arXiv:2305.03472v2 [cs.MM] 6 Sep 2023

生成式隐写扩散

Ping Wei
p
wei17@fudan.edu.cn 复旦大学 中国上海

Qing Zhou
21110240055
@m.fudan.edu.cn 复旦大学 中国上海

Zichi Wang
w
angzichi@shu.edu.cn
上海大学 中国上海

Zhenxing Qian*
z
xqian@fudan.edu.cn
复旦大学 中国上海

Xinpeng Zhang*
zhangxinpeng
@fudan.edu.cn 复旦大学 中国上海

Sheng Li
lisheng
@fudan.edu.cn 复旦大学，中国上海

摘要

生成式隐写术（GS）是一种新兴技术，可直接从秘密数据生成隐写图像。近期已有多种基于生成对抗网络或 Flow 的生成式隐写方法问世。然而，由于缺乏网络可逆性，现有基于生成对抗网络的生成式隐写方法无法完整恢复隐藏的秘密数据；而基于 Flow 的方法则因每个模块受到严格的可逆性限制，生成的图像质量较差。为解决这一问题，我们设计了一种名为“生成式隐写扩散”（GSD）的新型生成式隐写方案，并提出了名为“StegoDiffusion”的可逆扩散模型。该模型不仅能生成逼真的隐写图像，还能实现对隐藏秘密数据的 100% 恢复。所提出的 StegoDiffusion 模型利用非马尔可夫链和快速采样技术，高效生成隐写图像。通过基于 StegoDiffusion 生成过程中转移概率构建常微分方程（ODE），秘密数据与隐写图像可通过 ODE 的近似求解器——欧拉迭代公式相互转换，从而能够使用不可逆但表达能力更强的网络结构实现模型可逆性。我们提出的 GSD 兼具可逆性和高性能，在所有指标上均显著优于现有生成式隐写方法。

CCS 概念

• **信息系统** → 多媒体信息系统； • **安全与隐私** → 安全服务。

关键词

S隐写术、生成式隐写术、扩散 模型

1 引言

隐写术通过将秘密数据隐藏在载体媒体中来实现隐蔽通信。它可应用于各种类型的媒体，包括数字图像 [14, 23], 视频 [9, 16], 音频 [26], 和文本 [13]。其中，数字图像是最常用的载体媒介，而包含隐藏数据的图像被称为隐写图像。隐写术的目的是防止未获授权的人员发现秘密数据。

防止秘密数据传输遭到未授权访问。隐写分析 [3, 22]可检测载体图像或隐写图像是否包含隐藏数据。传统隐写术可以被隐写分析检测到，因此提出生成式隐写术（GS）作为替代方法，以增强安全性。

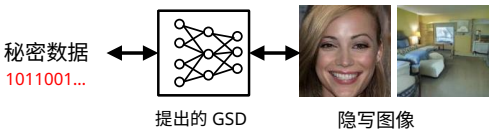


图 1：具有可逆性的生成式隐写术。我们提出的 GSD方案可以将二进制秘密数据转换为隐写图像，然后完整恢复隐藏的秘密数据。

Liu 等人 [10] 提出了一种生成式隐写方法，将秘密数据隐藏在生成对抗网络的类别标签中，通过生成不同类别的隐写图像来表示秘密数据。类似地，Hu 等人 [6] 和 Liu 等人 [11] 将秘密数据映射到生成对抗网络不同的潜在值范围，利用图像生成器生成隐写图像，并训练提取器以提取隐藏数据。然而，与传统的基于载体修改的隐写方法相比，这些方法在载荷、秘密数据提取准确率和图像质量方面表现较差。随后，为了提高秘密数据的载荷，Wei 等人 [24] 提出了一种生成器，在图像生成过程中将秘密数据直接集成到卷积特征图中，从而生成载荷更高、质量更好的隐写图像，并训练提取器来提取隐藏数据。然而，该方法在高载荷容量下的数据提取准确性相对较低。为提高提取准确率，Wei 等人 [25] 基于 Flow [7] 模型将生成器和提取器集成到一个网络中，建立了隐写图像与秘密数据之间的双向映射。然而， [25] 生成的隐写图像质量略低于 [24]生成的图像。这是因为 Flow 模块对可逆性的要求降低了网络的能力。为取得更好的性能，我们通过构建可逆扩散模型，提出了一种创新的生成式隐写方案 GSD。如图 1 所示，该方案可以将二进制秘密数据转换为量化隐写图像，并准确恢复隐藏的秘密数据。与采用可逆网络结构的 Flow 方法不同，GSD 构建了一个常微分方程（ODE），以确保模型的可逆性。其中，隐写图像生成和秘密数据提取通过 ODE 的近似求解器——欧拉迭代公式完成。这使得模型能够使用不可逆但表达能力强的网络，从而大幅提升模型性能。

扩散模型目前是最有前景的图像生成方法 [1, 27]。其图像质量超过了生成对抗网络和

*Xinpeng Zhang 和 Zhenxing Qian 是通讯作者。

展示出惊艳的视觉效果 [2, 15, 18–20]。扩散模型由前向扩散过程和反向图像生成过程组成。在扩散过程中，逐渐向图像中添加噪声，得到符合高斯分布的含噪图像。在生成过程中，逐渐对服从高斯分布的含噪图像进行去噪，从而得到清晰的自然图像。DDPM [5] 利用扩散模型生成高分辨率图像，推动了扩散模型成为当前趋势。然而，DDPM 的图像生成步骤过多，耗时过长。为了更有效地采样，Song 等提出 DDIM [21], ，利用非马尔可夫链缩短采样步数。OpenAI 采用分类引导的条件图像生成方法改进 DDPM，使其图像质量超越生成对抗网络生成的图像 [2]。他们还推出 GLIDE 模型 [15], ，利用 CLIP [17] 模型或无分类器引导，通过文本引导图像生成。2022年，OpenAI 发布 DALLE-2 [18],，这是一种两阶段模型，先根据文本描述生成 CLIP 图像嵌入，再使用解码器生成图像。随后，Google 提出 Imagen [20],，这是一种文本到图像的扩散模型，通过语言模型将输入文本转换为文本嵌入，并利用稳健的级联扩散模型生成高质量图像。

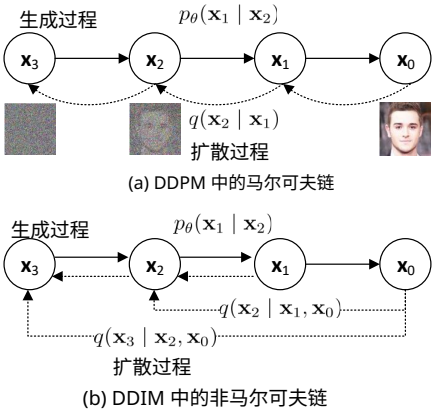


图2：存在两种不同的扩散方法。子图(a)展示了DDPM中马尔可夫链的扩散模式，其中$\mathbf{x}_t$的生成仅取决于$\mathbf{x}_{t-1}$。子图(b)展示了DDIM中的非马尔可夫扩散方法，其中$\mathbf{x}_t$的生成取决于$\mathbf{x}_{t-1}$和$\mathbf{x}_0$。

本文首先开发了一种用于生成式隐写术的新型扩散模型（命名为 StegoDiffusion），随后提出了一种基于所提出的 StegoDiffusion 和 DCT/IDCT 变换的新型生成式隐写方法，称为“生成式隐写扩散”（GSD）。数据发送方可以通过 DCT/IDCT 变换将秘密数据隐藏在所提出的 StegoDiffusion 的输入潜变量中，然后通过模型的生成过程生成高质量的隐写图像。为了提取隐藏数据，接收方可以将隐写图像输入 StegoDiffusion，该模型通过扩散过程恢复输入的隐写潜变量，然后从恢复后的潜变量的 DCT 系数中提取秘密数据。我们提出的 GSD 方案的主要创新如下：

1) 我们首先提出一种用于生成式隐写术的可逆扩散模型，该模型能够将秘密数据转换为隐写图像，并随后恢复隐藏的秘密数据。

2) 我们构建一个常微分方程（ODE），并使用其

近似求解器（欧拉迭代公式），以可逆的方式生成隐写图像，从而能够使用不可逆的网络结构实现可逆的隐写图像生成。

3）我们探索使用非马尔可夫链结合快速采样技术，以加快隐写图像的生成。

4）我们的 GSD 方案在所有指标上均优于现有的生成式隐写方法，包括载荷、提取准确率、图像质量和安全性。

2 预备知识

2.1 DDPM

DDPM [5] 涉及两个马尔可夫链：扩散过程和生成过程，如图 2a所示。扩散过程向原始图像$\mathbf{x}_0$ 添加噪声，并通过转移概率$q(\mathbf{x}_t \mid \mathbf{x}_{t-1})$生成接近高斯分布的噪声图像$\mathbf{x}_T$。另一方面，生成过程使用反向马尔可夫链 $p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t)$，从高斯噪声图像$\mathbf{x}_T$中获得清晰图像$\mathbf{x}_0$。

输入图像分布为$\mathbf{x}_0 \sim q(\mathbf{x}_0)$，扩散过程通过转移概率 $q(\mathbf{x}_t \mid \mathbf{x}_{t-1})$生成与$\mathbf{x}_0$ 维度相同的中间潜变量$\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_T$：

$$q(\mathbf{x}_t \mid \mathbf{x}_{t-1}) = N\left(\mathbf{x}_t; \sqrt{\alpha_t}\mathbf{x}_{t-1}, \beta_t \mathbf{I}\right), \tag{1}$$

其中，$\alpha_t \in (0,1)$是一个超参数，且 $\beta_t = 1 - \alpha_t$。特别地，当 $\bar{\alpha}_t = \prod_{i=0}^{t} \alpha_i$时，可以推导出：

$$q(\mathbf{x}_t \mid \mathbf{x}_0) = N\left(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I}\right). \tag{2}$$

展开上述公式，可将$\mathbf{x}_t$ 表示为$\mathbf{x}_0$ 与噪声变量 ϵ的线性组合：

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, \tag{3}$$

其中，$\epsilon \sim N(0,1)$是符合高斯分布的噪声。

图像生成过程通过可学习的马尔可夫链，逐步从随机噪声图像$\mathbf{x}_T \sim N(0,1)$中去除噪声，其转移概率为：

$$p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t) = N\left(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t, t), \Sigma(\mathbf{x}_t, t)\right), \tag{4}$$

其中，θ 表示可学习参数。如果学习到的反向转移概率 $p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t)$ 接近 $q(\mathbf{x}_{t-1} \mid \mathbf{x}_t)$，则可以实现生成过程。

扩散模型通过最大化似然函数 $\max_\theta \log p_\theta(\mathbf{x}_0)$ 进行训练，但 $p_\theta(\mathbf{x}_0)$ 无法直接优化。因此，DDPM 使用变分推断来优化负对数似然的通常变分下界：

$$-\mathbb{E}_{\mathbf{x}_0} \log p_\theta(\mathbf{x}_0) \le \mathbb{E}_{q(\mathbf{x}_{0:T})}\left[\log \frac{q(\mathbf{x}_{1:T} \mid \mathbf{x}_0)}{p_\theta(\mathbf{x}_{0:T})}\right] = L_{VLB}, \tag{5}$$

如果将以$\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{x}_0$为条件的扩散过程 q 和生成过程 $p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t)$ 建模为具有可训练均值函数和固定方差的高斯分布，则公式 5 中的目标函数可简化为：

$$L_{simple} = \mathbb{E}_{\mathbf{x}_0, \epsilon}\left[\left\|\epsilon - \epsilon_\theta\left(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, t\right)\right\|^2\right], \tag{6}$$

其中， $\epsilon_\theta(\cdot)$ 是生成过程中预测的噪声值，ϵ 是扩散过程中注入的噪声值。

2.2 DDIM

为了更有效地采样，DDIM [21]定义了一个非马尔可夫链来实现前向扩散过程，如图 2b 所示。前向扩散过程可由贝叶斯定理推导得出：

$$q_\sigma(\mathbf{x}_t \mid \mathbf{x}_{t-1}, \mathbf{x}_0) = \frac{q_\sigma(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{x}_0)\, q_\sigma(\mathbf{x}_t \mid \mathbf{x}_0)}{q_\sigma(\mathbf{x}_{t-1} \mid \mathbf{x}_0)}. \tag{7}$$

其中，$\mathbf{x}_t$ 的生成在非马尔可夫扩散过程中同时依赖于 $\mathbf{x}_0$ 和 $\mathbf{x}_{t-1}$。

给定输入图像 $\mathbf{x}_0$ 和 $\epsilon_t \sim N(0,1)$，可以通过前向扩散过程中的公式 3 获得$\mathbf{x}_t$。训练完成后，神经网络将噪声 ϵ_t的值预测为 $\epsilon_\theta(\mathbf{x}_t, t)$。因此，可以使用 $\mathbf{x}_t$和 $\epsilon_\theta(\mathbf{x}_t, t)$预测原始输入图像 $\mathbf{x}_0$：

$$\mathbf{x}_0' = \frac{\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\epsilon_\theta(\mathbf{x}_t, t)}{\sqrt{\bar{\alpha}_t}}. \tag{8}$$

图像生成过程中的转移概率 $p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t)$ 可通过扩散过程中的转移概率 $q_\sigma(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{x}_0 = \mathbf{x}\prime_0)$ 进行近似。因此，可以通过从 $\mathbf{x}_t$ 中采样得到 $\mathbf{x}_{t-1}$：

$$\mathbf{x}_{t-1} = \sqrt{\bar{\alpha}_{t-1}} \underbrace{\left(\frac{\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\epsilon_\theta(\mathbf{x}_t, t)}{\sqrt{\bar{\alpha}_t}}\right)}_{\text{predicted } \mathbf{x}_0} + \underbrace{\sqrt{1-\bar{\alpha}_{t-1}-\sigma_t^2}\cdot\epsilon_\theta(\mathbf{x}_t, t)}_{\text{direction pointing to } \mathbf{x}_t} + \underbrace{\sigma_t\epsilon_t}_{\text{random noise}}, \tag{9}$$

其中 $\sigma_t = \eta\sqrt{(1-\bar{\alpha}_{t-1})/(1-\bar{\alpha}_t)}\sqrt{1-\bar{\alpha}_t/\bar{\alpha}_{t-1}}, \eta \in [0,1]$. $\epsilon_t \sim N(0,1)$ $\epsilon_\theta(\mathbf{x}_t, t)$，而 ϵ_t 是 $\mathbf{x}_t$ 的预测值。

DDIM 仍使用变分下界 J_σ 来近似图像分布的似然函数：

$$-\mathbb{E}_{\mathbf{x}_0} \log p_\theta(\mathbf{x}_0) = J_\sigma = \gamma \cdot L_{simple} + C, \tag{10}$$

其中 $\gamma \in R^+$ 和 $C \in R$， L_{simple} 是公式 6 中 DDPM 的简化损失函数。因此，DDIM 中的变分下界 J_σ 可以使用与 DDPM 相同的优化目标，训练过程也相同。

3 提出的方案

本文提出了一种名为 GSD 的新型生成式隐写方案，其中开发了一种可逆扩散模型（名为 StegoDiffusion），用于可逆隐写图像合成。StegoDiffusion 采用确定性非马尔可夫链，并结合快速采样技术来加快图像生成。在该方案中，隐写图像与秘密数据通过欧拉迭代相互转换。此外，采用离散余弦变换/离散余弦逆变换，以减小图像量化的影响并提高数据提取准确性。

图 3 展示了所提出的 GSD 的整体框架，主要包括两个重要部分：秘密数据的隐藏/提取和预训练的 StegoDiffusion 模型。在生成隐写图像时，首先将秘密数据 **d** 隐藏在潜变量 **z**的离散余弦变换系数中，然后通过离散余弦逆变换获得隐写潜变量 $\mathbf{z}_s$。

算法 1 GSD 中 StegoDiffusion 的训练策略。

1: **输入**：超参数 $\bar{\alpha}_t$ ($t \in [1, T]$)；神经网络 U-net；采样步数 T。
2: **对**每个步骤**执行**
3: 随机采样输入图像 $\mathbf{x}_0 \sim q(\mathbf{x}_0)$；
4: 向 $\mathbf{x}_0$ 添加随机噪声，$\mathbf{x}_0 := \mathbf{x}_0 + \mathbf{n}$，其中 $\mathbf{n} \sim N(0, 0.01_2)$；
5: 随机采样扩散步 $t \sim U(1, T)$；
6: 随机采样注入噪声 $\epsilon \sim N(0,1)$；
7: 计算损失 L_{simple} 并通过梯度更新 U-net：$\nabla_\theta \left\|\epsilon - \epsilon_\theta\left(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, t\right)\right\|^2$。
8: **输出**：训练好的 StegoDiffusion 模型。

变换。接下来，将 $\mathbf{z}_s$ 用作初始噪声图像 $\mathbf{x}_S$，然后采用快速采样技术生成隐写图像 **x0**，将图像生成的采样步长从原来的 $T = 1000$ 减少至 S（S远小于 T），其中采样图像序列为：$\mathbf{x}_S \rightarrow \mathbf{x}_{S-1} \cdots \rightarrow \mathbf{x}_2 \rightarrow \mathbf{x}_1 \rightarrow \mathbf{x0}$。经过量化（将像素值转换为 [0, 255]范围内的量化图像）后，即可生成最终的隐写图像 $\mathbf{x}^s$。提取隐藏数据时，将预处理后的 $\mathbf{x}^s$ （将量化后的隐写图像转换为浮点图像，像素值范围为 [-1, 1]）输入 StegoDiffusion，通过扩散过程逐步恢复初始噪声图像，得到 $\mathbf{x}^s \rightarrow \mathbf{x}_0 \rightarrow \mathbf{x}1 \cdots \rightarrow \mathbf{x}_{S-1} \rightarrow \mathbf{x}_S$。然后，将 $\mathbf{x}_S$用作恢复的隐写潜变量 $\mathbf{z}\prime_s$，并从 $\mathbf{z}\prime_s$的 DCT 系数中提取隐藏数据 $\mathbf{d}\prime$。

我们方案中的 StegoDiffusion 采用与 DDPM 类似的训练策略。主要区别在于，为确保数据提取准确性，我们在图像预处理阶段向$\mathbf{x}_0$ 添加了高斯噪声。StegoDiffusion 的训练策略如算法 1 所示。首先必须确定超参数 $\bar{\alpha}_t$。每一步都会随机采样一张图像$\mathbf{x}_0$、一个时间步 t以及注入噪声 ϵ。神经网络 U-net 输出 $\epsilon_\theta(\mathbf{x}_t, t)$，作为此时注入噪声 ϵ的预测值。随后更新 U-net 的参数，使预测噪声 $\epsilon_\theta(\mathbf{x}_t, t)$ 接近注入噪声 ϵ，其中根据公式 3，将$\mathbf{x}_t$替换为 $\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon$。

3.1 StegoDiffusion 的可逆性

StegoDiffusion 的生成过程和扩散过程经过设计，具有可逆性，能够从生成的隐写图像中准确地提取数据。在公式 9 中，添加的噪声的~~方差为~~$\sigma_t = \eta\sqrt{(1-\bar{\alpha}_{t-1})/(1-\bar{\alpha}_t)}\sqrt{1-\bar{\alpha}_t/\bar{\alpha}_{t-1}}$，其中 $\eta \in [0,1]$。当$\eta = 0$时，添加的噪声为零，生成过程的采样变为确定性过程。

$$\mathbf{x}_{t-1} = \frac{\mathbf{x}_t}{\sqrt{\alpha_t}} + \left(\sqrt{1-\bar{\alpha}_{t-1}} - \sqrt{\frac{1-\bar{\alpha}_t}{\alpha_t}}\right)\epsilon_\theta(\mathbf{x}_t, t), \tag{11}$$

其中，$\mathbf{x}_{t-1}$可以由$\mathbf{x}_t$生成。当迭代次数为 T且$\mathbf{x}_T \sim N(0,\ 1)$时，可以生成一系列内容确定性图像，$\mathbf{x}_T \rightarrow \cdots \mathbf{x}_t \rightarrow \mathbf{x}_{t-1} \cdots \rightarrow \mathbf{x}0$。

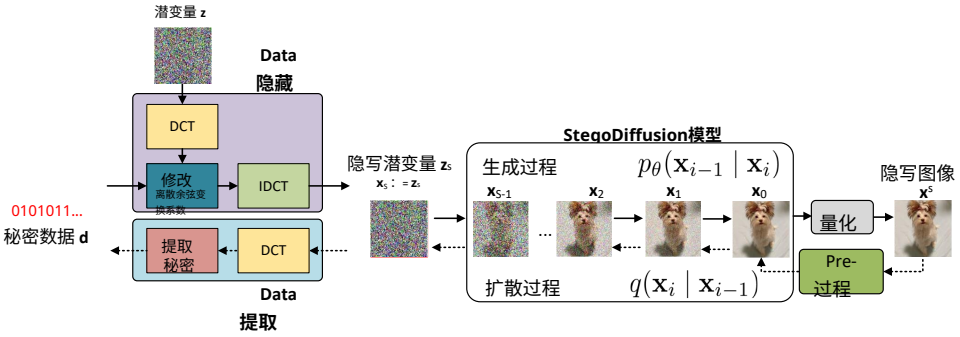


图 3：所提出的生成式隐写术 GSD 方案流程图。

通过重新整理上面的公式 11，图像生成过程中的转移概率 $p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t)$ 为：

$$\frac{\mathbf{x}_{t-1}}{\sqrt{\bar{\alpha}_{t-1}}} = \frac{\mathbf{x}_t}{\sqrt{\bar{\alpha}_t}} + \left(\sqrt{\frac{1-\bar{\alpha}_{t-1}}{\bar{\alpha}_{t-1}}} - \sqrt{\frac{1-\bar{\alpha}_t}{\bar{\alpha}_t}}\right)\epsilon_\theta\left(\mathbf{x}_t, t\right), \tag{12}$$

令 $\Delta t = 1$，公式 12 可写为：

$$\frac{\mathbf{x}_{t-\Delta t}}{\sqrt{\bar{\alpha}_{t-\Delta t}}} - \frac{\mathbf{x}_t}{\sqrt{\bar{\alpha}_t}} = \left(\sqrt{\frac{1-\bar{\alpha}_{t-\Delta t}}{\bar{\alpha}_{t-\Delta t}}} - \sqrt{\frac{1-\bar{\alpha}_t}{\bar{\alpha}_t}}\right)\epsilon_\theta\left(\mathbf{x}_t, t\right), \tag{13}$$

令 $\sigma = (\sqrt{1-\bar{\alpha}}/\sqrt{\bar{\alpha}})$ 和 $\bar{\mathbf{x}} = \mathbf{x}/\sqrt{\bar{\alpha}}$（$\sigma$ 和 $\bar{\mathbf{x}}$ 均为 t 的函数），则公式 13 可视为一个由欧拉方法求解的常微分方程（ODE）：

$$\mathrm{d}\bar{\mathbf{x}}(t) = \epsilon_\theta\left(\frac{\bar{\mathbf{x}}(t)}{\sqrt{\sigma^2+1}}, t\right)\mathrm{d}\sigma(t), \tag{14}$$

其中，$\bar{\mathbf{x}}_{t+1} = \bar{\mathbf{x}}_t + \mathrm{d}\bar{\mathbf{x}}(t)$，且 $\bar{\mathbf{x}}_{t-1} = \bar{\mathbf{x}}_t - \mathrm{d}\bar{\mathbf{x}}(t)$。然后，根据 $\mathbf{x}_t = \sqrt{\bar{\alpha}}\bar{\mathbf{x}}_t$，可使用 ODE 求解器求出任意时间 t 的图像 $\mathbf{x}_t$。这意味着所提出的 GSD 方案中使用的 StegoDiffusion模型具有可逆性。

若在公式 13 中设为 $\Delta = -1$，则可得到扩散过程的转移概率$q(\mathbf{x}_{t+1} \mid \mathbf{x}_t)$：

$$\begin{aligned}\frac{\mathbf{x}_{t+1}}{\sqrt{\bar{\alpha}_{t+1}}} - \frac{\mathbf{x}_t}{\sqrt{\bar{\alpha}_t}} &= \left(\sqrt{\frac{1-\bar{\alpha}_{t+1}}{\bar{\alpha}_{t+1}}} - \sqrt{\frac{1-\bar{\alpha}_t}{\bar{\alpha}_t}}\right)\epsilon_\theta\left(\mathbf{x}_t, t\right), \\ \mathbf{x}_{t+1} &= \sqrt{\alpha_t}\mathbf{x}_t + \left(\sqrt{1-\bar{\alpha}_{t+1}} - \sqrt{\alpha_t - \bar{\alpha}_{t+1}}\right)\epsilon_\theta\left(\mathbf{x}_t, t\right).\end{aligned} \tag{15}$$

总之，StegoDiffusion 的生成过程和扩散过程是常微分方程的一种离散形式。当迭代次数为 T 且 $\mathbf{x}_T \sim N \in (0,\ 1)$ 时，带噪图像 $\mathbf{x}_T$可通过公式 12 逐步去噪，得到图像 $\mathbf{x}_0$，这就是图像生成过程。另一方面，给定 $\mathbf{x}_0$，图像 $\mathbf{x}_T$可通过公式 15 逐步恢复，这就是扩散过程。这两个过程互为可逆，$\mathbf{x}_T \leftrightarrow \mathbf{x}_{T-1} \cdots \leftrightarrow \mathbf{x}_1 \leftrightarrow \mathbf{x}_0$。

3.2 隐写图像生成

为抵抗图像量化造成的失真，我们将秘密数据隐藏在频域中。首先，将秘密数据隐藏在潜变量 $\mathbf{z}$ 的 DCT 系数中，其中潜变量 $\mathbf{z}$ 是一种服从标准高斯分布的随机噪声，$\mathbf{z} \sim N(0,1)$。

$$\begin{aligned}\mathbf{Coeff_m} &= 1.0 \odot \mathrm{reshape}(2 \times \mathbf{d} - 1), \\ \mathbf{z}_s &= \mathrm{IDCT}(\mathbf{Coef_m}),\end{aligned} \tag{16}$$

其中，$\mathbf{Coeff_m}$ 表示隐藏秘密数据后潜变量 $\mathbf{z}$ 的 DCT 系数；$\odot$ 表示逐元素相乘。运算符 $\mathrm{reshape}(\cdot)$ 将 $\mathbf{d}$ 重塑为与 $\mathbf{z}$ 的 DCT 系数相同的形状。$\mathbf{d}$ 是输入的二进制秘密数据。当需要隐藏秘密数据“1”时，$\mathbf{z}$ 对应的 DCT 系数会被修改为“+1.0”；当需要隐藏秘密数据“0”时，$\mathbf{z}$ 对应的 DCT 系数会被修改为“-1.0”。在 DCT 系数中嵌入秘密数据后，通过离散余弦逆变换可得到隐写潜变量 $\mathbf{z}_s$。

以$\mathbf{z}_s$ 作为初始噪声图像$\mathbf{x}_S$，并通过图像 StegoDiffusion 的生成过程 $p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t)$ 中，初始噪声图像可以逐渐去噪，生成一系列清晰图像，$\mathbf{x}_S := \mathbf{z}_s \rightarrow \mathbf{x}_{S-1} \cdots \mathbf{x}_2 \rightarrow \mathbf{x}_1 \rightarrow \mathbf{x}_0$。经过量化后，即可得到隐写图像$\mathbf{x}^s$。量化操作会将浮点图像$\mathbf{x}_0$ 转换为数字图像$\mathbf{x}^s$，其像素值范围为 $[0, 255]$。

StegoDiffusion 中的最大迭代步数 T 设为一个较大的值（$T = 1000$），以确保最终噪声图像$\mathbf{x}_T$接近标准高斯分布，即 $\mathbf{x}_T \sim N(0,1)$。然而，这也意味着图像生成过程需要逐步采样 T 步。漫长的采样过程会降低图像生成速度，生成每张隐写图像需要数秒甚至数十秒，这不利于隐蔽通信。

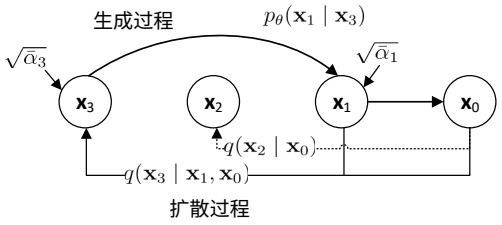


图 4：StegoDiffusion 中快速采样的示意图。展示了采样步数 $\tau = \{t_1, t_2\} = \{1, 3\}$。

可以通过缩短采样步数来加快图像生成。假设 $\tau=\{t_1,t_2,\cdots\ t_{i-1},t_i,\cdots\ t_S\}$ 是原始采样序列 $\mathrm{T}=\{1,2,\cdots\ T\}$ 的一个子序列，其中 $\tau\subseteq\mathrm{T}$。通过在加速图像生成过程中将步数从 T 减少到 S，图像生成速度会显著提升。例如，若原始步序列为 $\{1,2,\cdots,1000\}$，加速后的采样步序列为 $\{t_1=100,t_2=200,\cdots\ t_{i-1}=100\times(i-1)\ t_i=100\times i,\cdots\ t_{10}=1000\}$，图像生成所需时间仅为原来的百分之一。加速图像生成过程如图 4 所示。在图中，扩散过程跳过 x2，根据 x0 和 x1 生成 x3，扩散过程中的转移概率变为 $q(\mathrm{x}_3\mid\mathrm{x}1,\mathrm{x}0)$。图像生成时，生成过程的转移概率为$p_\theta(\mathrm{x}_1\mid\mathrm{x}3)$，也可以直接根据 x3 得到 x1，从而跳过 x2。需要注意的是，StegoDiffusion 训练期间仍需考虑每个步骤 $t\in[1,T]$，如算法 1 所示。只有在模型训练完成后，才会省略若干中间节点以加快采样过程。采样节点的超参数 $\sqrt{\bar{\alpha}_{t_i}}$由其索引值 t_i决定。在图 4 中，$\tau=\{t_1,t_2\}=\{1,3\}$，节点 $t_1=1$，因此 $\sqrt{\bar{\alpha}_{t_1}}=\sqrt{\bar{\alpha}_1}$；节点 $t_2=3$，因此超参数 $\sqrt{\bar{\alpha}_{t_2}}=\sqrt{\bar{\alpha}_3}$。根据公式 9，对于任意图像采样序列 $\tau=\{t_1,t_2,\cdots\ t_{i-1},t_i,\cdots\ t_S\}$，生成过程为：

$$\mathbf{x}_{t_{i-1}}=\sqrt{\bar{\alpha}_{t_{i-1}}}\left(\frac{\mathbf{x}_{t_i}-\sqrt{1-\bar{\alpha}_{t_i}}\epsilon_\theta\left(\mathbf{x}_{t_i},t_i\right)}{\sqrt{\bar{\alpha}_{t_i}}}\right)+\sqrt{1-\bar{\alpha}_{t_{i-1}}-\sigma_{t_i}^2}\cdot\epsilon_\theta\left(\mathbf{x}_{t_i},t_i\right)+\sigma_{t_i}\epsilon,\tag{17}$$

其中 $\sigma_{t_i}=\eta\sqrt{(1-\bar{\alpha}_{t_{i-1}})/(1-\bar{\alpha}_{t_i})}\sqrt{1-\bar{\alpha}_{t_i}/\bar{\alpha}_{t_{i-1}}}$，且 $\eta\in[0,1]$。当 $\eta=0$ 时，添加的噪声为零。此时，图像生成变为确定性过程。其转移概率 $p_\theta(\mathbf{x}_{t_{i-1}}\mid\mathbf{x}^{t_i})$ 可表示为：

$$\mathbf{x}_{t_{i-1}}=\sqrt{\bar{\alpha}_{t_{i-1}}}\left(\frac{\mathbf{x}_{t_i}-\sqrt{1-\bar{\alpha}_{t_i}}\epsilon_\theta\left(\mathbf{x}_{t_i},t_i\right)}{\sqrt{\bar{\alpha}_{t_i}}}\right)+\sqrt{1-\bar{\alpha}_{t_{i-1}}}\cdot\epsilon_\theta\left(\mathbf{x}_{t_i},t_i\right).\tag{18}$$

生成隐写图像的过程如算法2所示。首先，将输入的秘密数据嵌入潜变量 z 中，得到隐写潜变量 z_s。然后，利用公式 18中的转移概率，通过逐步去噪生成隐写图像：$\mathrm{x}_{t_S}\rightarrow\mathrm{x}^{t_{S-1}}\cdots\rightarrow\mathrm{x}_{t_2}\rightarrow\mathrm{x}_{t_1}\rightarrow\mathrm{x}0$。采样序列 $\tau=\{t_1,t_2,\cdots\ t_{i-1},t_i,\cdots\ t_S\}\subseteq\{1,2,\cdots\ T\}$。为方便起见，$\mathrm{x}_{t_S}\rightarrow\mathrm{x}^{t_{S-1}}\cdots\rightarrow\mathrm{x}_{t_2}\rightarrow\mathrm{x}_{t_1}\rightarrow\mathrm{x}0$ 可以重新编号为 $\mathrm{x}_S\rightarrow\mathrm{x}_{S-1}\cdots\rightarrow\mathrm{x}_2\rightarrow\mathrm{x}_1\rightarrow\mathrm{x}0$，转移概率 $p_\theta(\mathrm{x}_{t_{i-1}}\mid\mathrm{x}^{t_i})$ 可简写为 $p_\theta(\mathrm{x}_{i-1}\mid\mathrm{x}_i)$，如算法2所示。例如，当 $\tau=\{100,200,\cdots,900,1000\}$ 且 $S=10$ 时，表示图像生成时仅选取下标为这些数字的十个节点。由此生成相应的隐写图像 $\mathrm{x}_{10}\rightarrow\mathrm{x}_9\rightarrow\mathrm{x}8\cdots\rightarrow\mathrm{x}_1\rightarrow\mathrm{x}0$。然后，将 x0 量化，得到隐写图像 x^s。

算法2 隐写图像生成

输入： 预训练的 StegoDiffusion；预设超参数 $\bar{\alpha}_t$ $(t\in[1,T])$；采样序列 $\{t_1、\cdots\ t_S\}\subseteq\{1、2、\cdots\ T\}$。
对于每一步**执行**
  输入秘密数据$\mathbf{d}\in\{0,1\}^n$；
  随机采样潜变量$\mathbf{z}\sim N(0,1)$；
  将$\mathbf{d}$ 隐藏到潜变量$\mathbf{z}$ 中，得到隐写潜变量$\mathbf{z}_s$；
  将$\mathbf{z}_s$ 用作初始噪声图像$\mathbf{x}_S$，并将其输入预训练的 StegoDiffusion；
  迭代生成过程 $p_\theta(\mathbf{x}_{i-1}\mid\mathbf{x}_i)$ 以生成隐写图像，$\mathbf{x}_S\cdots\rightarrow\mathbf{x}_1\rightarrow\mathbf{x}_0$；
  对 $\mathbf{x}_0$ 进行量化，得到最终隐写图像 $\mathbf{x}^s$。
输出： 隐写图像集合。

算法 3 秘密数据提取

1：输入：与发送方相同的预训练 StegoDiffusion 模型、预设超参数 $\bar{\alpha}_t(t\in[1,T])$、采样序列 $\tau=\{t_1,\cdots\ t_S\}\subseteq\{1,2,\cdots\ T\}$，以及接收到的隐写图像。2：对每个步骤执行 3：输入一张隐写图像 x^s；4：对 x^s 进行预处理，并将其转换为浮点图像 x0；5：使用扩散过程 $q(\mathrm{x}_{i+1}\mid\mathrm{x}_i)$ 恢复原始图像序列 $\mathrm{x}_0\rightarrow\mathrm{x}1\cdots\rightarrow\mathrm{x}_S$；6：将 x_S 作为恢复的隐写潜变量 $\mathrm{z}'s$。7：从 $\mathrm{z}'s$ 的 DCT 系数中提取秘密数据 d′。8：输出：提取出的秘密数据。

3.3 秘密数据提取

$$\mathbf{Coeff_r}=\mathrm{DCT}(\mathbf{z}'_s),$$
$$\mathbf{d}'=\mathrm{reshape}\left(\left\lceil\frac{\mathrm{Sign}(\mathbf{Coeff_r})+1}{2}\right\rceil\right),\tag{19}$$

其中，$\mathbf{Coeff_r}$ 表示包含秘密数据的 DCT 系数，Sign(·) 表示输入的符号矩阵，$\lceil\cdot\rceil$ 表示向上取整操作。reshape(·) 操作用于将输出 $\mathbf{d}'$ 重塑为与原始秘密数据 $\mathbf{d}$相同的形状。

接收到隐写图像 $\mathbf{x}^s$ 后，接收方可以使用 StegoDiffusion 的扩散过程提取秘密数据，如算法 3 所示。

根据公式 18 和第 3.1 节所述的 StegoDiffusion 的可逆性，加速扩散过程 $q(\mathbf{x}_{t_{i+1}}\mid\mathbf{x}^{t_i})$ 的转移概率为：

$$\mathbf{x}_{t_{i+1}}=\sqrt{\bar{\alpha}_{t_{i+1}}}\left(\frac{\mathbf{x}_{t_i}-\sqrt{1-\bar{\alpha}_{t_i}}\epsilon_\theta\left(\mathbf{x}_{t_i},t_i\right)}{\sqrt{\bar{\alpha}_{t_i}}}\right)+\sqrt{1-\bar{\alpha}_{t_{i+1}}}\cdot\epsilon_\theta\left(\mathbf{x}_{t_i},t_i\right).\tag{20}$$

提取隐藏的秘密数据时，起始图像是隐写图像 x^s，通过迭代式 20 中的扩散过程，可以逐步恢复噪声图像，即 $\mathrm{x}^s\rightarrow\mathrm{x}^{t_1}\cdots\rightarrow\mathrm{x}_{t_i}\rightarrow\mathrm{x}_{t_{i+1}}\rightarrow\mathrm{x}_{t_{S-1}}\rightarrow\mathrm{x}^{t_S}$。为方便起见，将其重新编号如下：$\mathrm{x}^s\rightarrow\mathrm{x}1\cdots\rightarrow\mathrm{x}_i\rightarrow\mathrm{x}_{i+1}\cdots\rightarrow\mathrm{x}_S:=\mathrm{z}'s$。并将 $q(\mathrm{x}_{t_{i+1}}\mid\mathrm{x}^{t_i})$ 简化为 $q(\mathrm{x}_{i+1}\mid\mathrm{x}_i)$。噪声图像 x_S是恢复出的隐写潜变量 $\mathrm{z}'s$，并可根据式 19 提取隐藏的秘密数据。

4 实验结果

4.1 实施细节

实验在 Ubuntu 18.04 上进行，使用 Nvidia 3090 显卡和 Pytorch 1.8 平台。模型训练采用默认学习率为 4e-2 的 Adam优化器。训练期间，扩散步 $T = 1000$ 和超参数设置为 $\alpha_t = 1 - \frac{0.02t}{T}(t \in [1, T])$，以及 $\bar{\alpha}_t = \prod_{i=1}^{T} \alpha_i$。加速图像生成过程需要 S 采样步数来生成隐写图像，其中 $\tau = \{t_1, t_2, \cdots t_{i-1}, t_i, \cdots t_S\}$，以及 $\tau \subseteq \{1, 2, \cdots T\}$。对于 S 采样步数，采用均匀采样间隔，$\Delta t = t_i - t_{i-1} = \frac{T}{S}$。例如，当 $T = 1000$ 和 $S = 100$ 时，采样间隔为 $\Delta t = \frac{1000}{100} = 10$，采样步序列为 $\tau = \{10, 20, 30 \cdots 990, 1000\}$。在实验中，使用了不同的采样步数 S，包括 10、50、100、200、500 和 1000，Δt 100, 20, 10, ，相应的采样间隔分别为 5, 2, 1。

CelebA [12], Cifar10 [8], 和 Lsun-Bedroom [29] 用于模型训练。CelebA 和 Lsun-Bedroom 用于生成64×64 隐写图像，而 Cifar10 用于在实验中生成 32×32 隐写风景图像。每个数据集约包含 1.3 亿张图像，用于模型训练。

4.2 指标

我们使用五种评估指标——每像素比特数、FID、准确率、Pe 和时间——来评估 GSD 方案的性能，它们分别表示数据隐藏容量（载荷）、图像质量、秘密数据的提取准确率、安全性和耗时。

指标每像素比特数表示每个像素隐藏的数据比特数（bit-per-pixel），每像素比特数 $= \text{len(d)}/C \times H \times W$，其中 len(d) 表示隐藏数据的比特数；$C$、$H$ 和 W 分别表示隐写图像的通道数、高度和宽度。

FID 用于评估生成图像的质量，FID $= \|\mu_r - \mu_g\|^2 + \text{Tr}\left(\Sigma_r + \Sigma_g - 2\left(\Sigma_r \Sigma_q\right)^{1/2}\right)$，其中 μ_r 和 μ_g 分别表示从真实图像和生成图像中提取的特征的均值；Σ_r 和 Σ_g 表示提取特征的协方差；Tr 表示矩阵的迹。FID 值越小，生成图像的质量越好。

准确率表示隐藏的秘密数据的提取准确率，准确率 $= \text{d} \odot \text{d}' / \text{len(d)}$，其中 d 和 d′ 分别表示隐藏的秘密数据和提取的隐藏数据。$\odot$ 表示逐元素异或非（XNOR）运算。

Pe 是隐写分析安全性最常用的评价指标，Pe $= \min_{P_{FA}} \frac{1}{2}(P_{FA} + P_{MD})$，其中 P_{FA} 和 P_{MD} 分别是隐写分析中的虚警率和漏检率。Pe 的取值范围为 0 到 1，最优值为 0.5，表示隐写算法具有最高的安全性。

时间表示生成隐写图像或提取隐藏数据的平均时间，时间 $= \frac{t_{gen} + t_{ext}}{2}$，其中 t_{gen} 是从秘密数据生成隐写图像的平均时间，t_{ext} 是从隐写图像中提取隐藏数据的平均时间。时间的单位为秒/图像。

表 1：所提出的 GSD方案在不同数据集上的综合性能

数据集	S	准确率(%) ↑	Pe→ 0.5	Fid↓	时间↓
Cifar10 32×32	10	99.66	0.4552	15.34	7.33e-3
	50	99.91	0.4574	7.57	3.62e-2
	100	99.93	0.4595	7.45	0.0861
	200	99.93	0.4664	8.08	0.1605
	500	99.94	0.4796	9.10	0.3509
	1000	99.97	0.4857	13.55	0.7710
CelebA 64×64	10	99.84	0.4743	15.77	4.02e-2
	50	100.00	0.4850	8.10	0.1346
	100	100.00	0.4575	9.28	0.2520
	200	100.00	0.4967	11.61	0.4893
	500	100.00	0.4781	13.91	1.2071
卧室 64×64	10	99.51	0.4530	17.24	1.78e-2
	50	99.93	0.4620	11.38	0.1094
	100	99.98	0.4698	12.46	0.2128
	200	99.99	0.4647	14.20	0.4684
	500	100.00	0.4730	16.17	0.9954

4.3 GSD 的有效性

如图 5 和图 6 所示，GSD方案在 Cifar10 [8], CelebA [12], 和 Lsun-bedroom [29], 上进行训练，并且能够生成尺寸为 32×32 和 64×64、载荷为 1 每像素比特数的隐写图像。

图 5 展示了 10 类随机生成的、尺寸为 32×32 的隐写风景图像，每一类都清晰逼真。图 6a 展示了尺寸为 64×64 的隐写人脸图像，其面部特征合理，外观自然。图 6b 展示了尺寸为 64×64 的隐写卧室图像，几乎与真实图像难以区分。高质量的隐写图像能够欺骗第三方，使其难以检测出图像中的任何异常，从而提高隐蔽通信的安全性。

所提出的GSD方案在不同数据集和不同采样步数 S 下的综合性能评估结果如表1所示。在不同数据集上，生成的不同尺寸隐写图像所承载的秘密数据载荷始终为1 每像素比特数（bpp），这意味着尺寸越大的图像能够隐藏越多的秘密数据。随着图像生成采样步数 S 的增加，准确率（Acc）、Pe和时间（Time）指标的数值逐渐增大，而FID值则先降低，随后逐渐升高。当 S 介于50至100之间时，各项指标达到平衡：图像生成速度较快，Acc可达到约100%，通过SCRMQ1 [4] 测试得到的Pe值介于0.45至0.5之间，且FID值较小。此时，该方案在图像生成速度、提取准确率、图像质量和安全性方面均表现出优异的性能。

为了研究采样步数 S 对生成的隐写图像内容的影响，我们在 CelebA 数据集上使用相同设置、但改变 S 生成了一些隐写图像，如图 7 所示。可以观察到，使用不同的 S生成的隐写图像，其内容极为相似。只有使用 $S = 10$ 生成的图像（第一行）与其他图像略有不同。当 $S = 50$时，生成图像的内容与使用较大 S生成的图像相似，但图像生成速度快得多。

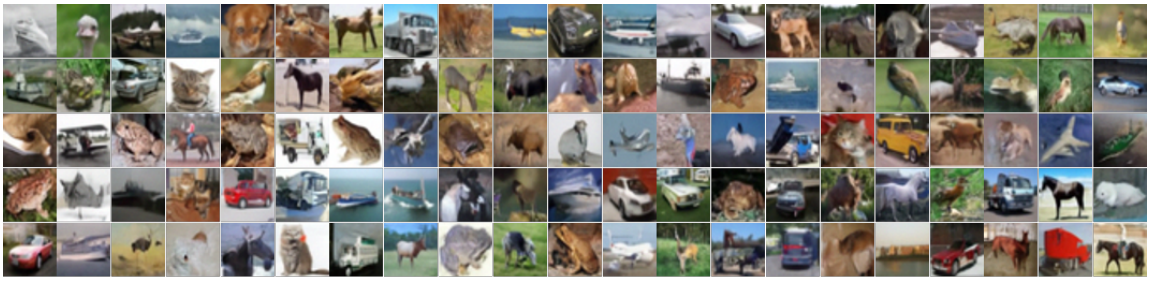
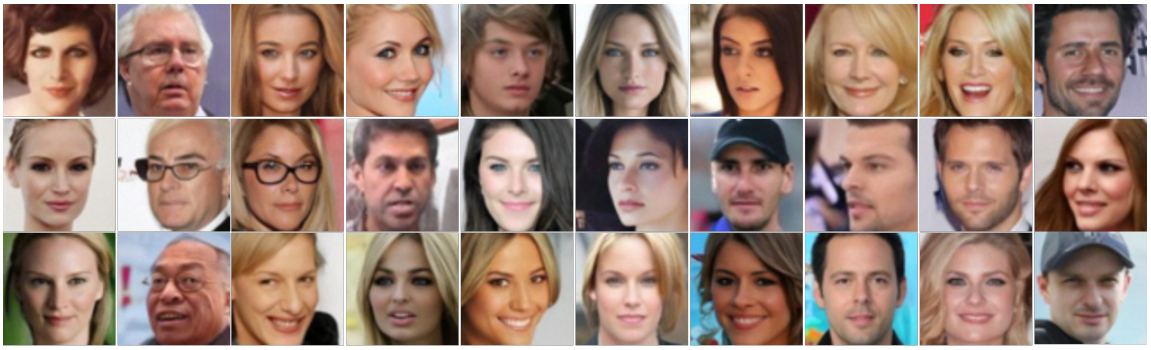


图 5：在 Cifar10 数据集上生成的 32×32 隐写风景图像。载荷=1每像素比特数，准确率=99.91%，Pe=0.4574，FID=7.57，S=50。



(a) CelebA 数据集上的隐写人脸图像



(b) 卧室数据集上的隐写卧室图像

图 6：GSD 在不同数据集上生成的 64×64 隐写图像。对于 CelebA 数据集：载荷=1 每像素比特数，准确率=100.00%，Pe=0.4850，FID=8.10，以及 S=50；对于卧室数据集：载荷=1 每像素比特数，准确率=100.00%，Pe=0.4620，FID=11.38，以及S=50。

4.4 可逆性可视化

为直观展示 GSD方案的可逆性，我们在图 8 中呈现了生成隐写图像和提取隐藏的秘密数据的过程，其中图像生成步骤 $S = 10$。从左到右，第 1、3、5 行的图像展示了使用所提出的 StegoDiffusion 的生成过程，从隐写潜变量 $\mathbf{z}_S$ 生成隐写图像的过程。根据式 18 生成一系列去噪隐写图像：$\mathbf{x}_{10} := \mathbf{z}_S \rightarrow \mathbf{x}_9 \rightarrow \mathbf{x}_8 \cdots \rightarrow \mathbf{x}_1 \rightarrow \mathbf{x}_0$。从右到左，第 2、4、6 行的图像展示了通过 StegoDiffusion 的扩散过程（如式 20 所述），从量化隐写图像逐步恢复隐写潜变量的过程：$\mathbf{x}_0 \rightarrow \mathbf{x}_1 \cdots \rightarrow \mathbf{x}_9 \rightarrow \mathbf{z}_S := \mathbf{x}_{10}$。从图 8 可以看出，隐写图像中的噪声潜变量 $\mathbf{z}_S$ 在生成过程中逐渐消除，图像也变得越来越清晰。经过 10 次迭代后，即可获得一幅逼真且自然的隐写图像。当生成的隐写图像经历扩散过程时，被去除的噪声可以逐渐恢复。输入的隐写潜变量 $\mathbf{z}_S$ 与恢复的隐写潜变量 $\mathbf{z}_{\prime S}$ 之间的绝对平均误差小于 1e-2。这表明图像生成过程与秘密数据提取过程几乎是可逆的，并且隐藏在隐写潜变量中的秘密数据可以以接近或达到 100% 的准确率提取出来。使用更大的步长 S时，可以获得更高的隐写图像质量和数据提取准确性。

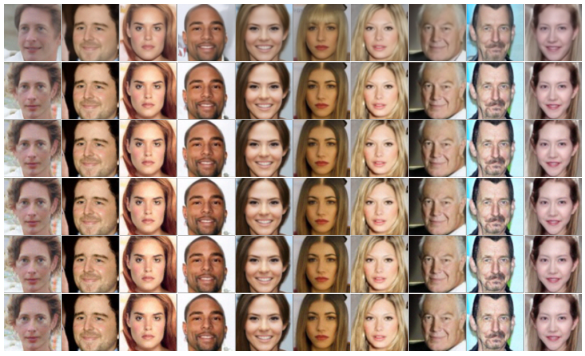


图 7：GSD 在不同采样步数下生成的隐写人脸图像 S。从第一行到最后一行，每行使用的S 分别为 10、50、100、200、500 和 1000。

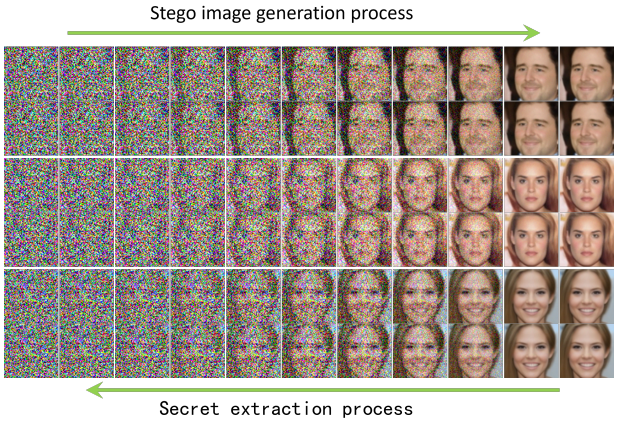


图 8：GSD 中逐步生成隐写图像和提取秘密数据的过程。第一、第三和第五行展示了从隐写潜变量 z_S 逐步去噪以生成隐写图像的过程，而第二、第四和第六行则展示了从量化隐写图像逐步恢复 z_S 的过程。

4.5 性能比较

如表 2 所示，我们将提出的 GSD 与其他先进方法进行了比较。方法 Hu [6], Liu [11], 和 GSN [24] 均使用生成对抗网络（GAN）生成隐写图像。GSF [25] 是一种基于 Flow 模型的生成式隐写方案。GSD 是我们提出的生成式隐写方案，它基于所提出的 StegoDiffusion 模型，并在图像生成过程中采用采样步数 $S = 50$ 。

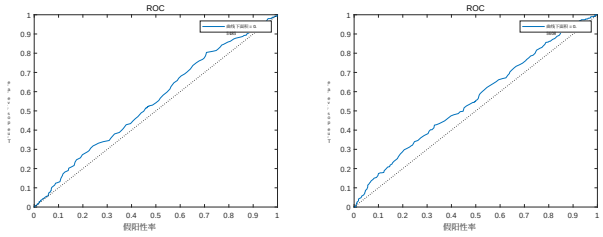
在 CelebA [12] 数据集上进行对比实验，以生成尺寸为 64×64 的隐写图像。计算 FID 时，使用 50,000 张生成图像和真实图像。Pe 通过隐写分析网络 Ye-net [28], 得出，该网络使用 10,000 张生成的隐写图像/载体图像进行训练，并使用 1,000 张隐写图像/载体图像进行测试；其中，载体图像由潜变量 z 直接生成，未隐藏秘密数据。不同指标之间存在权衡，例如，提高载荷可能会导致其他指标下降。Liu [11] 和 Hu [6] 将秘密数据载荷降低至

表 2：我们提出的 GSD 与最先进方法的比较

方法	每像素比特数 ↑	准确率 (%) ↑	FID ↓	Pe → 0.5
Liu [11]	4.17e-2	98.26	26.37	0.5350
Hu [6]	2.44e-2	91.73	30.81	0.4700
GSN [24]	0.33	98.15	12.88	0.4756
GSF [25]	1	99.96	22.45	0.4864
所提出的 GSD	**1**	**100.00**	**8.10**	**0.4960**

提高提取准确率和图像质量，这体现在每像素比特数低于 1e-2 且 FID 约为 30。与 Liu [11] 和 Hu[6], 相比，GSN [24] 和 GSF [25] 在保持相当的提取准确率和安全性的同时，提高了载荷和图像质量。然而，上述四种方案均无法实现 100% 的提取准确率。我们的 GSD 在所有指标上均优于其他方法：载荷最高可达 1 每像素比特数，提取准确率可达到 100%，同时保持最佳的图像质量和安全性。

4.6 安全性分析



(a) CelebA　　(b) 卧室

图 9：不同数据集上的检测 ROC 曲线。

为验证所提出的 GSD 方案的安全性，我们在 CelebA 和 Bedroom 数据集上测试了 GSD 对隐写分析算法 SCRMQ1 [4] 的抵抗能力。生成隐写图像/载体图像时，采样步数 S 设为 50。测试中，我们对 1,000 张包含秘密信息的隐写图像和 1,000 张不含秘密信息的生成载体图像进行分类，并绘制检测 ROC 曲线，如图 9 所示。在隐写术中，最优 ROC 曲线为反对角线（图中以虚线表示），最优曲线下面积值为 0.5。如图所示，我们的 ROC 曲线和曲线下面积值接近理想结果，这表明我们提出的 GSD 方案具有很高的安全性。

4.7 结论

本文通过设计可逆的 StegoDiffusion 模型，提出了一种名为“生成式隐写扩散”（GSD）的高性能生成式隐写方案。GSD方案采用带有快速采样技术的非马尔可夫扩散过程，以加快图像生成。该方案基于 StegoDiffusion 中的转移概率构建常微分方程，并通过欧拉迭代法实现隐写图像与秘密数据的逆变换。实验结果表明，在所有评估指标上，GSD 均优于现有方法。

参考文献

[1] Florinel-Alin Croitoru、Vlad Hondru、Radu Tudor Ionescu 和 Mubarak Shah。2023年。视觉中的扩散模型：综述。IEEE模式分析与机器智能汇刊（2023年）。[2] Prafulla Dhariwal 和 Alexander Nichol。2021年。扩散模型在图像合成方面胜过生成对抗网络。神经信息处理系统进展 34（2021年），8780–8794。[3] Numrena Farooq 和 Arvind Selwal。2023年。基于深度学习的图像隐写分析：系统综述与开放研究挑战。《环境智能与人性化计算杂志》（2023年），1–33。[4] Miroslav Goljan、Jessica Fridrich 和 Rémi Cogranne。2014年。用于彩色图像隐写分析的富模型。载于2014年IEEE信息取证与安全国际研讨会（WIFS）。IEEE，185–190。[5] Jonathan Ho、Ajay Jain 和 Pieter Abbeel。2020年。去噪扩散概率模型。神经信息处理系统进展 33（2020年），6840–6851。[6] Donghui Hu、Liang Wang、Wenjie Jiang、Shuli Zheng 和 Bin Li。2018年。一种基于深度卷积生成对抗网络的新型图像隐写方法。IEEE Access 6（2018年），38303–38314。[7] Diederik P Kingma 和 Prafulla Dhariwal。2018年。Glow：采用可逆1×1卷积的生成流。[8] Alex Krizhevsky、Geoffrey Hinton 等。2009年。从微小图像中学习多层特征。（2009年）。[9] Jayakanth Kunhoth、Nandhini Subramanian、Somaya Al-Maadeed 和 Ahmed Bouridane。2023年。视频隐写：最新进展与挑战。《多媒体工具与应用》（2023年），1–43。[10] Ming-Ming Liu、Min-qing Zhang、Jia Liu、Ying-nan Zhang 和 Yan Ke。2017年。基于生成对抗网络的无载体信息隐藏。arXiv预印本 arXiv:1712.06951（2017年）。[11] Xiyao Liu、Ziping Ma、Junxing Ma、Jian Zhang、Gerald Schaefer 和 Hui Fang。2022年。用于无嵌入隐写的图像解耦自编码器。载于IEEE/CVF计算机视觉与模式识别会议论文集。2303–2312。[12] Ziwei Liu、Ping Luo、Xiaogang Wang 和 Xiaoou Tang。2015年。野外人脸属性的深度学习。载于 IEEE计算机视觉国际会议论文集。3730–3738。[13] Mohammed Abdul Majeed、Rossilawati Sulaiman、Zarina Shukur 和 Mohammad Kamrul Hasan。2021年。文本隐写技术综述。《数学》9，21（2021年），2829。[14] Pratap Chandra Mandal、Imon Mukherjee、Goutam Paul 和 BN Chatterji。2022年。数字图像隐写：文献调查。《信息科学》（2022年）。[15] Alex Nichol、Prafulla Dhariwal、A. Ramesh、Pranav Shyam、Pamela Mishkin、Bob McGrew、Ilya Sutskever 和 Mark Chen。2021年。GLIDE：采用文本引导扩散模型实现照片级真实感图像生成与编辑。国际机器学习会议（2021年）。[16] Rachna Patel、Kalpesh Lad 和 Mukesh Patel。2021年。未压缩域和压缩域视频隐写的研究与探讨：全面综述。《多媒体系统》27，5（2021年），985–1024。[17] Alec Radford、Jong Wook Kim、Chris Hallacy、Aditya Ramesh、Gabriel Goh、Sandhini Agarwal、Girish Sastry、Amanda Askell、Pamela Mishkin、Jack Clark 等。2021年。通过自然语言监督学习可迁移的视觉模型。载于国际机器学习会议。PMLR，8748–8763。[18] Aditya Ramesh、Prafulla Dhariwal、Alex Nichol、Casey Chu 和 Mark Chen。2022年。基于 CLIP 潜变量的分层文本条件图像生成。arXiv预印本 arXiv:2204.06125（2022年）。[19] Robin Rombach、Andreas Blattmann、Dominik Lorenz、Patrick Esser 和 Björn Ommer。2022年。利用潜扩散模型进行高分辨率图像合成。载于IEEE/CVF计算机视觉与模式识别会议论文集。10684–10695。[20] Chitwan Saharia、William Chan、Saurabh Saxena、Lala Li、Jay Whang、Emily Denton、Seyed Kamyar Seyed Ghasemipour、Burcu Karagol Ayan、S Sara Mahdavi、Rapha Gontijo Lopes 等。2022年。采用深度语言理解的照片级真实感文生图扩散模型。arXiv预印本 arXiv:2205.11487（2022年）。[21] Jiaming Song、Chenlin Meng 和 Stefano Ermon。2020年。去噪扩散隐式模型。arXiv预印本 arXiv:2010.02502（2020年）。[22] Nandhini Subramanian、Omar Elharrouss、Somaya Al-Maadeed 和 Ahmed Bouridane。2021年。图像隐写：最新进展综述。IEEE Access 9（2021年），23409–23423。[23] Jinyuan Tao、Sheng Li、Xinpeng Zhang 和 Zichi Wang。2018年。迈向鲁棒图像隐写。IEEE视频技术电路与系统汇刊 29，2（2018年），594–600。[24] Ping Wei、Sheng Li、Xinpeng Zhang、Ge Luo、Zhenxing Qian 和 Qing Zhou。2022年。生成式隐写网络。载于第30届ACM多媒体国际会议论文集。1621–1629。[25] Ping Wei、Ge Luo、Qi Song、Xinpeng Zhang、Zhenxing Qian 和 Sheng Li。2022年。生成式隐写流。载于2022年IEEE国际多媒体与博览会会议（ICME）。IEEE，1–6。[26] Junqi Wu、Bolin Chen、Weiqi Luo 和 Yanmei Fang。2020年。基于针对卷积神经网络的迭代对抗攻击的音频隐写术。*IEEE* 信息取证与安全汇刊 15（2020年），2282–2294。[27] Ling Yang、Zhilong Zhang、Yang Song、Shenda Hong、Runsheng Xu、Yue Zhao、Yingxia Shao、Wentao Zhang、Bin Cui 和 Ming-Hsuan Yang。2022年。扩散模型：方法与应用的全面综述。*arXiv* 预印本 *arXiv:2209.00796*（2022年）。[28] Jian Ye、Jiangqun Ni 和 YangYi。2017年。利用深度学习学习用于图像隐写分析的层次化表示。*IEEE* 信息取证与安全汇刊 12，11（2017年），2545–2557。[29] Fisher Yu、Ari Seff、Yinda Zhang、Shuran Song、Thomas Funkhouser 和 Jianxiong Xiao。2015年。Lsun：利用人类参与的深度学习构建大规模图像数据集。*arXiv* 预印本 *arXiv:1506.03365*（2015年）。