，导致计算的 FID 波动较大。在 Celebra-HQ 数据集上，复现的 O-BELM 采样器的 FID 评分与原文中的结果则相符。

表 4. 不同采样器对无条件生成任务的 FID 评分 (\downarrow) 的比较

	CIFAR10 (32×32)			CelebA-HQ (256×256)		
	DDIM	O-BELM(原文)	O-BELM(复现)	DDIM	O-BELM(原文)	O-BELM(复现)
10 steps	17.45	10.98	18.82	27.13	19.13	18.94
20 steps	10.60	7.17	16.33	16.33	11.54	11.44
50 steps	6.96	5.24	13.51	10.77	10.41	10.72
100 steps	5.72	4.18	12.64	10.19	10.17	11.49

5.3 条件图像生成结果

在本节中，我们在条件图像生成任务下进一步评估复现的 O-BELM 采样器。我们使用 StableDiffusion V1.5 和 V2 基础模型，根据 COCO-14 验证集的文本提示生成分辨率为 512×512 的 30k 图像。所有方法都使用相同的种子和相同的文本提示集。从表??中可以明显看出，复现的 O-BELM 采样器在条件图像生成任务下的采样质量和原文相符。根据文本提示生成图像的部分结果可视化如图3所示。

表 5. 使用预训练的稳定扩散模型对文本到图像生成任务的 FID 评分 (\downarrow)

	SD-1.5 (512×512)			SD-2.0-base (512×512)		
	DDIM	O-BELM(原文)	O-BELM(复现)	DDIM	O-BELM(原文)	O-BELM(复现)
10 steps	21.44	18.19	19.32	20.40	17.01	17.94
20 steps	19.45	17.92	18.85	18.57	16.53	17.74
50 steps	18.93	17.96	18.33	17.82	16.52	17.32
100 steps	18.83	18.19	18.47	17.64	16.75	16.97

5.4 免训练图像编辑结果

在本节中，我们使用复现的 O-BELM 采样器在 SD1.5 扩散模型上进行图像编辑任务，得到的结果如图4所示。从结果图中可以看到，对于图片中物品变化较大的编辑，免训练的方式并不能很好的进行正确的编辑，对于细微的编辑，这种编辑方式也不能保证生成图像的稳定，即有可能会出现蓝色伪影（图中用红框框出）。但从编辑的任务中确实可以体现 O-BELM 的精确反转特性可以在保留辅助细节（第一行的背景、第二行的发型、第三行的交通标志、第四行的树木和作物、最后一行的构图）的同时进行图像更改。

将 Direct Inversion 中的 DDIM 采样器换成复现的 O-BELM 采样器，部分结果如图5所示，可以看出并没有达到很好的成效，但相比只使用 O-BELM 采样器会有细微变化。

6 总结与展望

综上所述，我们根据论文复现出了最优双向显式线性多步 (O-BELM) 采样器。而后，我们在 COCO 数据集上进行了图像重建实验，得到了和论文一致的精确反演。此外，我们还在 CIFAR10、CelebA-HQ 数据集上预训练的扩散模型中进行无条件图像生成以及 SD1.5 和 SD2.0 这类大规模预训练的扩散模型中进行条件图像生成实验，也得到了和论文中相同的采样质量。最后在复现图像编辑下游任务中，我们发现了直接通过修改文本并利用 O-BELM 采样器编辑后的图片容易保持不变或者生成不了想要的内容，这与论文中的效果有所出入，因此本工作也根据此问题采样 Direct Inversion 结合 O-BELM 采样器进行编辑的创新改进。O-BELM 采样器的精确反转特性在编辑任务中可以在保留辅助细节的同时进行图像更改，虽然只使用文本引导并不能进行准确编辑，并且 Direct Inversion 结合 O-BELM 采样器编辑结果也没有达到很好的成效，但未来或许可以使用掩码辅助或者微调的方式更好地发挥 O-BELM 采样器的优势。

参考文献

- [1] Brian DO Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982.
- [2] Francis Bashforth and John Couch Adams. *An attempt to test the theories of capillary action by comparing the theoretical and measured forms of drops of fluid*. University Press, 1883.
- [3] Robin Bombach, Andreas Blatmann, Dominik Lorenz, Patrick Esser, and Bjorn Omer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10684–10695, 2022.
- [4] John Charles Butcher. *Numerical methods for ordinary differential equations*. John Wiley & Sons, 2016.
- [5] Hyungjin Chung, Byeongsu Sim, Dohoon Ryu, and Jong Chul Ye. Improving diffusion models for inverse problems using manifold constraints. *Advances in Neural Information Processing Systems*, 35:25683–25696, 2022.
- [6] M Crouzeix and FJ Lisbona. The convergence of variable-stepsize, variable-formula, multistep methods. *SIAM journal on numerical analysis*, 21(3):512–534, 1984.
- [7] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems*, 34:8780–8794, 2021.
- [8] Laurent Dinh, David Krueger, and Yoshua Bengio. Nice: Non-linear independent components estimation. *arXiv preprint arXiv:1410.8516*, 2014.
- [9] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*, 2016.
- [10] Amir Hertz, Ron Mokady, Jay Tenenbaum, Kfir Aberman, Yael Pritch, and Daniel Cohen-or. Prompt-to-prompt image editing with cross-attention control. In *International Conference on Learning Representations*, 2023.
- [11] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.
- [12] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.
- [13] Jonathan Ho, Chitwan Saharia, William Chan, David J Fleet, Mohammad Norouzi, and Tim Salimans. Cascaded diffusion models for high fidelity image generation. *Journal of Machine Learning Research*, 23(47):1–33, 2022.

- [14] Seongmin Hong, Kyoonghyun Lee, Suh Yoon Jeon, Hyewon Bae, and Se Young Chun. On exact inversion of dpm-solvers. *arXiv preprint arXiv:2311.18387*, 2023.
- [15] Xuan Ju, Ailing Zeng, Yuxuan Bian, Shaoteng Liu, and Qiang Xu. Pnp inversion: Boosting diffusion-based editing with 3 lines of code. *International Conference on Learning Representations (ICLR)*, 2024.
- [16] Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. *arXiv preprint arXiv:2206.00364*, 2022.
- [17] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *Advances in Neural Information Processing Systems*, 35:23593–23606, 2022.
- [18] Bahjat Kawar, Gregory Vaksman, and Michael Elad. Snips: Solving noisy inverse problems stochastically. *Advances in Neural Information Processing Systems*, 34:21757–21769, 2021.
- [19] Diederik Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. In *Advances in Neural Information Processing Systems*, volume 34, pages 21696–21707. Curran Associates, Inc., 2021.
- [20] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollar, and CY2002rec Zitnick. Microsoft coco: Common objects in context. In *Computer Vision-ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pages 740–755. Springer, 2014.
- [21] Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Dpm-solver: A fast ode solver for diffusion probabilistic model sampling in around 10 steps. *Advances in Neural Information Processing Systems*, 35:5775–5787, 2022.
- [22] Barak Meiri, Dvir Samuel, Nir Darshan, Gal Chechik, Shai Avidan, and Rami Ben-Ari. Fixed-point inversion for text-to-image diffusion models. *arXiv preprint arXiv:2312.12540*, 2023.
- [23] William Edmund Milne. Numerical integration of ordinary differential equations. *The American Mathematical Monthly*, 33(9):455–460, 1926.
- [24] Ron Mokady, Amir Hertz, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or. Null-text inversion for editing real images using guided diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6038–6047, 2023.
- [25] Forest Ray Moulton. *New methods in exterior ballistics*. University of Chicago Press, 1926.
- [26] Alexander Quinn Nichol, Prafulla Dhariwal, Aditya Ramesh, Pranav Shyam, Pamela Mishkin, Bob Megrew, Ilya Sutskever, and Mark Chen. Glide: Towards photorealistic

- image generation and editing with text-guided diffusion models. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162, pages 16784–16804, 2022.
- [27] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 2022.
 - [28] Chitwan Saharia, William Chen, Saurabh Saxena, Lala Li, Jay Whang, Emily L Denton, Kamyar Ghasemipour, Raphael Gontijo Lopes, Burcu Karagol Ayan, Tim Salimans, et al. Photorealistic text-to-image diffusion models with deep language understanding. In *Advances in Neural Information Processing Systems*, volume 35, pages 36479–36494, 2022.
 - [29] Jascha Sohl-Dickstein, Eric Weiss, Nira Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, pages 2266–2265. PMLR, 2015.
 - [30] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021.
 - [31] Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. In *Proceedings of the 40th International Conference on Machine Learning, ICML’23*. JMLR.org, 2023.
 - [32] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32, 2019.
 - [33] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.
 - [34] Xuan Su, Jiaming Song, Chenlin Meng, and Stefano Ermon. Dual diffusion implicit bridges for image-to-image translation. *arXiv preprint arXiv:2203.08382*, 2022.
 - [35] Endre Suli and David F Mayers. *An introduction to numerical analysis*. Cambridge university press, 2003.
 - [36] Bram Wallace, Akash Gokul, and Nikhil Naik. Edict: Exact diffusion inversion via coupled transformations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22532–22541, 2023.
 - [37] Fangyikang Wang, Hubery Yin, Yuejiang Dong, Huminhao Zhu, Chao Zhang, Hanbin Zhao, Hui Qian, and Chen Li. Belm: Bidirectional explicit linear multi-step sampler for exact inversion in diffusion models. *arXiv preprint arXiv:2410.07273*, 2024.
 - [38] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Ergan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*, pages 0–0, 2018.

- [39] Guoqiang Zhang, Jonathan P Lewis, and W Bastiaan Kleijn. Exact diffusion inversion via bi-directional integration approximation. *arXiv preprint arXiv:2307.10829*, 2023.