PRoADS：采用潜变量优化和后向欧拉反演的可证明安全且鲁棒的音频扩散隐写术

Yongpeng Yan[2], Yanan Li[2],Qiyang Xiao[2],Yanzhen Ren[1,2,]*

[1]空天信息安全与可信计算重点实验室，教育部。[2]武汉大学网络空间安全学院

arXiv:2603.10314v1 [cs.CR] 11 Mar 2026

摘要

本文提出 PRoADS，这是一种基于音频扩散模型、可证明安全且鲁棒的音频隐写框架。作为一种生成式隐写方案，PRoADS 通过正交矩阵投影将秘密信息嵌入扩散模型的初始噪声中。为解决扩散反演中的重构误差导致比特错误率（误码率）较高这一问题，我们引入潜变量优化和后向欧拉反演，以尽可能减小潜变量重构误差和扩散反演误差。全面的实验表明，在 64 kbps MP3 压缩下，该方案的误码率仍低至 0.15%，显著优于现有方法，并展现出强大的鲁棒性。

索引词— 音频隐写，扩散模型

1. 引言

随着生成式人工智能的发展，互联网上AI生成内容的比例持续增长。这为生成式隐写创造了有利的通信信道。生成式隐写在多媒体内容生成过程中嵌入消息，与基于修改的隐写[1]相比，具有更优的安全性和样本多样性。当前的生成式音频隐写主要利用生成对抗网络（GANs）[2, 3] 和流模型[4], , 但生成内容在语音及专用数据集方面存在局限[5]。扩散模型在实践中已被证明具有强大的生成能力，音频生成领域涌现出大量出色的工作 [6, 7, 8]。

当前基于扩散模型的隐写方法可分为三类：基于生成内容的隐写[9, 10], 基于中间状态的隐写 [11,12, 13], 以及基于初始噪声嵌入的隐写 [14, 15,16, 17]。其中，基于初始噪声嵌入的方法具有显著优势：其生成过程与常规生成过程完全相同，且生成时无需同步随机种子。这些方法通过扩散反演获得重建的初始噪声，并从中提取秘密消息。不同隐写方案的主要区别在于如何将隐藏消息映射到扩散模型的初始噪声中。Wei [14]提出了 GSD 方法，将秘密消息映射到初始噪声的 DCT系数，并通过 IDCT 重建高斯分布噪声，作为扩散模型的初始噪声。Kim [15]提出了 DiffStega 方法，该方法将秘密消息映射到二进制中心，在分布保持与鲁棒性之间取得平衡。Yang [16]提出了区间映射方法，利用分位数函数将均匀分布的消息映射为高斯分布采样点。Hu [17] 通过构造正交矩阵并进行矩阵乘法，将秘密消息映射到标准高斯分布。尽管这些方法通过精心设计的映射算法提升了隐写系统的鲁棒性，但并未充分解决扩散反演过程中引入的潜变量重建误差和扩散反演误差。这些误差会降低消息提取时初始噪声的重建准确率。

本文提出一种基于音频扩散模型潜空间初始噪声映射的隐写方法。通过引入潜空间优化和后向欧拉反演，该方法提升了初始噪声的重建准确率，从而以更高精度实现鲁棒隐写。主要贡献如下：

1）我们提出一种基于音频扩散模型的生成式隐写方法，通过正交矩阵投影将秘密消息嵌入初始噪声中，实现鲁棒且可证明安全的隐写术。

2）我们在从隐写音频中提取消息的过程中引入潜空间优化和后向欧拉迭代，分别用于潜变量重构和扩散反演。这两种方法减少了初始噪声重构误差，从而降低了消息提取过程中的误码率。

3）实验表明，我们的方法在 64 kbps MP3压缩下仍能将误码率保持在 0.15%，展现出卓越的鲁棒性。

2. 预备知识

DPM-Solver 是一系列用于扩散模型的数值求解器，其一阶和二阶公式广泛应用于加速扩散采样。其主要目标是通过优化扩散常微分方程（ODE）的求解过程，提高采样的计算效率。对 Lu 等人提出的基于指数积分器的扩散 ODE 框架进行高阶泰勒展开后，可根据展开阶数推导出不同形式的求解器，其中一阶公式（等价于 DDIM采样）为：[18],

$$x_{t_i} = \frac{\sigma_{t_i}}{\sigma_{t_{i-1}}} x_{t_{i-1}} - \alpha_{t_i}(e^{-h_i} - 1)x_\theta(x_{t_{i-1}}, t_{i-1}) \quad (1)$$

本工作得到国家自然科学基金（NSFC）资助，资助号为 NO.62572358,621723064,62372334。* 通讯作者。

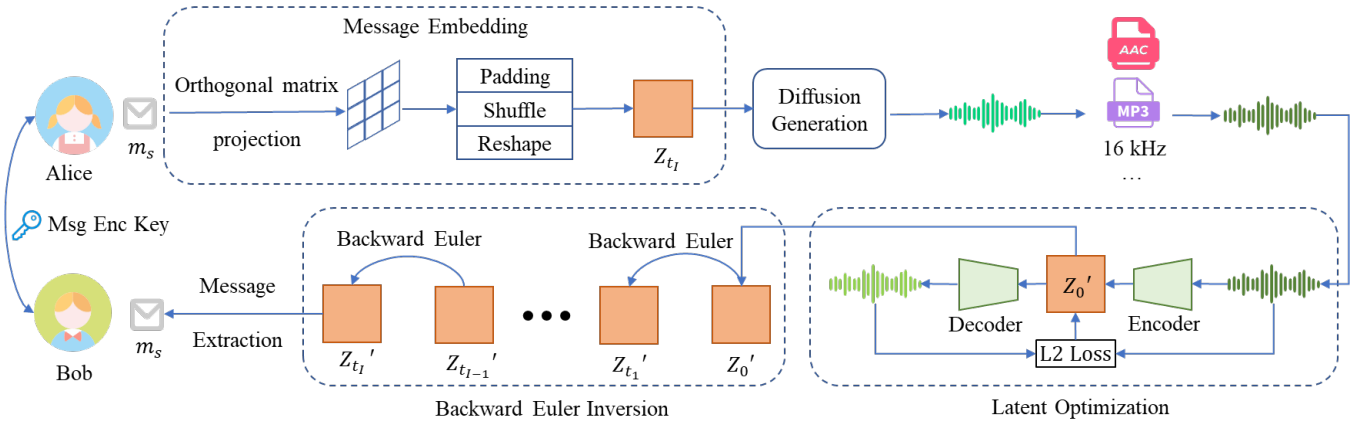


图 1。PRoADS 的整体框架。

以及二阶公式：

$$x_{t_i} = \frac{\sigma_{t_i}}{\sigma_{t_{i-1}}} x_{t_{i-1}} - \alpha_{t_i}(e^{-h_i} - 1)x_\theta(x_{t_{i-1}}, t_{i-1}) - \frac{\alpha_{t_i}(e^{-h_i} - 1)}{2r_i}(x_\theta(x_{t_{i-1}}, t_{i-1}) - x_\theta(x_{t_{i-2}}, t_{i-2})) \tag{2}$$

具体而言，扩散反演是一种求解上述扩散采样方程逆过程的方法。在扩散反演过程中引入后向欧拉迭代，可以获得更准确的解。

3. 所提方法

本文提出的隐写方法整体框架如图 1 所示，包含三个主要组成部分：消息嵌入、潜变量优化和后向欧拉反演。

3.1. 消息嵌入

给定一串消息比特，消息嵌入会将其映射为初始扩散噪声。设潜空间的维度为 $[F, T]$。首先初始化一个正交矩阵 $A \in \mathbb{R}^{N\times N}$，并定义 $C = F \times T \bmod (N \times N)$。秘密消息表示为二进制矩阵 $M \in \{0,1\}^{C\times N\times N}$。据此，我们计算秘密嵌入噪声矩阵如下：

$$z_{\text{secret}} = A \cdot M \cdot A^T. \tag{3}$$

随后，将 z_{secret} 展平为一个向量，并在其后追加从标准高斯分布中采样的额外噪声，将向量长度扩展至 $F \times T$。然后对所得向量执行洗牌操作，并将其重塑为 $[F, T]$ 潜空间格式，从而得到用于生成的最终噪声张量 z_s。这一过程可简洁地表示为：

$$z_s = \text{Reshape}\big(\text{Shuffle}\big(\text{Padding}(A \cdot M \cdot A^T)\big)\big). \tag{4}$$

在提取过程中，逆向操作按相反顺序执行：

$$M = A^T \cdot \text{Clip}\big(\text{Shuffle}^{-1}\big(\text{Flatten}(z_s)\big)\big) \cdot A. \tag{5}$$

3.2. 潜变量优化

潜扩散模型采用编码器-解码器架构，将音频信号压缩为潜表示，随后在潜空间中执行扩散生成过程。在隐写应用中，编码器用于在反演前将隐写音频转换为潜表示。然而，由于编码器相对于解码器不可逆，编码器重建的潜变量与原始潜表示之间会出现差异，从而影响消息提取的准确性。

为解决这一问题，我们对重建潜变量采用神经网络梯度优化，通过迭代近似，使其在解码器计算前逐步收敛至原始潜表示。详细计算过程见算法 1。

算法 1幂法

输入： 编码器 $E(\cdot)$，解码器 $D(\cdot)$，接收的音频 x，迭代步数 n，迭代步长 h；
输出： 重建潜变量 z；

```
z ← E(x);
对于 i = 1, 2, ··· , n do
    z = z − h × ∇_z ‖x − D(z)‖_2^2;
结束循环
返回 z。
```

3.3. 后向欧拉反演

当前的初始噪声嵌入隐写术方法采用朴素反演方法进行初始噪声重建。这些方法通过近似策略将隐式反演方程转化为显式计算形式，从而提高计算效率，但会牺牲反演精度。尽管这类反演方法在基于扩散模型的编辑应用中十分常见，隐写任务仍需优先考虑重建准确率，而非计算效率。

我们采用后向欧拉法求解隐式反演方程，提出了一种更精确的求解方法，并详细介绍了一阶和二阶求解器的步骤。

3.3.1. 一阶求解器

通过重新整理公式 1，可得到如下精确的隐式反演方程：

$$x_{t_{i-1}} = \frac{\sigma_{t_{i-1}}}{\sigma_{t_i}} \left(x_{t_i} + \alpha_{t_i}(e^{-h_i} - 1)x_\theta(x_{t_{i-1}}, t_{i-1}) \right) \quad (6)$$

隐式方程无法直接计算，但可以通过后向欧拉法迭代求解。求解过程分为两步：首先，使用前向欧拉法计算初始近似值；其次，通过牛顿法或不动点迭代更新隐式方程，直至收敛到目标精度。设迭代步长为 h；优化后的 DDIM 反演过程如算法 2 所示。

算法 2 后向欧拉DDIM反演

输入： 解码潜变量 $\hat{z}_{t_N}$，时间步 $\{t_i\}_{i=0}^N$，数据预测模型 $z_\theta(\cdot)$，迭代步长 h；

输出： 初始噪声 $\hat{z}_{t_0}$；

对于 $i \leftarrow N$ *to* 1 **执行**
  $\hat{z}_{t_{i-1}} \leftarrow \frac{\sigma_{t_{i-1}}}{\sigma_{t_i}}(\hat{z}_{t_i} + \alpha_{t_i}(e^{-h_i} - 1)z_\theta(\hat{z}_{t_i}, t_{i-1}))$
  当 $|\hat{z}_{t_i} - z'_{t_i}| > \epsilon$ **执行**
    $z'_{t_i} \leftarrow \frac{\sigma_{t_i}}{\sigma_{t_{i-1}}}\hat{z}_{t_{i-1}} - \alpha_{t_i}(e^{-h_i} - 1)z_\theta(\hat{z}_{t_{i-1}}, t_{i-1})$
    $\hat{z}_{t_{i-1}} \leftarrow \hat{z}_{t_{i-1}} - h(z'_{t_i} - \hat{z}_{t_i})$
  结束 while
结束 for
返回 $\hat{z}_{t_0}$.

由于后向欧拉方法具有无条件稳定性，迭代的收敛性与步长无关。通过后向欧拉反演，相邻时间步之间的误差可被约束在 ϵ以内。

3.3.2. 二阶求解器

所提出的优化方法从估计二阶项开始，并结合前向欧拉法（传统的 DDIM 反演）与后向欧拉法，以兼顾准确性和效率。

考虑公式 2。线性项和一阶展开项引入的误差相对较大，因此我们采用后向欧拉法进行更精确的计算。相比之下，二阶项引入的近似误差相对较小；因此，我们使用步长较小的前向欧拉法对其进行近似，并在后向欧拉迭代过程中将其视为常数。详细步骤见算法 3。

由于 DPM-Solver 的前向采样过程最初会通过低阶形式采样高阶采样所需的状态，因此后向欧拉反演也分为两个阶段。在算法 3 中，第 16–20 行对应从 z_{t_0} 到 z_{t_1} 的初始一阶采样，而第 2–15 行对应随后的二阶采样过程。

我们对算法的分析如下：二阶采样过程的反演包含两个步骤。第一步使用细粒度步长前向欧拉法，近似公式 2 中的二阶项，记为$\hat{y}_{ti-1}$ 和$\hat{y}_{ti-2}$。第二步将该二阶项近似值视为常数，并对一阶项迭代应用后向欧拉法。

Algorithm 3 Second-order DPM-Solver Inversion

Input: Initial latent variable decoded $\hat{z}t_N$, sampling time steps ${t_i}^N i = 0$, diffusion state update function $z_\theta(\cdot)$, backward Euler iteration update function UPDATE$(\cdot)$, sampling step size J;

Output: Reconstructed initial state $\hat{z}t_0$;

1: $\lambda_{t_i} = \log \frac{\alpha_{t_i}}{\sigma_{t_i}}, h_i = \lambda_{t_i} - \lambda_{t_{i-1}}, r_i = \frac{h_{i-1}}{h_i}$
2: **for** $i \leftarrow N$ *to* 2 **do**
3: $\hat{y}_{t_i} \leftarrow \hat{z}_{t_i}$
4: **for** $j \leftarrow 1$ *to* $2J$ **do**
5: $m_0 = \alpha_{t_{i-j/J}}(e^{-h_{i-j/J}} - 1)z_\theta(\hat{y}_{t_{i-(j-1)/J}}, t_{i-j/J})$
6: $\hat{y}_{t_{i-j/J}} \leftarrow \frac{\sigma_{t_{i-j/J}}}{\sigma_{t_{i-(j-1)/J}}}(\hat{y}_{t_{i-j/J}} + m_0)$
7: **end for**
8: $\hat{z}_{t_{i-1}} \leftarrow \hat{y}_{t_{i-1}}$
9: **while** $|\hat{z}_{t_i} - z'_{t_i}| > \epsilon$ **do**
10: $m_0 = \alpha_{t_i}(e^{-h_i} - 1)z_\theta(\hat{z}_{t_{i-1}}, t_{i-1})$
11: $m_1 = \frac{\alpha_{t_i}(e^{-h_i} - 1)}{2r_i}(z_\theta(\hat{y}_{t_{i-1}}, t_{i-1}) - z_\theta(\hat{y}_{t_{i-2}}, t_{i-2}))$
12: $z'_{t_i} = \frac{\sigma_{t_i}}{\sigma_{t_{i-1}}}\hat{z}_{t_{i-1}} - m_0 - m_1$
13: $\hat{z}_{t_{i-1}} \leftarrow \hat{z}_{t_{i-1}} - h(z'_{t_i} - \hat{z}_{t_i})$
14: **end while**
15: **end for**
16: $\hat{z}_{t_0} \leftarrow \frac{\sigma_{t_0}}{\sigma_{t_1}}(\hat{z}_{t_1} + \alpha_{t_1}(e^{-h_i} - 1)z_\theta(\hat{z}_{t_1}, t_0))$
17: **while** $|\hat{z}_{t_1} - z'_{t_1}| > \epsilon$ **do**
18: $z'_{t_1} \leftarrow \frac{\sigma_{t_1}}{\sigma_{t_0}}\hat{z}_{t_0} - \alpha_{t_1}(e^{-h_1} - 1)z_\theta(\hat{z}_{t_0}, t_0)$
19: $\hat{z}_{t_0} \leftarrow \hat{z}_{t_0} - h(z'_{t_1} - \hat{z}_{t_1})$
20: **end while**
21: **return** $\hat{z}_{t_0}$

细粒度前向欧拉计算如第 4–7 行所示。为便于说明，采样步长设为 50，这意味着 DPM-Solver 在 50 个离散时间间隔上进行采样。为得到$\hat{y}_{t_{i-1}}$和$\hat{y}_{t_{i-2}}$，该方法不会直接一次性对 50 个时间步进行传统反演，而是以更细的步长迭代五次，每次包含 10 个时间步。得到$\hat{y}_{t_{i-1}}$后，再以类似方式根据$\hat{y}_{t_{i-1}}$计算$\hat{y}_{t_{i-2}}$。

第 8–14 行描述了后向欧拉迭代过程
迭代的初始值$\hat{z}_{t_{i-1}}$ 设为$\hat{y}_{t_{i-1}}$。在第 10 行，m_0 表示泰勒展开的一阶项，而第 11 行的 m_1 对应二阶项，此处将其视为常数。最终状态$\hat{z}_{t_{i-1}}$ 通过第 12 行和第 13 行的后向欧拉迭代步骤计算得出。第 16–20 行实现了时间步 t_0处的后向欧拉过程，这与算法 2 等效。

3.4. 安全性

从结构角度来看，我们方法中的载体生成过程与 [17]提出的方法相似；唯一的区别在于扩散模型的潜在形状，因此引入了填充和重塑方法。 [17] 给出了分布一致性的证明，表明 z_s 服从标准高斯分布，从而确保正常生成过程与隐写过程不可区分，并确保隐写音频的分布和感知质量不可区分。

表 1。不同方法在各种攻击方法下的隐写消息误码率（BER）（%）↓

调度器	方法	ori	AAC压缩（bps）				MP3压缩（bps）				重采样		high	low
			320k	192K	128K	64K	320k	192K	128K	64K	up	down	衰减	增强
DDIM	Yang[16]	6.55	6.81	6.81	6.81	9.57	6.65	6.65	6.74	8.31	6.55	7.20	7.23	6.53
	Kim[15]	1.44	1.54	1.54	1.54	2.58	1.46	1.47	1.50	2.22	1.43	1.74	2.42	1.48
	Hu[17]	0.11	0.12	0.12	0.13	0.26	0.12	0.11	0.12	0.19	0.11	0.13	0.46	0.11
	PRoADS	**0.09**	**0.10**	**0.10**	**0.11**	**0.21**	**0.10**	**0.09**	**0.10**	**0.15**	**0.09**	**0.11**	**0.25**	**0.09**
DPMSolver	Yang[16]	7.17	7.44	7.44	7.45	9.09	7.27	7.27	7.33	8.68	7.16	7.74	7.69	7.10
	Kim[15]	1.71	1.80	1.81	1.82	2.97	1.78	1.80	1.82	2.56	1.70	2.09	2.44	1.66
	Hu[17]	0.62	0.64	0.63	0.62	0.84	0.64	0.63	0.63	0.75	0.62	0.67	0.83	0.61
	PRoADS	**0.12**	**0.15**	**0.15**	**0.15**	**0.30**	**0.14**	**0.14**	**0.13**	**0.24**	**0.12**	**0.16**	**0.29**	**0.13**

4. 实验

4.1. 设置

在初步实验中，我们评估了几种公开可用的音频扩散模型，包括 AudioLDM[6], StableAudio[8], MusicLDM[19], 和 EzAudio[7]。最终，我们选择了 EzAudio，因为它采用端到端训练的编解码器，且参数量相对较小。我们在 AudioCaps[20] 数据集上进行了实验。该模型用于生成时长为 10 秒、采样率为 24 kHz 的音频样本。为确保实验公平，所有隐写系统均采用了统一的 57344($14 \times 64 \times 64$) 比特嵌入容量。

为评估我们提出的方法和基线方法在各种攻击场景下的鲁棒性，我们引入了一套全面的扰动，包括四种不同比特率下的高级音频编码和 MP3 压缩、上采样至 44.1 kHz、下采样至 16 kHz、高频衰减（8 kHz 以上，-6 dB）以及低频增强（200 Hz 以下， +3 dB）。

4.2. 结果

4.2.1. 鲁棒性

表 1 展示了我们提出的方法和对比方法在各种攻击条件下的消息提取误码率。Yang[16] 和 Kim[15]的消息映射算法对初始噪声重建中的错误高度敏感，因此在不同攻击场景下的比特错误率始终较高。

我们的方法基于 Hu[17]的映射方法进行了三项关键改进：调整嵌入算法，使其适用于形状各异的潜空间，并采用另外两种旨在减少提取错误的技术。与 Hu[17], 在去噪扩散隐式模型采样方案下的方法相比，大多数攻击下的误码率降低了约 0.02%；在 64 kbps 压缩攻击下，误码率降低了 0.04%，在高频衰减攻击下则显著降低了 0.21%。在二阶 DPMSolver采样方案下，各种攻击中的误码率均降低了约 0.5%。这些实验结果表明，我们提出的方法有效增强了隐写系统抵御多种攻击方法的鲁棒性。

4.2.2. 消融实验

为验证潜变量优化（L.O.）和后向欧拉反演（B.E.）的有效性，我们进行了消融实验，

表2。消融实验的误码率（%）↓

攻击	基线	L.O.	B.E.	L.O.+B.E.
None	0.62	0.22	0.19	**0.12**
高级音频编码（192 kbps）	0.63	0.23	0.22	**0.15**
高级音频编码（64 kbps）	0.84	0.39	0.39	**0.30**
降采样	0.67	0.25	0.23	**0.16**
高频衰减	0.83	0.39	0.39	**0.29**
低频增强	0.61	0.21	0.19	**0.13**

并移除 L.O. 和 B.E. 方法以建立基线方法。比较聚焦于若干攻击场景，在二阶 DPMSolver 采样方法下，观察到鲁棒性显著提升。结果如表2所示。

对于失真影响较小的攻击（192 Kbps AAC压缩、降采样和低频增强），B.E. 方法取得了更好的性能提升，使误码率降低了 0.01%–0.03%。对于 64 Kbps AAC压缩和高频衰减攻击，两种优化方法的效果相当。两种技术协同应用取得了最佳结果：与单一方法相比，误码率降低了 0.07%–0.1%，相较于基线则提升了约 0.5%。

4.3. 计算成本

PRoADS 的生成过程与常规生成过程完全相同。生成一段 10 秒的音频片段总共需要 6.8 秒，这表明发送方能够连续流式生成秘密音频。然而，对于基于反演的提取过程，由于求解逆向生成过程需要大量迭代，因此所需时间为 106 秒，但仍在可接受范围内。尽管如此，在隐写术应用中，提升重建准确率和降低误码率被认为更加重要。

5. 结论

本文提出 PRoADS，这是一种可证明安全且鲁棒的音频扩散隐写方案。该方案基于鲁棒正交矩阵投影嵌入算法，通过引入潜变量优化和后向欧拉反演，最大限度地降低初始噪声的重构误差，并减少消息的误码率。实验表明，即使遭受 64 kbps 高级音频编码压缩攻击，我们的隐写方案仍能将误码率保持在 0.3 %。

6. 参考文献

[1] Ping Wei、Sheng Li、Xinpeng Zhang、Ge Luo、Zhenxing Qian 和 Qing Zhou，《生成式隐写网络》，载于第 *30*届 *ACM* 多媒体国际会议论文集，2022年，第 1621–1629 页。[2] Jing Li、Kaixi Wang 和 Xiaozhu Jia，《一种基于生成对抗网络的无载体音频隐写术》，《电子学》，第 12 卷，第 5 期，第 1253 页，2023年。[3] Lang Chen、Rangding Wang、Li Dong 和 Diqun Yan，《基于心理声学模型的不可感知对抗音频隐写术》，《多媒体工具与应用》，第 82 卷，第 17 期，第 26451–26463 页，2023年。[4] Kejiang Chen、Hang Zhou、Hanqing Zhao、Dongdong Chen、Weiming Zhang 和 Nenghai Yu，《基于文本转语音生成模型的分布保持隐写》，《IEEE可靠与安全计算汇刊》，第 19 卷，第 5 期，第 3343–3356 页，2021年。[5] Sanfeng Zhang、Baiyu Tian、Yang Gao、Mengyao Dai 和 Wang Yang，《Birdssong：一种风格化生成音频隐写术》，《计算机与电气工程》，第 123 卷，第 110112 页，2025年。[6] Haohe Liu、Zehua Chen、Yi Yuan、Xinhao Mei、Xubo Liu、Danilo Mandic、Wenwu Wang 和 Mark D Plumbley，《AudioLDM：基于潜扩散模型的文本转音频生成》，*arXiv* 预印本 *arXiv:2301.12503*，2023年。[7] Jiarui Hai、Yong Xu、Hao Zhang、Chenxing Li、Helin Wang、Mounya Elhilali 和 Dong Yu，《EzAudio：利用高效扩散变换器增强文本转音频生成》，*arXiv* 预印本 *arXiv:2409.10819*，2024年。[8] Zach Evans、CJ Carr、Josiah Taylor、Scott H Hawley 和 Jordi Pons，《快速时间条件潜音频扩散》，载于第*41*届国际机器学习会议，2024年。[9] Jiansong Zhang、Kejiang Chen、Weixiang Li、Weiming Zhang 和 Nenghai Yu，《生成图像隐写术：利用波动性增强安全性》，《*IEEE*可靠与安全计算汇刊》，第 21 卷，第 4 期，第 3994–4005 页，2023年。[10] Jiwen Yu、Xuanyu Zhang、Youmin Xu 和 Jian Zhang，《Cross：扩散模型实现可控、鲁棒且安全的图像隐写》，《神经信息处理系统进展》，第 36 卷，第 80730–80743 页，2023年。[11] Tushar M Jois、Gabrielle Beck 和 Gabriel Kaptchuk，《Pulsar：面向扩散模型的安全隐写术》，载于*2024*年 *ACM SIGSAC* 计算机与通信安全会议论文集，2024年，第 4703–4717 页。[12] Yinyin Peng、Donghui Hu、Yaofei Wang、Kejiang Chen、Gang Pei 和 Weiming Zhang，《Stegaddpm：基于去噪扩散概率模型的生成图像隐写术》，载于第*31*届 *ACM* 多媒体国际会议论文集，2023年，第 7143–7151 页。[13] Tian Wu、Xuan Hu、Yijun Hu、Yizhe Wang、Chunnian Liu 和 Yiping Zhu，《通过中间状态正态分布实现鲁棒生成式隐写》，载于国际智能计算会议。Springer，2024年，第 416–426 页。[14] Ping Wei、Qing Zhou、Zichi Wang、Zhenxing Qian、Xinpeng Zhang 和 Sheng Li，《生成式隐写扩散》，*arXiv* 预印本 *arXiv:2305.03472*，2023年。

[15] Daegyu Kim、Chaehun Shin、Jooyoung Choi、Dahuin Jung 和 Sungroh Yoon，《Diffusion-stego：通过消息投影实现无需训练的扩散生成式隐写术》，*Information Sciences*，第 122358 页，2025年。[16]Zijin Yang、Kai Zeng、Kejiang Chen、Han Fang、Weiming Zhang 和 Nenghai Yu，《高斯着色：针对扩散模型的可证明无性能损失图像水印》，载于*IEEE/CVF*计算机视觉与模式识别会议，2024年，第 12162–12171 页。[17] Xiaoxiao Hu、Sheng Li、Qichao Ying、Wanli Peng、Xinpeng Zhang 和 Zhenxing Qian，《通过流行的稳定扩散建立稳健的生成式图像隐写术》，*IEEE*信息取证与安全汇刊，2024年。[18] Cheng Lu、Yuhao Zhou、Fan Bao、Jianfei Chen、Chongxuan Li 和 Jun Zhu，《DPM-Solver：一种快速常微分方程求解器，可在约 10 步内完成扩散概率模型采样》，载于《神经信息处理系统进展》，第 35 卷，第 5775–5787 页，2022年。[19] Ke Chen、Yusong Wu、Haohe Liu、Marianna Nezhurina、Taylor Berg-Kirkpatrick 和 Shlomo Dubnov，《MusicLDM：利用节拍同步混合策略提升文本到音乐生成的新颖性》，载于*ICASSP 2024-2024 IEEE*国际声学、语音与信号处理会议（*ICASSP*）。IEEE，2024年，第 1206–1210 页。[20] Chris Dongjoo Kim、Byeongchang Kim、Hyunmin Lee 和 Gunhee Kim，《AudioCaps：为真实场景中的音频生成描述文本》，载于*2019*年北美计算语言学协会年会论文集：人类语言技术，第*1*卷（长篇与短篇论文），2019年，第 119–132 页。