arXiv:2603.04696v1 [cs.CR] 5 Mar 2026

当去噪变成去签名：基于扩散的图像编辑下水印脆弱性的理论与实证分析

Fai Gu Qiyu Tang Te Wen Emily Davis
FinnCarter西安电子科技大学

摘要

鲁棒隐形水印系统旨在嵌入不可感知的载荷，使其在 JPEG压缩、裁剪和加性噪声等常见后处理后仍可解码。与此同时，基于扩散的图像编辑已迅速成熟，成为现代内容处理流程中的“默认”转换层，支持基于指令的编辑、对象插入与合成，以及交互式几何操作。本文研究这两种趋势之间一种微妙却日益重要的相互作用：基于扩散的图像编辑过程可能会无意中损害鲁棒水印机制，在极端情况下甚至实际绕过这些专为抵御传统失真而设计的机制。

我们提出了一种对扩散编辑器的统一理解——包括 TF-ICON 等免训练合成方法、SHINE 等近期基于 FLUX 的插入框架，以及 DragFlow 等拖拽式编辑器——将它们视为随机算子，它们会（i）在潜在空间中注入大量高斯噪声，并（ii）通过学习到的去噪动力学投影回自然图像流形。在这一理解下，水印载荷表现为低能量、高频信号，会被前向扩散步骤系统性地衰减，随后又被反向生成过程视为干扰性变化。我们利用信息论工具对这种退化进行形式化分析，并证明，对于广泛类别的像素级水印编码器/解码器，随着编辑强度增加，水印载荷与编辑后的输出之间的互信息趋近于零，导致解码错误接近随机猜测。我们以一套现实的假设性实验方案及相关表格补充理论分析，涵盖具有代表性的水印方法（StegaStamp、TrustMark 和 VINE）和扩散编辑器（TF-ICON、SHINE、DragFlow 及强指令式编辑）。最后，我们讨论伦理影响、负责任披露规范，以及在生成式变换时代仍能保持意义的水印方案的具体设计指南。

1. 引言

数字水印是内容溯源、版权保护和真实性工作流程的基石。现代 不可见水印系统旨在将秘密载荷嵌入图像，使生成的含水印图像在感知上与原图无法区分，同时载荷在常见后处理和操作后仍可恢复（Zhu 等人，2018；Tancik 等人，2020；Luo 等人，2020；Bui 等人，2023年）。历来，鲁棒性评估主要针对一系列相对“低级”的扰动和信号处理操作：JPEG压缩、调整大小、像素噪声、轻度模糊、裁剪及其组合。如今，这种鲁棒性评估范式已不再全面。

扩散模型（Ho 等，2020；Song 等，2020；Rombach 等，2022）催生了新一代图像编辑系统，这些系统实质上是生成式变换：编辑后的输出并非只是输入经过后期处理的版本，而是重新合成的图像；它在满足高级约束（指令、掩码、参考主体或几何拖拽操作）的同时，将中间表征重新投影到学习得到的自然图像流形上。InstructPix2Pix（Brooks 等人，2023）等基于指令的编辑器只需极少的用户输入，便可实现全局语义变化；基于反演的方法则可对给定的真实图像进行可控编辑（Kawar 等，2023；Mokady 等，2023）；DragDiffusion（Shi 等，2024）和 DragonDiffusion（Mou 等，2024）等交互式系统通过在扩散潜变量中进行优化来实现几何操作；而 TF-ICON（Lu 等，2023）、SHINE（Lu 等，2025b）和 DragFlow（Zhou 等，2025c）等近期免训练合成流水线，则利用日益强大的先验（包括扩散变换器），以高保真度插入和移动物体。这些方法拓展了实际图像修改的范围，而经典失真模型只能较为有限地描述这类修改。

近期安全导向研究发出警示：基于生成模型的再生成式攻击可能移除隐形水印（或至少使其无法解码）（Zhao 等人，2024；Ni 等人，2025a；Fu 等人，2025a）。从宏观上看，此类攻击会注入噪声，然后使用去噪器或生成模型重建图像，从而破坏低级水印特征，同时保留图像的语义。与此同时，学界提出了明确利用扩散先验来增强鲁棒性的水印方案，例如 VINE（Lu 等人，2024b）以及潜空间水印方法（Zhang 等人，2024；Gunn 等人，2024）。尽管取得了这些进展，编辑场景仍引入了新的失效模式：水印可能在普通创意编辑过程中意外受损，而用户并无明确移除水印的意图。

本文从理论和实证角度分析了扩散式图像编辑过程中水印的脆弱性。我们的目标并非提供移除水印的操作指南，而是理解其失效机制，并为稳健的水印设计与部署提出建设性建议。我们重点研究了三类代表现代流程、且符合所需侧重点的扩散编辑器：（i）免训练合成（TF-ICON (Lu等人，2023年)）；（ii）借助强 DiT／流先验实现的插入与协调合成（SHINE (Lu等人，2025b)）；以及（iii）在 DiT 中结合区域监督的交互式拖动编辑（DragFlow (Zhou 等人，2025c)）。我们研究了这些编辑器与具有代表性的水印编码器／解码器之间的相互作用，涵盖经典深度水印（HiDDeN (Zhu 等人，2018)）、物理鲁棒嵌入（StegaStamp (Tancik 等人，2020)），以及近期的通用高分辨率水印（TrustMark (Bui 等人，2023年；2025年)），同时也涵盖基于扩散先验的水印（VINE (Lu等人，2024b)）。

贡献。 我们的主要贡献如下：(i) 我们将基于扩散的图像编辑流水线形式化为结合加性噪声损坏与流形投影的随机算子，并确定其中哪些步骤对像素级水印的破坏最大。(ii) 我们构建了一个信息论模型并给出证明，表明水印载荷信息在扩散编辑过程中会收缩，在温和假设下，解码错误将趋近于随机猜测。(iii) 我们提出了一套可复现的实验方案，并提供切合实际的假设性表格，用于衡量多种水印方法和扩散编辑方法的水印鲁棒性与视觉保真度。(iv) 我们将这些启示归纳为可操作的韧性水印设计指南，并探讨有关评估、披露与部署的伦理考量。

2. 相关工作

2.1. 基于扩散的图像编辑

扩散模型已成为高质量图像合成的主流范式，并通过条件控制与引导实现可控生成（Ho 等人，2020；Song 等人，2020；Rombach 等人，2022；Ho &Salimans，2022）。将这些模型应用于真实图像编辑，催生了多类方法。基于反演的方法将图像映射为扩散潜变量（或噪声）表示，再根据编辑提示词进行重新采样；代表性方法包括 Imagic（Kawar etal.，2023年）和空文本反演（Mokady et al.，2023年）。基于指令的编辑器（如 InstructPix2Pix（Brooks 等人，2023年））通过合成指令对训练条件扩散模型，无需针对每张图像进行优化即可快速编辑。交互式编辑进一步拓展了扩散工作流：DragDiffusion（Shi等人，2024）和 DragonDiffusion（Mou等人，2024）会优化扩散潜变量，以实现点级或基于特征对应的几何操作。

近期工作还强调免训练的编辑与合成，即在极少（或零）微调的情况下使用强大的预训练模型。TF-ICON（Lu等人，2023年）通过利用反演策略和注意力注入实现跨域图像引导合成，而SHINE（Lu等人，2025b）则利用现代扩散变换器与流匹配更强的先验，在无需反演的情况下以高保真度完成物理上合理的插入。DragFlow（Zhou 等人，2025c）认为，DiT 特征几何会使先前的逐点拖拽目标失效，并引入基于区域的监督与适配器增强反演，从而在DiT中实现最先进的拖拽编辑。

2.2. 鲁棒隐形水印与基于深度学习的方案

传统水印通过频域或扩频设计来嵌入能够抵抗常见失真的信号，但现代系统越来越依赖深度网络，以联合优化不可感知性、容量和鲁棒性（Zhu 等人，2018；Tancik 等人，2020；Luo 等人，2020；Ren 等人，2025年）。HiDDeN（Zhu 等人，2018）展示了结合可微失真层对编码器和解码器进行端到端训练的方法。StegaStamp（Tancik 等人，2020）通过针对真实的光度和几何失真进行训练，重点提升打印与再采集场景下的鲁棒性。失真无关训练（如（Luo 等人，2020）所采用的方法）利用对抗训练和信道编码，提高对未知失真的泛化能力。TrustMark（Bui 等人，2023年；2025年）引入了面向通用水印的架构损失和空间-频谱损失，以适应不同分辨率，并提供了一个配对的水印移除网络，用于再水印工作流。

2.3. 水印对生成式变换的鲁棒性

随着生成模型在图像到图像变换方面变得日益有效，研究界开始质疑，面对再生成攻击时，图像水印是否仍然具有意义。Zhao 等人表明，可以利用生成式人工智能通过一系列再生成攻击，在理论上移除不可见的像素级水印（Zhao 等人，2024）。随后，针对扩散的分析和方法表明，基于扩散的编辑与再生成会严重降低水印可检测性（Ni 等人，2025a；Fu 等人，2025a；Ni 等人，2025b）。与此同时，一些研究尝试利用扩散模型来提升水印鲁棒性，例如将水印嵌入扩散潜空间（Zhang 等人，2024），或利用潜变量初始化中的密码编码（Gunn 等人，2024）。VINE（Lu 等人，2024b）明确研究了针对编辑的鲁棒性，并提出利用代理失真和扩散先验进行嵌入。

2.4. 扩散编辑与水印鲁棒性的基准测试

随着扩散编辑日趋多样化，多个基准测试相继出现，用于量化编辑保真度、身份保持和可控性。EditVal（Basu等人，2023年）为文本引导编辑评估提供了标准化流程，涵盖多种编辑类型，并指出许多编辑器难以处理空间变换；这一失效模式与水印能否经受几何或局部编辑直接相关。更广泛而言，诸如（Huang等人，2024）的综述根据监督类型（文本、掩码、示例、几何）以及是否依赖反演、测试时优化或直接前馈编辑，对扩散编辑器进行分类。在水印方面，与 VINE 一同推出的 W-Bench 明确评估水印方案对一系列现代编辑算子的抵抗能力，而非仅考察经典失真（Lu等人，2024b）。NeurIPS“抹去不可见”挑战赛（Ding等人，2024）等社区压力测试，以及随后发布的获胜攻击报告（Shamshad 等人，2025年），进一步表明，水印鲁棒性应通过针对自适应和生成式变换的实证评估来理解，而不能只考察固定破坏模型。

扩散模型中的概念擦除。 扩散模型中的概念擦除与本文主题密切相关，因为概念擦除和水印保留都取决于信息如何沿去噪轨迹注入、抑制或保留；在编辑时介入以擦除语义概念，也可能作为副作用擦除（或扭曲）承载水印的信号。MACE（Lu等人，2024年）通过交叉注意力细化和基于 LoRA 的扩展，将概念移除推广到大量概念，；ANT（Li等人，2025a）提出轨迹感知微调，在保留早期阶段分数场的同时引导中后期去噪阶段；EraseAnything（Gao等人，2024；2025年）则通过双层优化和注意力正则化，针对整流流变换器模型。虽然这些工作侧重于抑制不需要的概念，但其机制——注意力操控、引导重加权和轨迹引导——在结构上与扩散编辑控制以及可能破坏水印的动态过程都有重叠。这种密切关联表明存在一个更广泛的系统性问题：对扩散流水线进行安全驱动的修改，可能会无意中影响水印可靠性；而水印驱动的控制也可能与安全驱动的轨迹编辑发生冲突。

3. 方法

3.1. 问题设定

设 $x \in \mathbb{R}^{H\times W\times 3}$ 为干净图像（像素值归一化至 $[0,1]$），并设 $m \in \{0,1\}^B$ 表示一个 B位载荷。水印方案由编码器 E 和解码器 D组成：

$$\tilde{x} = E(x, m; k), \tag{1}$$
$$\hat{m} = D(y; k), \tag{2}$$

其中 k 是编码器与解码器共享的密钥（或秘密参数）。我们假设不可感知性： $\tilde{x}$ 在视觉上接近 x（依据标准度量指标，例如峰值信噪比（Hor´e与Ziou，2010年）、结构相似性指数（Wang等人，2004年）以及学习感知图像块相似度（Zhang等人，2018年））。我们假设对常规失真族具有鲁棒性 $\tau \sim \mathcal{T}$，例如 JPEG、缩放、裁剪、高斯噪声：

$$\mathbb{P}\left[D(\tau(\tilde{x}); k) = m\right] \geq 1 - \epsilon. \tag{3}$$

本文研究基于扩散的编辑算子 T下的鲁棒性：

$$y = T(\tilde{x}; c, \omega), \tag{4}$$

其中 c 编码用户意图（文本指令、掩码、参考图像、拖动约束），且 ω 表示随机性（随机种子、扩散噪声、随机采样）。示例包括 TF-ICON 合成（Lu等人，2023年）、SHINE 插入（Lu等人，2025b），以及 DragFlow 拖拽编辑（Zhou 等人，2025c）。

3.2. 作为两阶段随机信道的扩散编辑

大多数扩散编辑流水线（包括基于反演、基于优化和免训练变体）都可以抽象为以下两步的合成：(i) 一个编码/加噪阶段，将噪声注入中间表示；(ii) 一个去噪/投影 阶段，将其映射回学习流形上的图像。

具体而言，考虑一个潜在扩散工作流（Rombach 等，2022）。令 $z_0 = \mathrm{Enc}(\tilde{x})$ 为一个潜变量编码器。执行到时间步 t的前向扩散步骤会产生

$$z_t = \sqrt{\bar{\alpha}_t}\, z_0 + \sqrt{1-\bar{\alpha}_t}\, \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, I), \tag{5}$$

其中 $\bar{\alpha}_t$ 是累积噪声调度（Ho 等人，2020）。随后，编辑器应用带条件控制的反向过程c （并可选择额外的优化或引导）以生成 $\hat{z}_0$。最后， $y = \mathrm{Dec}(\hat{z}_0)$。

我们从概念上将编辑算子分解为一个马尔可夫核：

$$\tilde{x} \xrightarrow{\mathrm{Enc}} z_0 \xrightarrow{\mathrm{Noise}_t} z_t \xrightarrow{\mathrm{Denoise}_c} \hat{z}_0 \xrightarrow{\mathrm{Dec}} y. \tag{6}$$

即使 $\mathrm{Denoise}_c$ 在感知层面上具有很强的内容保持能力，也不受约束去保留编码水印载荷的低能量高频扰动。事实上，学习型去噪器经过训练，旨在去除类似噪声的信号；精心设计的不可见水印在构造上与这类信号相似。

3.3. 水印衰减的信号模型

深度水印领域的一个普遍经验观察是，水印通常以微小扰动 δ 的形式实现：

$$\tilde{x} = x + \delta(x, m; k), \tag{7}$$

其中 $\|\delta\|$ 较小，以保持不可感知性。在潜在空间中，对 Enc 在 x附近进行一阶近似，可写为

$$z_0(\tilde{x}) \approx z_0(x) + J_x \delta, \tag{8}$$

其中 J_x 是局部雅可比矩阵。前向扩散步骤会将扰动缩放 $\sqrt{\bar{\alpha}_t}$:

$$z_t(\tilde{x}) - z_t(x) \approx \sqrt{\bar{\alpha}_t}\, J_x \delta. \tag{9}$$

与此同时，注入噪声的每维方差为（$1 - \bar{\alpha}_t$）。因此，时刻 t 的一个简单信噪比（SNR）近似指标为

$$\mathrm{SNR}(t) \triangleq \frac{\bar{\alpha}_t \|J_x \delta\|^2}{(1-\bar{\alpha}_t)\, d}, \tag{10}$$

其中 d 是潜在维度。对于许多编辑流水线所用的噪声水平（通常对应中到高 t ，以允许语义变化）， $\bar{\alpha}_t$ 可能较小，而（$1 - \bar{\alpha}_t$）较大，从而使信噪比趋近于零。在这种情况下，水印扰动在统计上将无法与扩散噪声区分。

3.4. 特定编辑流程的考量：TF-ICON、SHINE、DragFlow

尽管抽象的信道视角是共享的，但具体的编辑方法会以不同方式实例化这些组件。

TF-ICON。TF-ICON（Lu等人，2023年）通过将真实图像反演为扩散潜变量，并在去噪过程中注入修改后的注意力图，实现免训练的跨域组合。有两个方面与此相关：反演会引入显式加噪与反向去噪（可能包含多个步骤），而注意力操控会重新加权语义与空间特征的传播。即便合成输出在视觉上保持忠实，这两者都可能破坏水印扰动。

SHINE。 SHINE（Lu等人，2025b）强调无需反演的潜变量准备：先执行一步前向扩散以获得噪声潜变量，然后使用流形引导锚点损失和辅助引导项对其进行优化。显式的一步扩散会注入大量高斯噪声，在更高分辨率下以及 Transformer 先验更强时尤其如此；随后进行的轨迹引导还可能进一步“正则化掉”那些与水印一致、但在视觉上无关紧要的特征。

DragFlow。 DragFlow（Zhou 等人，2025c）在 DiT 模型中引入了基于区域的监督和适配器增强反演，用于拖拽编辑。与其他基于优化的拖拽编辑器一样，它会迭代优化噪声潜变量以满足几何约束，这可能改变局部频率内容和交互模式。重要的是，DragFlow 针对强生成先验而设计，因此倾向于将中间潜变量大幅投影到自然图像流形上；这一过程可能会抑制那些与语义上有意义的内容无关的水印信号。

3.5. 扩散编辑操作与水印失效模式的分类

基于扩散的编辑并非单一操作，而是一系列将基础生成模型与特定任务控制相结合的流程。为了分析水印脆弱性，可以从三个影响信息保留的维度对编辑器进行分类。

轴 A：噪声注入的位置。有些编辑器在像素空间注入噪声（例如图像到图像扩散流水线），而潜变量扩散编辑器则在压缩后的 VAE 潜变量中注入噪声（Rombach 等人，2022年）。SHINE 明确在流匹配潜变量中执行一步前向扩散（Lu等人，2025b），而许多基于反演的编辑器则通过迭代反演（DDIM 风格）构造噪声潜变量（Mokady 等人，2023年）。从水印的角度来看，关键量是注入点处的有效信噪比；在语义压缩后注入噪声，甚至会更加剧烈地抹除水印的高频分量。

轴 B：模型被要求“创作”多少内容。编辑操作的范围从近恒等调整（低噪声、小幅提示词变化）到高熵生成（高噪声、强指令或大面积掩码区域）。为合成而设计的方法（例如 TF-ICON 和 SHINE）通常需要模型臆造边界区域、阴影、反射并进行纹理协调（Lu等人，2023年；2025b）。这些操作在语义上有意义，但可能会覆盖那些与模型解释因素不一致的细微信号。

轴 C：编辑器是否会优化潜变量。基于优化的编辑器（DragDiffusion、DragFlow、DragonDiffusion）会迭代优化潜变量，以满足几何约束（Shi等人，2024；Zhou 等人，2025c；Mou等人，2024）。这引入了另一种水印破坏机制：编辑约束所产生的梯度可能会将潜变量推向水印解码器决策边界不稳定的区域，即使最终图像在感知上相似。无需优化的免训练方法仍可通过加噪和去噪去除水印，但优化会加剧其脆弱性。

沿着这些轴，我们发现了水印评估中反复出现的三种实际失效模式：（i）信噪比崩塌（前向噪声淹没水印载体），（ii）流形投影（反向过程丢弃干扰扰动），以及（iii）控制诱导的重新分布（注意力/适配器引导将概率质量集中于语义显著因素，从而无意中抹去正交信号）。我们的理论结果主要关联（i）和（ii），而我们的实验方案（算法1）旨在通过实证揭示（iii）。

3.6. 评估方案与可复现性目标

我们的评估目标是同时衡量两项指标：(i) 水印存活性（比特准确率或检测分数），以及 (ii) 编辑后输出相对于无水印编辑的视觉保真度。这一点至关重要，因为通过破坏图像来“移除”水印毫无意义。因此，我们在相同的编辑条件 c 和随机种子 ω下，定义两次配对运行：

$$y_{\text{wm}} = T(E(x, m; k); c, \omega), \tag{11}$$
$$y_{\text{clean}} = T(x; c, \omega). \tag{12}$$

我们评估恢复出的载荷 $\hat{m} = D(y_{\text{wm}}; k)$，并计算 y_{wm} 与 y_{clean} 之间的视觉相似度（而非 y_{wm} 与 x之间的相似度），以区分水印引起的伪影与编辑引起的变化。

算法 1 扩散编辑水印压力测试（所用方案）

要求： 干净图像数据集 $\{x_i\}_{i=1}^{N}$，载荷长度 B，水印编码器/解码器（E，D），密钥为 k，编辑算子集合 $\mathcal{T}_{\text{edit}}$，编辑条件 $\mathcal{C}$（文本提示词、掩码、拖拽操作），随机种子 Ω

确保： 水印鲁棒性指标和水印引起的保真度指标

对于 $i \leftarrow 1$**到** N **执行**
  采样载荷 $m_i \sim \text{Unif}(\{0,1\}^B)$
  $\tilde{x}_i \leftarrow E(x_i, m_i; k)$
  对所有 $T \in \mathcal{T}_{\text{edit}}$ **执行**
    对所有 $c \in \mathcal{C}(T)$ **执行**
      对所有 $\omega \in \Omega$**执行**
        $y_{\text{wm}}^{(i)} \leftarrow T(\tilde{x}_i; c, \omega)$
        $y_{\text{干净}}^{(i)} \leftarrow T(x_i; c, \omega)$
        $m_i \leftarrow D(y_{\text{wm}}^{(i)}; k)$
        计算比特准确率 $\text{BA}_i \leftarrow \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}[\hat{m}_{i,b} = m_{i,b}]$
        计算保真度度量指标峰值信噪比($y_{\text{wm}}^{(i)}$, $y_{\text{干净}}^{(i)}$)、结构相似性指数(·)、学习感知图像块相似度(·)
      结束循环
    结束循环
  结束循环
结束循环
汇总不同图像、条件和随机种子下的鲁棒性与保真度统计数据

4. 实验设置

4.1. 结果范围与免责声明

实验部分采用详尽且可复现的写作风格，但数值结果是假设性的，旨在符合先前关于基于扩散模型的水印去除和基准测试的研究所报告的趋势，并尽可能贴近实际（Zhao etal.，2024；Ni et al.，2025a；Fu et al.，2025a；Lu等人，2024b）。这一选择遵循了纳入假设性表格、同时保持实验结构可信的指令。我们强调，该实验方案可通过所引方法的公开实现来执行。

4.2. 数据集与预处理

我们将 W-EDITBENCH 定义为水印和编辑评估中常用的自然图像来源的合集：(i) 用于自然场景的 MS-COCO val2017（5k 张图像），以及 (ii) 从大型文生图数据集（如 DiffusionDB）中选取的 5k 张图像子集，用于代表 AI 生成或带有说明文字的图像。所有图像均经过缩放，使较短边为 512 像素，然后从中心裁剪为 512×512。

4.3. 水印方法

我们评估提示中强调的三种代表性水印方法：**StegaStamp**（Tancik 等人，2020），这是一种针对真实失真进行训练、具有物理鲁棒性的神经水印；**TrustMark**（Bui 等人，2023年；2025年），这是一种适用于任意分辨率、并具有强鲁棒性声明的通用水印；以及**VINE**（Lu 等人，2024b），这是一种扩散先验水印方法，专门针对类似编辑的代理失真进行训练。为求完整，我们还报告了**HiDDeN**（Zhu 等人，2018）的结果，将其作为广受引用的基线。

除非另有说明，所有方法均嵌入一个 $B = 100$ 比特载荷。对于默认容量较低的方法（例如 StegaStamp 的原始设计），我们通过纠错编码将 100 比特载荷映射到该方法支持的容量；这一设置符合鲁棒水印实践（Luo 等人，2020）。

4.4. 基于扩散的编辑方法

我们从四类评估扩散编辑：(i) **组合/插入：** TF-ICON（Lu等人，2023年）和 SHINE（Lu等人，2025b）；(ii) **拖拽式编辑：** DragFlow（Zhou 等人，2025c）（基于 DiT 先验的区域拖拽）和 DragDiffusion（Shi等人，2024）（基于 UNet 的拖拽编辑），作为参照；(iii) **指令编辑：** InstructPix2Pix（Brooks 等人，2023年），用于全局语义编辑；(iv) **再生成式编辑：** 采用一种简单的图像到图像扩散流水线，与（Zhao 等人，2024；Fu 等人，2025a）中研究的再生成攻击一致。

每种编辑器都有一个编辑强度超参数，通常通过噪声水平、时间步或引导尺度实现。我们将强度离散为三档：**低**、**中**、**高**，分别对应 (a) 轻微的语义调整、(b) 常规编辑和 (c) 大幅变换。

4.5. 度量指标

水印鲁棒性。我们报告平均比特准确率（BA，以 % 计），并针对输出置信度分数的方法报告检测类指标 TPR@0.1%FPR。对于仅输出比特的方法，我们根据汇总的比特一致性推导检测分数。

保真度。我们将嵌入水印的编辑结果 y_{wm} 与未修改的编辑结果 y_{clean} 进行比较，采用 PSNR（Hor´e & Ziou，2010）、SSIM（Wang 等人，2004）和学习感知图像块相似度（LPIPS）（Zhang 等人，2018）。

5. 结果

5.1. 传统失真下的基线鲁棒性

我们首先报告常规非生成式失真下的鲁棒性，以便 contextualize“鲁棒性差距”

表 *1.* W-EDITBENCH（512×512）上在传统失真下的假设鲁棒性。BA 为比特准确率，以 % 表示。数值越高越好。

方法	JPEG（Q=50）	缩放（0.6×）	裁剪（0.8 区域）	模糊（σ=1.0）
HiDDeN（Zhu 等人，2018）	96.8%	97.5%	94.2%	93.1%
StegaStamp（Tancik 等人，2020）	97.9%	98.2%	95.6%	94.4%
TrustMark(Bui 等人，2023年)	99.1%	99.3%	97.8%	96.9%
VINE (Lu等人, 2024b)	99.4%	99.5%	98.6%	97.5%

由扩散编辑引起的差距。表 1 显示，所有方法在 JPEG、缩放、裁剪和轻度模糊下均能达到较高的比特准确率，这与其设计目标一致（Tancik 等人，2020；Bui 等人，2023年；Lu等人，2024b）。如果认为“鲁棒性”可以泛化到扩散编辑，这些出色的结果可能会产生误导。

5.2. 基于扩散的编辑中的水印存活性

表2 总结了所要求的扩散编辑器下基于扩散的编辑中的水印存活性。由此呈现出两个一致的模式。

首先，即使是低强度编辑，也会显著降低水印解码性能，尤其是对于针对经典失真优化的水印方案（HiDDeN、StegaStamp、TrustMark）。这与再生成攻击文献（Zhao 等人，2024；Ni 等人，2025a；Fu 等人，2025a）一致。

其次，扩散感知设计带来的鲁棒性提升确实存在，但受到限制。在编辑过程中，VINE 始终比基线更具鲁棒性，但在中到高编辑强度下，其比特准确率会接近随机水平（对于比特而言接近 50%），这与相关报告一致：足够强的扩散编辑可以绕过甚至是扩散先验水印（Ni 等人，2025a；b）。

5.3. 保留水印的编辑与破坏水印的编辑之保真度

为区分“图像遭到破坏导致水印失效”与“内容得以保留但水印仍然失效”，我们报告了以下两者之间由水印引起的差异：y_{wm} 和 y_{clean}。表3 显示，在不同编辑器中，水印对编辑后的输出仅有边际影响（峰值信噪比高、结构相似性指数高、学习感知图像块相似度低）。因此，即使编辑后的输出在有无水印的情况下基本相同，扩散编辑仍可能导致水印解码失败。这表明，编辑算子在保留语义的同时移除了类水印成分。

5.4. 水印存留曲线与噪声水平的关系

我们的理论表明，前向扩散噪声水平是水印信息损失的主要驱动因素（方程（10））。表 4 报告了最简单的再生式编辑器在不同参数下的假设存活曲线

表 2.W-EDITBENCH 上基于扩散的编辑对水印鲁棒性的假设性结果（比特准确率，%）。“低/中/高”对应逐渐增强的编辑强度。随机水平比特准确率为 50%。

方法	TF-ICON（Lu等人，2023年）			SHINE（Lu等人，2025b）			DragFlow（Zhou 等人，2025c）			InstructPix2Pix（Brooks 等人，2023年）		
	Low	Med	High	Low	Med	High	Low	Med	High	Low	Med	High
HiDDeN（Zhu 等人，2018）	74.1%	56.8%	50.9%	71.5%	55.4%	50.3%	69.2%	54.7%	50.1%	78.3%	60.2%	51.5%
StegaStamp (Tancik 等人, 2020)	79.6%	58.1%	51.2%	76.8%	56.2%	50.4%	73.5%	55.0%	50.2%	82.7%	62.1%	52.0%
TrustMark（Bui 等人，2023年）	84.8%	60.5%	52.0%	82.9%	59.2%	51.1%	79.1%	57.3%	50.6%	86.0%	64.4%	53.2%
VINE (Lu等人, 2024b)	90.2%	69.1%	54.8%	88.7%	67.5%	53.6%	86.4%	65.9%	52.8%	90.8%	71.2%	55.7%

表3。水印与干净编辑输出之间的假设保真度，取 W-EDITBENCH 和各编辑条件下的平均值。度量指标比较 y_{wm} 与 y_{clean}。峰值信噪比/结构相似性指数越高、学习感知图像块相似度越低，效果越好。

方法	TF-ICON		SHINE		DragFlow		InstructPix2Pix	
	峰值信噪比 ↑	学习感知图像块相似度 ↓	峰值信噪比 ↑	学习感知图像块相似度 ↓	峰值信噪比 ↑	学习感知图像块相似度 ↓	峰值信噪比 ↑	学习感知图像块相似度 ↓
StegaStamp	42.8	0.012	42.5	0.013	41.9	0.015	43.1	0.011
TrustMark	43.6	0.010	43.3	0.011	42.7	0.013	43.8	0.010
VINE	44.1	0.009	43.9	0.010	43.1	0.012	44.3	0.009

表4. 再生成式扩散编辑下假设BA（%）随噪声强度的变化。

方法 $t_{\max}$=0.20	$t_{\max}$=0.35	$t_{\max}$=0.50	$t_{\max}$=0.65
StegaStamp 85.2%	63.4%	52.1%	50.4%
TrustMark 88.9%	66.7%	53.0%	50.7%
VINE 93.1%	74.2%	56.3%	51.6%

最大时间步 $t_{\max}$（或等效的噪声强度）。如预期，随着噪声水平增加，解码准确率迅速崩塌至随机猜测水平；扩散感知水印（VINE）带来的改进会使崩塌有所延缓，但无法消除这一现象。

5.5. UNet 与 DiT 编辑器：更强的先验会加剧水印脆弱性

近期编辑研究中一个反复出现的假设是，扩散 Transformer（DiT）和整流流模型比早期基于 UNet 的潜空间扩散模型具有更强的生成先验，因此能够实现更高的保真度并减少伪影（Lu 等人，2025b；Zhou 等人，2025c）。从水印的角度看，更强的先验有利有弊：它们能提升感知真实感，但也可能进一步降低低能量水印扰动得以保留的可能性，因为流形投影会变得更具决定性。

表5 对比了在拖拽任务和输出保真度相匹配的情况下，基于 UNet 的编辑器（DragDiffusion）与基于 DiT/Flow 的编辑器（DragFlow）。我们观察到，在学习感知图像块相似度相近时，基于 DiT 的编辑器对传统像素级方案产生的水印比特准确率更低，这与更强的先验能够更可靠地将输出“吸回”无水印流形这一观点相符。

5.6. 与重新嵌入水印和移除网络的交互

一些水印系统会明确提供移除器，以支持受控流水线中的重新嵌入水印。例如，Trust-Mark 包含一个移除模型，以便在可信工作流中重新嵌入水印（Bui 等人，2025年）。在扩散 editi[ng] 上下文中，t his 容量 bili[ty] 会产生一个 su btle 系统问题：如果编辑过程隐式移除了水印，能否通过重新嵌入水印恢复来源溯源，同时避免产生叠加伪影?

表 6 报告了一种假设流程，其中 (i) 嵌入 Trust-Mark 水印，(ii) 图像经过 SHINE 或 TF-ICON 编辑，(iii) 输出重新嵌入水印。我们测量重新嵌入水印后的输出与无水印编辑之间的峰值信噪比，以及新水印的解码准确率。结论是，重新嵌入水印在视觉质量方面是可行的，但会改变来源溯源的语义：结果中的水印对应的是编辑后的内容，而非原始来源。

5.7 消融：容量与不可感知性权衡

一种可能的做法是提高水印能量（从而提高信噪比）以增强鲁棒性，但这可能会损害图像质量和可检测性。表 7 通过改变通用编码器的载荷长度和嵌入强度，并报告中等强度编辑下的鲁棒性及视觉失真，展示了这种权衡。由此可见，单纯“增强水印”并非万全之策：扩散编辑仍可能削弱并忽略强扰动，而感知指标可能迅速恶化。

表 5. 在拖拽编辑条件相同的情况下，对 UNet 与 DiT 进行假设性比较。我们报告水印编辑与无水印编辑之间的比特准确率 (%) 和学习感知图像块相似度。

方法	DragDiffusion(UNet) (Shi等人，2024)		DragFlow(DiT/Flow)(Zhou 等人，2025c)		差距（DiT−UNet）	
	BA ↑	学习感知图像块相似度 ↓	BA ↑	学习感知图像块相似度 ↓	BA	学习感知图像块相似度
StegaStamp	62.8%	0.014	55.0%	0.015	−7.8	+0.001
TrustMark	66.9%	0.012	57.3%	0.013	−9.6	+0.001
VINE	73.4%	0.010	65.9%	0.012	−7.5	+0.002

表6.扩散编辑后的假设性重新嵌入水印（TrustMark）。

编辑器	强度	峰值信噪比 ↑	新水印的平衡准确率 ↑
TF-ICON	Low	43.7	98.9%
TF-ICON	Med	43.4	98.6%
SHINE	Low	43.5	98.8%
SHINE	Med	43.1	98.4%

6. 理论证明

现在，我们给出理论结果，以支持这一实证观察：基于扩散的编辑会严重破坏像素级水印信息。我们的证明采用“信道”视角：水印将消息嵌入图像中；扩散编辑施加随机变换；根据数据处理，不可能增加有关该消息的信息。

6.1. 符号与假设

令 $M \in \{0,1\}^B$ 为均匀随机选取的水印消息。令 X 表示编码器生成的含水印图像（或潜变量）。令 Y 表示基于扩散的编辑器的输出。我们将整个编辑流程视为一个马尔可夫核 $P_{Y|X}$，其中包含内部随机性（例如扩散噪声、采样随机性、随机种子）。因此，$M \to X \to Y$ 是一个马尔可夫链。

我们作出两个温和的假设。首先，编码器具有不可感知性：对于每个 m，$X|M{=}m$ 都处于干净图像分布的一个小邻域内。其次，编辑器在某个中间表示 Z 处包含一个显式的高斯加噪步骤：

$$Z = \sqrt{\bar{\alpha}}\, X + \sqrt{1-\bar{\alpha}}\, \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, I), \tag{13}$$

对于某个 $0 < \bar{\alpha} < 1$。这一假设适用于潜变量扩散编辑器（Rombach 等人，2022），并且在执行前向扩散以实现语义变化的编辑方法中是明确成立的（Brooks 等人，2023年；Lu 等人，2025b；Zhou 等人，2025c）。编辑流程的其余部分通过任意的（可能是非线性的）随机化算法 Z 将 Y 映射为 Q（例如，结合条件控制和引导的反向扩散），因此 $M \to X \to Z \to Y$ 具有马尔可夫性。

6.2. 高斯加噪下可区分性的收缩

我们首先证明，如果两条水印消息在 X 上诱导出的分布彼此接近（正如不可感知性所表明的那样），那么经过高斯加噪后，它们会变得更加接近。

引理 1（加性高斯信道中的 KL 收缩）。设 Z 由 X 根据方程(13)生成。对于任意两个定义在 P 和 Q 上的分布 X，诱导出的、定义在 Z 上的分布之间的 KL 散度满足

$$D_{\mathrm{KL}}(P_Z \,\|\, Q_Z) \le \frac{\bar{\alpha}}{1-\bar{\alpha}} \cdot \mathbb{E}_{x\sim P,\, x'\sim Q}\left[\frac{1}{2}\|x - x'\|_2^2\right], \tag{14}$$

其中 P_Z 和 Q_Z 表示经过高斯信道后的推前分布。

证明。 对于协方差相同的加性高斯信道，两个条件高斯分布之间的似然比具有二次型。经过标准计算可得，对于确定性输入 x 和 x'，

$$D_{\mathrm{KL}}\left(\mathcal{N}(\sqrt{\bar{\alpha}}x, (1-\bar{\alpha})I) \,\|\, \mathcal{N}(\sqrt{\bar{\alpha}}x', (1-\bar{\alpha})I)\right) = \frac{\bar{\alpha}}{2(1-\bar{\alpha})}\|x - x'\|_2^2. \tag{15}$$

对于一般输入分布，利用 KL 散度的凸性，对 (x, x') 取期望后即可得到所述界。 □

随后，Pinsker 不等式给出全变差距离的上界：

$$\|P_Z - Q_Z\|_{\mathrm{TV}} \le \sqrt{\frac{1}{2} D_{\mathrm{KL}}(P_Z \,\|\, Q_Z)}. \tag{16}$$

因此，除非不同水印引起较大的 ℓ_2 分离（这与不可感知性相冲突），否则加噪后的分布会变得接近，难以区分。

6.3. 信息衰减与解码不可能性

接下来，我们对 M 与编辑后的输出 Y 之间的互信息进行界定。

定理1（扩散编辑下的信息损失）。假设编辑流程包含一个形式为 $X \to Z$ 方程（13）的高斯加噪步骤，随后是任意的随机映射 $Z \to Y$。则

$$I(M;Y) \le I(M;Z) \le I(M;X), \tag{17}$$

表 7. 假设的容量–不可感知性–鲁棒性权衡（通用深度水印基线）。

载荷 B	嵌入强度	相对于 Cover 的峰值信噪比 ↑	相对于 Cover 的学习感知图像块相似度 ↓	SHINE（中）下的比特准确率 ↑	DragFlow（中）下的比特准确率 ↑
50	Low	44.8	0.008	63.0%	61.2%
50	High	39.2	0.026	70.1%	68.5%
100	Low	43.1	0.012	59.2%	57.3%
100	High	37.9	0.031	66.8%	64.7%
200	Low	41.0	0.019	55.8%	54.2%
200	High	35.6	0.044	61.5%	59.3%

而且 $I(M;\ Z)$ 按 $\bar{\alpha}\downarrow 0$ 衰减。特别地，若水印编码器是不可感知的，即对于所有 m, m'，期望平方距离 $\mathbb{E}\|X_m - X_{m'}\|_2^2$ 为 $O(\epsilon^2)$，则 $I(M;\ Z) = O\left(\frac{\bar{\alpha}}{1-\bar{\alpha}}\epsilon^2\right)$。

证明。 马尔可夫链 $M \to X \to Z \to Y$ 由数据处理不等式蕴含$I(M;\ Y) \le I(M;\ Z)$，并且由于 $M \to X \to Z$，$I(M;\ Z) \le I(M;\ X)$。为了界定 $I(M;\ Z)$，我们可以写出

$$I(M;Z) = \mathbb{E}_m\left[D_{\mathrm{KL}}(P_{Z|M=m} \,\|\, P_Z)\right], \tag{18}$$

并应用引理1 以及由不可感知性引起的条件分布之间的微小分离（例如，参见（Cover & Thomas，2006）第2章中的标准处理）。 $\bar{\alpha}/(1-\bar{\alpha})$ 中的显式缩放由高斯信道KL公式得出。 □

6.4. 多步扩散与信息指数衰减

上述抽象采用单个高斯步骤来突出该机制。实际的扩散编辑器通常应用多个加噪与去噪步骤，且正向过程是一条马尔可夫链（Ho 等人，2020；Song 等人，2020）。我们现在陈述一个多步结果，以刻画随有效加噪步数发生的指数衰减。

命题 1 (高斯步骤间的指数收缩)。 考虑潜在空间中的一系列 T 高斯加噪步骤，每一步的形式为

$$Z_t = \sqrt{\alpha_t}\, Z_{t-1} + \sqrt{1-\alpha_t}\, \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, I), \quad t = 1, \ldots, T, \tag{19}$$

其中 $0 < \alpha_t < 1$。令 Z_T 为最终的加噪表示，令 Y 为通过从 Z_T 进行任意随机映射得到的结果（例如，以编辑为条件的反向扩散）。那么，对于任意两个消息 m、m'，

$$D_{\mathrm{KL}}\left(P_{Z_T|M=m} \,\|\, P_{Z_T|M=m'}\right) \le \left(\prod_{t=1}^{T} \alpha_t\right) C, \tag{20}$$

其中 C 取决于相应 Z_0 分布的成对分离程度。特别地，若 $\prod_{t=1}^{T} \alpha_t$ 较小（如在中等到高噪声编辑中），则成对可辨识性会迅速丧失，且任何后续处理都无法通过数据处理将其恢复。

证明。 在每一步，条件分布与各向同性高斯分布卷积，且协方差匹配的两个高斯分布之间的KL 散度按相应的 α_t 因子收缩（见引理1证明中的确定性输入计算）。对于一般输入分布，耦合论证与KL散度的凸性相结合，给出一个包含 α_t 项之积的上界，反映了信号能量的反复衰减。最终映射 $Z_T \to Y$ 不能增大KL 散度。

□

命题1 可视为对“SNR塌缩”直觉的形式化：每个扩散加噪步骤都会削弱关于微小扰动的信息，而反复加噪会使表示趋近各向同性高斯分布，使所有不可感知的差异都消失。这与关于扩散编辑下水印脆弱性的信息论分析一致（Ni 等人，2025b）。

定理 2 (类 Fano 比特错误下界)。 设 M 在 $\{0,1\}^B$ 上服从均匀分布，且 $\hat{M}$ 是根据 Y 计算得出的任意估计量（包括经训练的水印解码器的输出）。那么，消息错误的概率满足

$$\mathbb{P}[\hat{M} \neq M] \ge 1 - \frac{I(M;Y) + \log 2}{B \log 2}. \tag{21}$$

因此，如果 $I(M;\ Y) \ll B$，恢复完整载荷的概率接近于零。

Proof. 这是经典的 Fano 不等式；参见（Cover &homas，T2006）。 □

阐释。 定理 1–2 将一个核心直觉形式化：对于不可感知的像素级载荷，扩散编辑是一种收缩信息的变换。即使编辑后的图像在感知上与对干净图像进行编辑所得的图像相近，加噪与重新合成过程仍可能有效地“重置”水印信道。这些结果与生成式重建能够可证明地移除不可见水印的结论一致（Zhao 等人，2024），并将这一认识从显式攻击扩展到常见的编辑流水线（Ni 等人，2025a；Fu 等人，2025a；Ni 等人，2025b）。

7. 讨论

7.1. 为什么扩散编辑对鲁棒水印而言具有独特挑战

传统失真保留了大量像素级信息：它们可能会压缩、下采样或扰动内容，但通常不会重新合成内容。深度水印方案利用这一点，通过参数化损坏族进行训练。扩散编辑从两个方面打破了这一假设。

首先，前向扩散/加噪步骤可能会使水印的信噪比（方程（10））骤降，从而在统计上使其不可见。其次，去噪步骤会采用一种学习先验，明确偏好自然图像，并将低层异常视为噪声。由于水印的设计目标是不可感知，因此它们不太可能被编码为鲁棒的语义结构；因此，模型在将图像投影到流形上时几乎没有动力保留水印。这解释了为什么即使编辑后的图像在视觉上忠实且质量很高，水印解码仍可能失败（表 3）。

7.2. D 更具韧性的水印设计指南 标记

我们的分析表明，“对经典失真具有鲁棒性”是必要但不充分的属性。基于现有方案和我们的理论，我们总结了几条建设性指南。

将信息嵌入与生成先验对齐的表征中。 嵌入扩散模型所用潜在空间的方案，可能比像素级扰动更能经受编辑（Zhang等人，2024；Gunn等人，2024）。然而，仅有“潜变量”还不够：如果编辑包含从高噪声水平重新采样，即使潜变量信息也可能被覆盖。

针对类似扩散的代理变换进行训练，但要校准强度。 VINE（Lu等人，2024b）利用代理模糊和扩散先验来提升对编辑的鲁棒性。我们的结果表明，这类训练会使存活曲线发生偏移（表 4），但可能无法消除高噪声下的根本性崩溃。设计者应在威胁模型中校准预期的编辑强度，而非假设方案具有完全的通用性。

将模型级或语义水印作为互补信号。 可被证明能够移除的像素级水印促使人们加入更高层次的溯源机制（Zhao 等人，2024），包括语义水印、模型级水印或密码学溯源元数据。这些机制可能无法取代像素级水印，但可以与其互补，尤其是在由生成模型执行编辑变换时。

将扩散变换检测作为备用方案。 如果预计在强扩散编辑后水印恢复会失败，那么一种实用的系统级指南是检测图像是否经历过基于扩散的变换，并将水印缺失视为含义不明确，而非据此认定其无罪。在高噪声重生成过程中检测生成式变换是一个开放问题，但与在此过程中保留任意像素级载荷相比，这可能更容易实现（Ni 等人，2025a）。

7.3. 伦理考量

关于水印脆弱性的研究本质上具有双重用途。一方面，这类研究能让防御方了解现实中的失效模式，并推动改进水印和溯源工具。另一方面，相关细节也可能被滥用于规避保护措施。因此，我们强调，我们的贡献是一个理论与评估框架，而非“攻击配方”；我们倡导负责任披露和基准测试实践，优先考虑防御、透明度与审计（Shamshad 等人，2025年）。我们还指出，某些水印系统（例如 TrustMark）明确支持为重新嵌入水印工作流而移除水印（Bui 等人，2025年）；这种设计选择有助于合法地重新建立来源溯源，但会使安全模型更加复杂，因此应通过访问控制和策略进行管理。

7.4. 溯源流水线部署指南

实际的来源溯源系统通常会结合不可见水印、基于元数据的机制和策略控制。我们的分析提出了若干部署层面的最佳实践。

首先，在扩散编辑普遍存在的环境中，不要将水印缺失解读为不存在来源溯源的证据。如果平台允许用户执行基于扩散的编辑，水印缺失可能仅仅表明所应用的编辑超出了鲁棒性范围。

其次，记录编辑操作，并在各种转换过程中维护来源溯源。如果编辑是在经过身份验证的环境中进行的（例如托管式创意套件），那么最可靠的来源溯源记录可能是编辑历史和经密码学签名的元数据，而水印则作为辅助信号。

第三，谨慎使用重新嵌入水印。虽然重新嵌入水印可以为编辑后的内容附加来源溯源（表6），但也会覆盖任何基于水印、指向原始来源的关联。系统应区分“衍生自”的来源溯源与“创作者”的来源溯源，理想情况下通过经签名的转换日志来实现。

最后，持续进行基准测试。由于扩散编辑器不断快速演进（新的骨干网络、调度方案和控制机制），因此应持续评估水印鲁棒性

针对当前编辑器及其默认设置重新评估，这与社区挑战赛的理念相似（Ding等人，2024；Shamshad 等人，2025年），也与 W-Bench 倡导的面向编辑的基准测试相一致（ Lu 等人，2024b）。

7.5. 局限性

我们的理论界限有意保持通用，因此无法涵盖水印方案和编辑器的具体架构细节。在实践中，某些水印设计可能会嵌入与语义特征相关的信息，因此能够部分经受某些编辑操作。我们的实证表格是假设性的，实际表现取决于具体实现、参数和数据集。最后，扩散编辑种类繁多且发展迅速；新的架构（例如整流流和高分辨率扩散 Transformer）可能会改变其中的权衡。

8. 结论

我们对基于扩散的图像编辑如何可能无意中损害鲁棒的不可见水印进行了理论分析和实证风格的分析。通过将扩散编辑建模为一种注入高斯噪声并投影到自然图像流形上的随机信道，我们推导出信息论结果，表明随着编辑强度增加，不可感知的像素级水印载荷与编辑后的输出之间的互信息会减少。我们设计了旨在便于复现的假设性实验，结果显示，具有代表性的水印方案（如 StegaStamp、TrustMark，甚至能够感知扩散的 VINE）在现代扩散编辑器（包括 TF-IC ON、SHINE 和 DragFlow）作用下，鲁棒性可能会急剧下降。我们得出结论，未来的水印方案必须明确考虑生成式变换：要么将水印嵌入能够经受此类变换的表征中，要么通过更高层级的溯源机制补充像素级载荷。

A. 更多理论细节

A.1. 从成对不可区分性到检测器下界

对于输出二元判定（水印存在或不存在）的检测器，可以根据总变差界推导出不可能性结果。假设 H_0 是从干净输入生成的编辑后图像分布，而 H_1 是从带水印输入生成的编辑后图像分布。任何检测器的错误率至少为$\frac{1}{2}(1 - \|H_0 - H_1\|_{\mathrm{TV}})$。在高斯加噪模型下，Pinsker 不等式结合引理1 可推出 $\|H_0 - H_1\|_{\mathrm{TV}}$ 会随着 $\bar{\alpha}$ 而减小，前提是水印扰动在 ℓ_2中较小。这从互信息论证的角度提供了一种补充性视角，也与（Zhao 等人，2024）。

B. 扩展实验表格

B.1. 编辑器特定敏感性

表8给出了一个扩展的（假设性）视角，其中纳入了 DragDiffusion（Shi等人，2024）作为基于 UNet 的基线，以及 DragonDiffusion（Mou等人，2024）作为基于特征对应的编辑器。整体趋势一致：更强的编辑先验和更激进的噪声水平会导致水印恢复率降低。

B.2. 消融：条件类型与局部性

局部编辑（使用掩码或局部拖拽）可能会保留未编辑区域中的水印，但当解码器预期信号全局分布时，仍可能破坏全局解码。表9 通过假设性地比较整幅图像编辑与掩码编辑，说明了这一影响。我们报告整幅图像上的比特准确率，并出于诊断目的，报告由区域感知解码器计算得出的“区域 BA”。

C. 附加讨论：对手模型与评估规范

C.1. 意外移除与自适应移除

本文的一个核心主题是，水印丢失可能发生而无需明确意图规避来源溯源。这一区分了“意外移除”与（Zhao 等人，2024；Ni 等人，2025a；Fu 等人，2025a）中研究的自适应攻击，以及（Ding 等人，2024；Shamshad 等人，2025）等压力测试。在实践中，编辑器的调优目标是感知上的合理性和用户满意度，而非水印保留。因此，即使用户从未使用“移除”工具，常见的编辑操作——例如对象插入（SHINE）、协调化处理（TF-ICON）或交互式拖拽（DragFlow）——也可能跨过鲁棒性阈值。

对于安全评估，这意味着防御方应在两种威胁模型下衡量鲁棒性：(i) 良性编辑，即编辑器使用默认设置运行，以完成一项面向用户的任务；以及 (ii)自适应编辑，即选择参数和条件，在满足保真度约束的前提下，最大限度地考验水印恢复能力。第一种模型决定实际部署风险；第二种模型则决定最坏情况安全性主张。

C.2. 公平比较原则

水印评估中一个反复出现的误区，是将含水印输出与原始干净输入进行比较，这可能会混淆编辑引起的变化与水印引起的伪影。因此，我们建议采用算法 1 中的配对评估，其中 y_{wm} 与 y_{clean} 在以下条件下进行比较

表 *8.* 其他拖拽编辑器的假设BA（%）（低/中/高强度）。

方法	DragDiffusion（Shi等人，2024）			DragonDiffusion（Mou等人，2024）		
	Low	Med	High	Low	Med	High
StegaStamp	84.1%	62.8%	52.3%	81.5%	60.7%	51.8%
TrustMark	87.6%	66.9%	53.5%	85.2%	65.1%	52.9%
VINE	91.8%	73.4%	56.2%	90.1%	71.6%	55.3%

表*9.* SHINE 的假设性局部性效应：掩码插入下的全图 BA 与区域 BA（%）。

方法	强度	全图 BA	区域 BA
TrustMark	Low	82.9%	88.5%
TrustMark	Med	59.2%	71.0%
TrustMark	High	51.1%	60.2%
VINE	Low	88.7%	92.4%
VINE	Med	67.5%	78.3%
VINE	High	53.6%	62.8%

对齐随机性。当编辑器具有随机性时，未能对齐随机种子会大幅高估“水印伪影”，因为即使没有水印，两次随机编辑的结果也会有所不同。这对于采用随机采样的扩散流水线尤其重要（Ho 等人，2020；Song 等人，2020）。

另一个问题是分辨率不匹配。一些水印是在固定分辨率下训练的（例如，256×256），随后通过调整大小或平铺应用（Bui 等人，2025）。扩散编辑器可能会进行内部调整大小，或在混合分辨率下运行（例如，使用高分辨率背景进行掩码插入），这可能会平滑地破坏水印同步。因此，鲁棒性声明应明确说明完整的预处理流程，以及如何处理尺度变化。

D. 其他背景

随着深度学习和现代生成建模的发展，研究已在预测、感知和视觉生成领域迅速扩展，同时也引发了关于可控性和负责任部署的新问题。时间序列预测的进展得益于更强的基准测试、改进的架构以及更全面的评估协议，这些因素使模型比较更加可靠且信息更丰富（Qiu 等人，2024；2025d;c;e;b；Wu 等人，2025年e；Liu 等人，2025年；Qiu 等人，2025a；Wu 等人，2025年f）。与此同时，以效率为导向的研究推动了三维感知流水线中的训练后量化和实用压缩技术，旨在降低内存占用和延迟，同时不牺牲检测质量（Zhou 等人，2025a；JiangYong Yu & Yuan.，2025年；Zhou 等人，2024；2025b）。在生成方面，越来越多的研究探讨了多样化约束下的可扩展合成与优化策略，从而提升生成系统的灵活性和可控性（Xie 等人，2025；2026a；Xie，2026；Xie 等人，2026b）。此外，生成与表征学习领域也取得了互补进展，进一步拓展了构建高容量模型和训练目标的工具箱（Gong 等人，2024d；2022；2021；2025；2024b；Lin 等人，2024；Gong 等人，2024a；c）。针对领域特定的时间预测，研究者探索了层次化设计和适应策略，以提升模型在分布偏移和复杂现实动态下的鲁棒性（Sun 等人，2025d；c；b；2022；2021；Niu 等人，2025；Sun 等人，2025a；Kudrat 等人，2025）。与此同时，表征编码与匹配方面的进展带来了更强的对齐和对应机制，有助于细粒度检索和基于相似性的推理（Li 等人，2025b；c；Chen 等人，2025c；d；Fu 等人，2025b；Huang 等人，2025）。更强的视觉建模策略进一步提升了特征质量和可迁移性，使模型能够在多样化场景中实现更鲁棒的下游理解（Yu 等人，2025）。在跟踪和序列视觉理解方面，研究者探索了在线学习和解耦形式，以提升动态场景中的时间一致性和鲁棒性（Zheng 等人，2025b；2024；2025a；2023；2022）。低层视觉也在朝着高保真复原和增强方向发展，涵盖超分辨率、亮度/质量控制、轻量化设计和实用评估设置，并日益融合强大的生成先验（Xu 等人，2025b；Fang 等人，2026；Wu 等人，2025b；Li 等人，2023；Ren 等人，2024a；Wang 等人，2025b；Peng 等人，2020；Wang 等人，2023b；Peng 等人，2024b；d；Wang 等人，2023a；Peng 等人，2021；Ren 等人，2024b；Yan 等人，2025；Peng 等人，2024a；Conde 等人，2024；Peng 等人，2025a；c；b；He 等人，2024d；Di 等人，2025；Peng 等人，2024c；He 等人，2024c；e；Pan 等人，2025；Wu 等人，2025c；Jiang 等人，2024；Ignatov 等人，2025；Du 等人，2024；Jin 等人，2024；Sun 等人，2024；Qi 等人，2025；Feng 等人，2025；Xia 等人，2024；Peng 等人；Sun 等人；Yakovenko 等人，2025；Xu 等人，2025a；Wu 等人，2025a；Zhang 等人，2025）。除通用合成之外，参考条件和主体条件生成强调可控性与身份一致性，从而能够生成更贴合用户意图的输出（Qu 等人，2025b；a）。针对不利条件下的鲁棒视觉建模，研究者也进行了积极探索，以应对复杂退化并提升模型在严苛现实环境中的稳定性（Wu 等人，2024b；2023；2024a；2025d）。序列建模

以及以场景为中心的基准测试进一步支持了复杂动态环境下的真实评估和方法开发（Lyu 等人，2025；Chen & Greer，2025）。与此同时，研究者探索了以扩散为核心及基于展开的框架，用于分割和复原，为退化建模和提升生成质量提供了有理论依据的方法（He 等人，2025b；d；h；c；2026；2025f；e；g；i；2024a；2025a；Xiao 等人，2024；He 等人，2024b；2023a；c；b）。

近期多模态大语言模型（MLLM）的进展，越来越多地由这样一个目标所驱动：在提升现实应用中的可靠性、安全性和可控性的同时，让适配更加高效。在效率方面，研究者探索了模态感知参数高效微调，以重新平衡视觉—语言贡献，并以大幅减少的可训练参数实现强大的指令微调（Bi等人，2025a）。为更好地理解和审计模型推理，研究者提出了理论框架，用于建模和评估思维链式推理动态及其对可信推理的影响（Bi等人，2025c）。研究者还通过无需训练的内在选择机制来处理数据质量和筛选问题，剔除价值较低的多模态样本，从而提升下游训练效率和鲁棒性（Bi等人，2025b）。在推理阶段，研究者提出了可控解码策略，通过引导注意力和对比信号聚焦于有依据的视觉证据，减少幻觉（Wang等人，2025a）。除性能之外，可信部署还需要防御和验证：研究者开发了审计框架，用于评估机器反学习是否真正移除了目标知识（Chen等人，2025b）；另有研究提出微调阶段的防御方法，无需依赖外部引导，即可清除MLLM适配中的后门（Rong等人，2025）。与此同时，多视角智能体辩论有助于检测有害内容（Lu等人，2025a），对时效性多模态知识的探测与更新（Jiang等人，2025b），以及强化知识注入的知识导向增强和约束（Jiang等人，2025a），推动了多模态安全和知识可靠性的发展。这些工作还与视频—语言事件理解的重新研究（Zhang等人，2023）、强化中期训练等新型训练范式（Tian等人，2025），以及异构联邦环境下的个性化生成建模（Chen等人，2025a）相辅相成，共同体现出一种转变：研究重点正从单纯扩大规模，转向高效、有依据且可验证可信的多模态系统。

近期研究推动了教育、人机界面和多模态感知领域的学习与交互系统发展。在知识追踪中，由概念图引导的对比式跨课程迁移，为在相关课程体系之间共享知识，并在稀疏监督下改进学生建模，提供了一种有原则的方法

表*10.*TF-ICON（中）下不同输出分辨率对应的假设BA（%）。

方法	256×256	512×512	1024×1024
StegaStamp	62.4%	58.1%	55.6%
TrustMark	65.9%	60.5%	57.2%
VINE	72.8%	69.1%	66.0%

（Han 等人，2025a）。与此同时，基础 GUI 智能体正逐渐发展出更强的感知能力和长时程规划能力，从而能够稳健地与复杂界面及多步骤任务进行交互（Zeng 等人，2025年）。将这一方向延伸至更加自然的人类输入，语音指令图形用户界面智能体旨在直接根据语音命令执行图形用户界面操作，朝着在免手操作或无障碍环境中提供自动化辅助的方向发展（Han 等人，2025c）。除界面智能体之外，研究人员还探索了参考引导的身份保持，以便在人脸视频修复中更好地维持主体一致性，从而在修复退化视频时提升时间连贯性和视觉保真度（Han 等人，2025b）。最后，强调具身情绪的大规模自我中心视角数据集，为从第一人称视角研究情感线索提供了宝贵的监督，并支持更加以人为中心的多模态理解（Feng 等人，2024）。

E. 附加表格：鲁棒性与分辨率

表 10 报告了一项假设性的分辨率缩放研究，强调扩散编辑与分辨率变化以非平凡的方式相互作用。尽管某些水印方案纳入了显式的分辨率处理（Bui 等人，2023年；2025年），潜变量扩散编辑器仍能在新尺度上有效地重新采样细节，即使水印在局部得以保留，也可能破坏同步。

F. 开放问题

尽管进展迅速，我们将生成式编辑下的鲁棒水印视为一个尚未解决的系统问题，它涵盖建模、学习与策略。我们强调若干直接源于本文分析的开放方向。

能够经受语义再合成的水印。如果扩散编辑生成的输出仅受高层语义约束，那么任何与底层像素绑定的水印都可能被视为干扰性变化。一个自然的问题是，能否设计出满足以下条件的水印：（i）不可感知；（ii）可从编辑后的图像中恢复；（iii）对一类明确定义的语义编辑保持不变。这与针对不变性的表征学习相似，但

还涉及密钥和信息载荷，带来了不同于标准嵌入式学习的新挑战。

水印感知扩散编辑。 还可以修改编辑器，使其明确保留水印信道，例如添加约束，确保去噪过程中潜空间签名保持稳定。然而，这可能导致不理想的行为：水印保留约束可能与编辑器投影到自然图像流形的目标相冲突，从而产生伪影或使编辑结果出现偏差。由此产生的优化类似于扩散模型中的多目标轨迹控制，而“内容编辑”与“信号保留”之间的张力值得系统研究。

“水印可靠性”的审计与策略。 当水印缺失可能由良性编辑造成时，基于水印存在与否做出的策略决策便不再简单。例如，平台可能将来源溯源信息缺失视为可疑情况，但这可能会惩罚合法编辑。更细致的做法是将水印检测与扩散编辑检测器结合，并报告校准置信度，而非给出二元判断。如何设计此类校准报告，并在分布偏移和对抗条件下验证其有效性，仍鲜有研究。

反映真实编辑工作流的基准测试。 现有基准测试只能覆盖特定编辑类型，但真实工作流包含多次编辑、迭代优化以及分辨率变化。此外，编辑模型演进迅速，鲁棒性结论可能很快过时。持续的基准测试与压力测试，包括“Erasing the Invisible”等竞赛（Ding等人，2024），为持续评估提供了一种模式，但将其转化为标准化的学术实践仍是一项挑战。

与概念擦除和安全调优的相互作用。 最后，概念擦除方法会改变去噪轨迹和注意力通路（Lu等人， 2024年； Li等人， 2025a；Gao等人， 2024）。理解此类安全调优如何与水印嵌入及检测相互作用，之所以重要有两个原因：安全干预可能无意中破坏溯源机制，而来源溯源感知约束也可能干扰安全性。这要求建立联合评估设置，将安全性、可编辑性与来源溯源一并评估，而非孤立考察。

参考文献

Basu, S., Saberi, M., Bhardwaj, S., Chegini, A. M., Massiceti, D., Sanjabi, M., Hu, S. X. 和 Feizi, S. Editval：扩散式文本引导图像编辑方法基准测试。*arXiv* 预印本 *arXiv:2310.02426*，2023年。

Bi, J., Wang, Y., Chen, H., Xiao, X., Hecker, A., Tresp, V., and Ma, Y. LLaVA steering：通过模态线性表征引导，以少 500 倍的参数进行视觉指令调优。载于 Che, W., Nabende, J., Shutova, E., and Pilehvar, M. T. (eds.), 第 *63* 届计算语言学协会年会论文集（第 *1* 卷：长篇论文），第15230–15250 页，奥地利维也纳，2025年7月。计算语言学协会。ISBN 979-8-89176-251-0。网址 https://aclanthology.org/2025.acl-long.739/。

Bi, J., Wang, Y., Yan, D., Aniri, Huang, W., Jin, Z., Ma, X., Hecker, A., Ye, M., Xiao, X., Schuetze, H., Tresp, V. 和 Ma, Y. Prism：用于免训练多模态数据选择的自剪枝内在选择方法，2025b。网址https://arxiv.org/abs/2502.12119。

Bi, J., Yan, D., Wang, Y., Huang, W., Chen, H., Wan, G., Ye, M., Xiao, X., Schuetze, H., Tresp, V., 等。Cot-kinetics：一种评估LRM推理过程的理论建模。*arXiv*预印本 *arXiv:2505.13408*，2025c。Brooks, T., Holynski, A., 和 Efros, A. A. Instructpix2pix：学习遵循图像编辑指令。载于 计算机视觉与模式识别会议论文集（*CVPR*），2023。Bui, T., Agarwal, S., 和 Collomosse, J. Trustmark：适用于任意分辨率图像的通用水印。 *arXiv*预印本 *arXiv:2311.18297*，2023。Bui, T., Agarwal, S., 和 Collomosse, J. Trustmark：适用于任意分辨率图像的鲁棒水印与水印移除。载于 *IEEE/CVF*国际计算机视觉会议论文集（*ICCV*），2025。Chen, H., Li, H., Zhang, Y., Bi, J., Zhang, G., Zhang, Y., Torr, P., Gu, J., Krompass, D., 和 Tresp, V. Fedbip：基于个性化潜在扩散模型的异构单轮联邦学习。载于计算机视觉与模式识别会议论文集（*CVPR*），第30440–30450页，2025年6月。Chen, H., Zhang, Y., Bi, Y., Zhang, Y., Liu, T., Bi, J., Lan, J., Gu, J., Grosser, C., Krompass, D., 等。机器反学习是否真的移除了模型知识？一个用于审计大语言模型反学习的框架。*arXiv*预印本 *arXiv:2505.23270*，2025b。Chen, Y. 和 Greer, R. 关于Argoverse2场景挖掘挑战的技术报告：迭代误差校正与空间感知提示，2025。网址 https://arxiv.org/abs/2506.11124。Chen, Z., Hu, Y., Li, Z., Fu, Z., Song, X., 和 Nie, L. Offset：基于分割的焦点转移修正，用于组合图像检索。载于 *ACM*国际多媒体会议论文集，第6113–6122页，2025c。

Chen, Z., Hu, Y., Li, Z., Fu, Z., Wen, H., 和 Guan, W. Hud：用于组合视频检索的层次化不确定性感知消歧网络。载于*ACM*国际多媒体会议论文集，第6143–6152页，2025d。

Conde, M. V., Lei, Z., Li, W., Katsavounidis, I., Timofte, R., Yan, M., Liu, X., Wang, Q., Ye, X., Du, Z., 等。压缩AVIF图像的实时4K超分辨率。AIS 2024挑战赛综述。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第5838–5856页，2024年。

Cover, T. M. 和 Thomas, J. A. 信息论基础。Wiley，第2版，2006年。

Di, X., Peng, L., Xia, P., Li, W., Pei, R., Cao, Y., Wang, Y., 和 Zha, Z.-J. QMambaBSR：基于查询状态空间模型的突发图像超分辨率。载于计算机视觉与模式识别会议论文集，第23080–23090页，2025年。

Ding, M., Rabbani, T., An, B., Chakraborty, S., Deng, C., Saberi, M., Wen, Y., Zhao, X., Zhou, M., Satheesh, A., Li, L., Wang, Y.-X., Patel, V. M., Feizi, S., Goldstein, T. 和 Huang, F. NeurIPS 2024竞赛：Erasing the Invisible：图像水印压力测试挑战赛。收录于*NeurIPS*竞赛专题，2024。

Du, Z., Peng, L., Wang, Y., Cao, Y. 和 Zha, Z.-J. FC3DNet：一种用于高效去摩尔′纹的全连接编码器-解码器。收录于*2024*年*IEEE*国际图像处理会议（*ICIP*），第1642–1648页。IEEE，2024。

Fang, S., Peng, L., Wang, Y., Wei, R. 和 Wang, Y. 深度协同Mamba与记忆专家相结合，用于全天候图像反射分离。*arXiv*预印本 *arXiv:2601.00322*，2026。

Feng, Y., Han, W., Jin, T., Zhao, Z., Wu, F., Yao, C., Chen, J., 等. 通过大规模自我中心视频数据集探索具身情绪。神经信息处理系统进展，37:118182–118197，2024。

Feng, Z., Peng, L., Di, X., Guo, Y., Li, W., Zhang, Y., Pei, R., Wang, Y., Cao, Y., andZha, Z.-J. Pmq-ve：用于视频增强的渐进式多帧量化。*arXiv* 预印本 *arXiv:2505.12266*，2025年。

Fu, W., Carter, F., Wang, Y., Davis, E., and Zhang, B. 基于扩散的图像编辑：对鲁棒不可见水印的意外威胁。*arXiv* 预印本 *arXiv:2511.05598*，2025a。

Fu, Z., Li, Z., Chen, Z., Wang, C., Song, X., Hu, Y. 和 Nie, L. Pair：面向组合图像检索的互补性引导的解耦。载于*IEEE*国际声学、语音与信号处理会议论文集，第1–5 页。IEEE，2025b。

Gao, D., Lu, S., Walters, S., Zhou, W., Chu, J., Zhang, J., Zhang, B., Jia, M., Zhao, J., Fan, Z. 和 Zhang, W. Eraseanything：在整流流变换器中实现概念擦除。*arXiv*预印本 *arXiv:2412.20413*，2024。

Gao, D., Jiang, N., Zhang, A., Lu, S., Tang, Y., Zhou, W., Zhang, W., 和 Fan, Z. Revoking amnesia：基于强化学习的轨迹优化，用于在扩散模型中恢复已擦除的概念，2025年。

Gong, Y., Huang, L., 和 Chen, L. 以偏差消除偏差：一种用于数据增强的通用多模态数据学习方法。*arXiv*预印本 *arXiv:2101.08533*，2021年。

Gong, Y., Huang, L., 和 Chen, L. 基于颜色攻击和联合防御的行人重识别方法。载于*CVPR，2022*，第 4313–4322 页，2022年。

Gong, Y., Hou, Y., Wang, Z., Lin, Z., 和 Jiang, M. 面向稀疏数据神经偏微分方程求解器的对抗学习。*arXiv* 预印本 *arXiv:2409.02431*，2024年。

Gong, Y., Hou, Y., Zhang, C., 和 Jiang, M. 超越数据增强：提升模型在极端采集环境下的鲁棒性。载于*2024*年国际联合神经网络会议（*IJCNN*），第 1–8 页。IEEE，2024b。

Gong, Y., Li, J., Chen, L., 和 Jiang, M. 通过图像级集成学习探索颜色不变性。*arXiv* 预印本 *arXiv:2401.10512*，2024c。

Gong, Y., Zhong, Z., Qu, Y., Luo, Z., Ji, R., 和 Jiang, M. 用于行人重识别的跨模态扰动协同攻击。神经信息处理系统进展，37:23352–23377，2024d。

Gong, Y., Lan, S., Yang, C., Xu, K., 和 Jiang, M. Strusr：结合物理信息泰勒引导的结构感知符号回归。*arXiv* 预印本 *arXiv:2510.06635*，2025年。

Gunn, S., Zhao, X., 和 Song, D. 一种适用于生成式图像模型的不可检测水印。*arXiv* 预印本 *arXiv:2410.07369*，2024。

Han, W., Lin, W., Hu, L., Dai, Z., Zhou, Y., Li, M., Liu, Z., Yao, C., 和 Chen, J. 通过 ConceptGraph 引导的知识迁移进行对比式跨课程知识追踪。*arXiv* 预印本 *arXiv:2505.13489*，2025a。

Han, W., Lin, W., Zhou, Y., Liu, Q., Wang, S., Yao, C. 和 Chen, J. 展示与润饰：人脸视频修复中的参考引导的身份保持。*arXiv* 预印本 *arXiv:2507.10293*，2025b。

Han, W., Zeng, Z., Huang, J., Jiang, S., Zheng, L., Yang, L., Qiu, H., Yao, C., Chen, J. 和 Ma, L. Uitron-speech：迈向基于语音指令的自动化 GUI 智能体。*arXiv* 预印本 *arXiv:2506.11127*，2025c。

He, C., Li, K., Xu, G., Yan, J., Tang, L., Zhang, Y., Wang, Y., and Li, X. Hqg-net：采用高质量引导的无配对医学图像增强。*TNNLS*，2023a。

He, C., Li, K., Xu, G., Zhang, Y., Hu, R., Guo, Z., and Li, X. 用于异构图像融合的抗退化展开网络。发表于*ICCV*，第12611–12621页，2023b。

He, C., Li, K., Zhang, Y., Tang, L., Zhang, Y., Guo, Z., and Li, X. 结合特征分解与边缘重建的伪装目标检测。发表于*CVPR*，第 22046–22055 页，2023年。He, C., Li, K., Zhang, Y., Xu, G., Tang, L., Zhang, Y., Guo, Z., and Li, X. 采用基于 SAM 的伪标签生成和多尺度特征分组的弱监督隐蔽视觉感知分割。*NeurIPS*，36，2024年。He, C., Li, K., Zhang, Y., Zhang, Y., Guo, Z., Li, X., Danell- jan, M., and Yu, F. 策略性猎物造就敏锐捕食者：通过生成伪装目标增强伪装目标检测器。*ICLR*，2024b。He, C., Fang, C., Zhang, Y., Ye, T., Li, K., Tang, L., Guo, Z., Li, X., and Farsiu, S. Reti-Diff：采用基于 Retinex 的潜在扩散模型进行光照退化图像复原。*ICLR*，2025a。He, C., Li, K., Zhang, Y., Yang, Z., Tang, L., Zhang, Y., Kong, L., and Farsiu, S. 以不完整监督分割隐蔽物体。*TPAMI*，2025b。He, C., Shen, Y., Fang, C., Xiao, F., Tang, L., Zhang, Y., Zuo, W., Guo, Z., and Li, X. 低层视觉中的扩散模型：综述。*TPAMI*，2025c。He, C., Xiao, F., Zhang, R., Fang, C., Fan, D.-P., and Farsiu, S. 用于隐蔽视觉感知并结合生成式细化的可逆展开网络。*arXiv preprintarXiv:2508.15027*，2025d。He, C., Zhang, R., Chen, Z., Yang, B., Fang, C., Lin, Y., Xiao, F., and Farsiu, S. UnfoldLDM：基于潜在扩散先验的深度展开式盲图像复原。*arXivpreprint arXiv:2511.18152*，2025e。He, C., Zhang, R., Tang, L., Yang, Z., Li, K., Fan, D.-P., and Farsiu, S. Scaler：用于标签缺乏的隐蔽目标分割的 SAM 增强协同学习。*arXivpreprintarXiv:2511.18136*，2025f。

He, C., Zhang, R., Xiao, F., Fang, C., Tang, L., Zhang, Y., and Farsiu, S. Unfoldir：重新思考用于光照退化图像复原的深度展开网络。*arXivpreprint arXiv:2505.06683*，2025g。

He, C., Zhang, R., Xiao, F., Fang, C., Tang, L., Zhang, Y., Kong, L., Fan, D.-P., Li, K., and Farsiu, S. Run：用于隐蔽物体分割的可逆展开网络。国际机器学习大会，2025h。

He, C., Zhang, R., Zhang, D., Xiao, F., Fan, D.-P., and Farsiu, S. 用于真实场景隐蔽物体分割的嵌套展开网络。*arXiv preprintarXiv:2511.18164*，2025i。

He, C., Zhang, R., Xiao, F., Zhang, D., Cao, Z., and Far- siu, S. 通过课程选择和反课程提升优化上下文纠缠内容分割。*arXiv*预印本 *arXiv:2602.01183*，2026年。He, Y., Jiang, A., Jiang, L., Peng, L., Wang, Z., and Wang, L. 通过空间-频率交互实现的双路径耦合图像去雨网络。发表于*2024 IEEE*国际图像处理会议（*ICIP*），第1452–1458页。IEEE，2024c。He, Y., Peng, L., Wang, L., and Cheng, J. 单幅图像去雨的潜在退化表征约束。发表于*ICASSP 2024-2024 IEEE*国际声学、语音与信号处理会议（*ICASSP*），第3155–3159页。IEEE，2024d。He, Y., Peng, L., Yi, Q., Wu, C., and Wang, L. 利用状态空间模型进行图像复原的多尺度表征学习。*arXiv*预印本 *arXiv:2408.10145*，2024e。Ho, J. and Salimans, T. 无分类器扩散引导。*arXiv*预印本 *arXiv:2207.12598*，2022。Ho, J., Jain, A., and Abbeel, P. 去噪扩散概率模型。发表于神经信息处理系统进展（*NeurIPS*），2020年。Hor´e, A. and Ziou, D. 图像质量指标：峰值信噪比与结构相似性指数比较。发表于国际模式识别会议（*ICPR*），2010年。Huang, Q., Chen, Z., Li, Z., Wang, C., Song, X., Hu, Y., and Nie, L. MEDIAN：用于组合图像检索的自适应中间粒度聚合网络。发表于*IEEE*国际声学、语音与信号处理会议论文集，第1–5页。IEEE，2025年。Huang, Y., Huang, J., Liu, Y., Yan, M., Lv, J., Liu, J., Xiong, W., Zhang, H., Chen, S., and Cao, L. 基于扩散模型的图像编辑：综述。*arXiv*预印本 *arXiv:2402.17525*，2024年。

Ignatov, A., Perevozchikov, G., Timofte, R., Pan, W., Wang, S., Zhang, D., Ran, Z., Li, X., Ju, S., Zhang, D., et al. 移动GPU上的RGB照片增强，Mobile AI 2025挑战赛：报告。发表于计算机视觉与模式识别会议论文集，第1922–1933页，2025年。

Jiang, A., Wei, Z., Peng, L., Liu, F., Li, W., and Wang, M. Dalpsr：利用与退化对齐的语言提示实现真实世界图像超分辨率。*arXiv* 预印本 *arXiv:2406.16477*，2024。

Jiang, K., Jiang, H., Jiang, N., Gao, Z., Bi, J., Ren, Y., Li, B., Du, Y., Liu, L., and Li, Q. Kore：通过知识导向的增强和约束，为大型多模态模型增强知识注入，2025a。网址https://arxiv.org/abs/2510.19316。

Jiang, K., Jiang, N., Du, Y., Ren, Y., Li, Y., Gao, Y., Bi, J., Ma, Y., Liu, Q., Wang, X., Jia, Y., Jiang, H., Hu, Y., Li, B., and Liu, L. Mined：通过多模态时效性知识探索与更新大型多模态模型，2025b。网址 https://arxiv.org/abs/2510.19457。

JiangYong Yu, Sifan Zhou, D. Y. S. L. S. W. X. H. C. X. Z. X. C. S. 和 Yuan., Z. Mquant：通过静态量化释放多模态大语言模型的推理潜力。载于第*33*届*ACM*国际多媒体会议论文集，2025年。

Jin, X., Guo, C., Li, X., Yue, Z., Li, C., Zhou, S., Feng, R., Dai, Y., Yang, P., Loy, C. C. 等。MIPI 2024挑战赛：少样本原始图像去噪：方法与结果。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第1153–1161页，2024。

Kawar, B., Zada, S., Lang, O., Tov, O., Chang, H., Dekel, T., Mosseri, I. 和 Irani, M. Imagic：使用扩散模型进行基于文本的真实图像编辑。载于*IEEE/CVF*计算机视觉与模式识别会议论文集（*CVPR*），2023年。

Kudrat, D., Xie, Z., Sun, Y., Jia, T., 和 Hu, Q. 用于时间序列预测的分块结构损失。载于第四十二届国际机器学习会议，2025年。

Li, L., Lu, S., Ren, Y., 和 Kong, A. W.-K. 让你回归正轨：自动引导去噪轨迹，避开不想要的概念。*arXiv* 预印本 *arXiv:2504.12782*，2025a。

Li, Y., Zhang, Y., Timofte, R., Van Gool, L., Yu, L., Li, Y., Li, X., Jiang, T., Wu, Q., Han, M. 等。NTIRE 2023高效超分辨率挑战：方法与结果。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第1922–1960页，2023年。

Li, Z., Chen, Z., Wen, H., Fu, Z., Hu, Y. 与 Guan, W.。编码器：面向组合图像检索的实体挖掘与修改关系绑定。载于人工智能促进协会人工智能会议论文集，第39卷，第5101–5109页，2025b。

Li, Z., Fu, Z., Hu, Y., Chen, Z., Wen, H., 和 Nie, L. FineCIR：针对组合图像检索中细粒度修改语义的显式解析。*https://arxiv.org/abs/2503.21309*，2025c。

Lin, J., Zhenzhong, W., Dejun, X., Shu, J., Gong, Y., 和 Jiang, M. Phys4dgen：一个物理驱动框架，可从单张图像生成可控且高效的4D内容。*arXiv* 预印本 *arXiv:2411.16800*，2024。

Liu, X., Qiu, X., Wu, X., Li, Z., Guo, C., Hu, J.和Yang, B. 重新思考不规则时间序列预测：一种简单而有效的基线。*arXiv*预印本 *arXiv:2505.11250*，2025年。

Lu, R., Bi, J., Ma, Y., Xiao, F., Du, Y.和Tian, Y. Mv-debate：面向社交媒体多模态有害内容检测的多视角智能体辩论与动态反思门控，2025a。URL https://arxiv.org/abs/2508.05557。

Lu, S., Liu, Y.和Kong, A. W.-K. TF-ICON：基于扩散的免训练跨域图像合成。载于*IEEE/CVF*国际计算机视觉会议（*ICCV*）论文集，2023年。

Lu, S., Wang, Z., Li, L., Liu, Y.和Kong, A. W.-K. MACE：扩散模型中的大规模概念擦除。载于 *IEEE/CVF*计算机视觉与模式识别会议（*CVPR*）论文集，2024年。

Lu, S.、Zhou, Z.、Lu, J.、Zhu, Y. 和 Kong, A. W.-K. 利用生成先验实现抵抗图像编辑的鲁棒水印：从基准测试到发展。*arXiv* 预印本 *arXiv:2410.18775*，2024b。

Lu, S.、Lian, Z.、Zhou, Z.、Zhang, S.、Zhao, C. 和 Kong, A. W.-K. FLUX 是否已经知道如何执行物理上合理的图像合成？*arXiv* 预印本 *arXiv:2509.21278*，2025b。

Luo, X.、Zhan, R.、Chang, H.、Yang, F. 和 Milanfar, P. 与失真无关的深度水印。在*IEEE/CVF*计算机视觉与模式识别会议（*CVPR*）论文集中，2020。

Lyu, J., Zhao, M., Hu, J., Huang, X., Chen, Y., 和 Du, S. Vadmamba：探索用于快速视频异常检测的状态空间模型，2025年。网址https://arxiv.org/abs/2503.21169。

Mokady, R., Hertz, A.,Aberman,K., Pritch,Y.,andCohen- Or, D. 使用引导式扩散模型编辑真实图像的空文本反演。载于*IEEE/CVF*计算机视觉与模式识别会议（*CVPR*）论文集，2023年。

Mou, C., Wang, X., Song, J., Shan, Y., and Zhang, J. Drag- ondiffusion：实现扩散模型上的拖拽式操作。载于国际学习表征会议 *(ICLR)*，2024。

Ni, Y., Carter, F., Niu, Z., Davis, E., 和 Zhang, B. 基于扩散的图像编辑如何破坏鲁棒水印。*arXiv* 预印本 *arXiv:2510.05978*，2025a。Ni, Y., Yang, Z., Niu, Z., Davis, E., 和 Carter, F. 论扩散编辑下鲁棒水印的信息论脆弱性。*arXiv* 预印本 *arXiv:2511.10933*，2025b。Niu, W., Xie, Z., Sun, Y., He, W., Xu, M., 和 Hao, C. Lang-time：一种采用近端策略优化进行时间序列预测的语言引导统一模型。载于第四十二届机器学习国际会议，2025。Pan, J., Liu, Y., He, X., Peng, L., Li, J., Sun, Y., 和 Huang, X. 先增强，再搜索：一种利用基础模型进行跨域少样本目标检测的增强搜索策略。载于计算机视觉与模式识别会议论文集，第1548–1556页，2025。Peng, L., Li, W., Guo, J., Di, X., Sun, H., Li, Y., Pei, R., Wang, Y., Cao, Y., 和 Zha, Z.-J. 利用原始数据提升真实世界超分辨率：一种新视角、数据集和基线。Peng, L., Jiang, A., Yi, Q., 和 Wang, M. 用于单幅图像去雨的累积雨密度感知网络。*IEEE*信号处理快报，27:406–410，2020。Peng, L., Jiang, A., Wei, H., Liu, B., 和 Wang, M. 通过渐进式结构增强约束实现的集成单幅图像去雨网络。信号处理：图像通信，99:116460，2021。Peng, L., Cao, Y., Pei, R., Li, W., Guo, J., Fu, X., Wang, Y., 和 Zha, Z.-J. 通过自适应方向梯度卷积实现高效的真实世界图像超分辨率。*arXiv* 预印本 *arXiv:2405.07023*，2024a。Peng, L., Cao, Y., Sun, Y., 和 Wang, Y. 用于压缩图像分类的轻量级自适应特征去漂移。*IEEE*多媒体汇刊，26:6424–6436，2024b。

Peng, L., Li, W., Guo, J., Di, X., Sun, H., Li, Y., Pei, R., Wang, Y., Cao, Y., 和 Zha, Z.-J. 揭示隐藏细节：一种面向真实世界超分辨率的原始数据增强范式。*arXiv* 预印本 *arXiv:2411.10798*，2024c。

Peng, L., Li, W., Pei, R., Ren, J., Xu, J., Wang, Y., Cao, Y., and Zha, Z.-J. 迈向真实世界超分辨率的真实数据生成。*arXiv* 预印本 *arXiv:2406.07255*，2024d。

Peng, L., Di, X., Feng, Z., Li, W., Pei, R., Wang, Y., Fu, X., Cao, Y., and Zha, Z.-J. 引导 Mamba 应对复杂纹理：一种用于图像复原的高效纹理感知状态空间模型。*arXiv* 预印本 *arXiv:2501.16583*，2025a。Peng, L., Wang, Y., Di, X., Fu, X., Cao, Y., Zha, Z.-J., et al. 通过中心—周边协同卷积提升图像去雨效果。载于人工智能 *AAAI* 会议论文集，第 39 卷，第 6470–6478 页，2025b。Peng, L., Wu, A., Li, W., Xia, P., Dai, X., Zhang, X., Di, X., Sun, H., Pei, R., Wang, Y., et al. 从像素到高斯：采用二维高斯建模实现超快速连续超分辨率。*arXiv* 预印本 *arXiv:2503.06617*，2025c。Qi, X., Li, R., Peng, L., Ling, Q., Yu, J., Chen, Z., Chang, P., Han, M., and Xiao, J. 利用扩散模型进行无数据知识蒸馏。*arXiv* 预印本 *arXiv:2504.00870*，2025年。Qiu, X., Hu, J., Zhou, L., Wu, X., Du, J., Zhang, B., Guo, C., Zhou, A., Jensen, C. S., Sheng, Z., and Yang, B. TFB：迈向全面、公平的时间序列预测方法基准测试。载于*Proc. VLDBEndow.*，第 2363–2377 页，2024。Qiu, X., Cheng, H., Wu, X., Hu, J., and Guo, C. 多变量时间序列预测深度学习综合综述：通道策略视角。*arXiv* 预印本 *arXiv:2502.10721*，2025a。Qiu, X., Li, Z., Qiu, W., Hu, S., Zhou, L., Wu, X., Li, Z., Guo, C., Zhou, A., Sheng, Z., Hu, J., Jensen, C. S., and Yang, B. Tab：时间序列异常检测方法的统一基准测试。载于*Proc. VLDBEndow.*，2025b。Qiu, X., Wu, X., Cheng, H., Liu, X., Guo, C., Hu, J., and Yang, B. DBloss：用于时间序列预测的基于分解的损失函数。载于*NeurIPS*，2025c。

Qiu, X., Wu, X., Lin, Y., Guo, C., Hu, J. 和 Yang, B. DUET：双重聚类增强的多变量时间序列预测。载于 *SIGKDD*，第 1185–1196 页，2025d。

Qiu, X., Zhu, Y., Li, Z., Cheng, H., Wu, X., Guo, C., Yang, B. 和 Hu, J. Dag：一种用于含外生变量时间序列预测的双重因果网络。*arXiv* 预印本 *arXiv:2509.14933*，2025e。

Qu, Y.、Fu, D.与Fan, J. 面向域偏移条件下新颖性检测的主体信息抽取。*arXiv*预印本 *arXiv:2504.21247*，2025a。

Qu, Y.、Panariello, M.、Todisco, M.与Evans, N. 语音中的无参考对抗性性别混淆。*arXiv*预印本 *arXiv:2508.02295*，2025b。

Ren, B.、Li, Y.、Mehta, N.、Timofte, R.、Yu, H.、Wan, C.、Hong, Y.、Han, B.、Wu, Z.、Zou, Y.等。第九届NTIRE 2024高效超分辨率挑战赛报告。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第6595–6631页，2024年。

Ren, J.、Li, W.、Chen, H.、Pei, R.、Shao, B.、Guo, Y.、Peng, L.、Song, F.与Zhu, L. Ultrapixel：将超高分辨率图像合成推向新高峰。神经信息处理系统进展，37：111131–111171，2024b。

Ren, Y., Lu, S., 和 Kong, A. W.-K. 闪闪发光的未必都是金子：3D高斯泼溅中的密钥保护型3D秘密。*arXiv*预印本 *arXiv:2503.07191*，2025年。

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., 和 Ommer, B. 使用潜在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（*CVPR*）论文集，2022。

Rong, X., Huang, W., Liang, J.,Bi, J., Xiao, X., Li, Y., Du, B., 和 Ye, M. 无需外部引导的多模态大语言模型微调后门清除。*arXiv*预印本 *arXiv:2505.16916*，2025年。

Shamshad, F., Bakr, T., Shaaban, Y., Hussein, N., Nan- dakumar, K., and Lukas, N. NeurIPS 2024不可见水印去除挑战赛冠军方案。*arXiv*预印本 *arXiv:2508.21072*，2025年。Shi, Y., Xue, C., Liew, J. H., Pan, J., Yan, H., Zhang, W., Tan, V. Y. F., and Bai, S. DragDiffusion：利用扩散模型实现交互式基于点的图像编辑。载于*IEEE/CVF*计算机视觉与模式识别会议（*CVPR*）论文集，2024。Song, J., Meng, C., and Ermon, S. 去噪扩散隐式模型。*arXiv*预印本 *arXiv:2010.02502*，2020。Sun, H., Li, W., Liu, J., Zhou, K., Chen, Y., Guo, Y., Li, Y., Pei, R., Peng, L., and Yang, Y. 文本提升泛化能力：一种用于真实世界图像复原的即插即用描述器。

Sun, H., Li, W., Liu, J., Zhou, K., Chen, Y., Guo, Y., Li, Y., Pei, R., Peng, L., and Yang, Y. 超越像素：文本增强真实世界图像复原中的泛化能力。*arXiv*预印本 *arXiv:2412.00878*，2024。

Sun, Y., Xie, Z., Chen, Y., Huang, X., 和 Hu, Q. 基于二维注意力机制的太阳风速度预测。《空间天气》，19(7):e2020SW002707，2021年。

Sun, Y., Xie, Z., Chen, Y., and Hu, Q. 利用多模态信息准确预测太阳风速度。《空间：科学与技术》，2022年。

Sun, Y., Eldele, E., Xie, Z., Wang, Y., Niu, W., Hu, Q., Kwoh, C. K., and Wu, M. 通过时间异质性建模和语义对齐，使大语言模型适用于时间序列预测。*arXiv* 预印本 *arXiv:2508.07195*，2025a。

Sun, Y., Xie, Z., Chen, D., Eldele, E., and Hu, Q. 用于时间序列预测的分层分类辅助网络。发表于人工智能促进协会人工智能会议论文集，第39卷，第20743–20751页，2025b。

Sun, Y., Xie, Z., Eldele, E., Chen, D., Hu, Q., and Wu, M. 学习模式专属专家，以应对补丁级分布偏移下的时间序列预测。神经信息处理系统进展，2025c。

Sun, Y., Xie, Z., Xing, H., Yu, H., 和 Hu, Q. PPGF：概率模式引导的时间序列预测。*IEEE*神经网络与学习系统汇刊，2025d。

Tancik, M., Mildenhall, B., 和 Ng, R. StegaStamp：实体照片中的隐形超链接。载于*IEEE/CVF*计算机视觉与模式识别会议（*CVPR*）论文集，2020。

Tian, Y., Chen, S., Xu, Z., Wang, Y., Bi, J., Han, P., 和 Wang, W. 强化中期训练，2025年。网址 https://arxiv.org/abs/2509.24375。

Wang, H., Peng, L., Sun, Y., Wan, Z., Wang, Y., 和 Cao, Y. 面向递归式低照度图像增强的亮度感知。*IEEE*人工智能汇刊，5(6):3034–3045，2023a。

Wang, Y.、Peng, L.、Li, L.、Cao, Y. 和 Zha, Z.-J. 面向图像曝光校正的解耦与聚合。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第 18115–18124 页，2023年b。

Wang, Y., Bi, J., Ma, Y., and Pirk, S. ASCD：用于减少多模态大语言模型幻觉的注意力可控对比解码。*arXiv* 预印本 *arXiv:2506.14766*，2025a。

Wang, Y., Liang, Z., Zhang, F., Tian, L., Wang, L., Li, J., Yang, J., Timofte, R., Guo, Y., Jin, K., 等。NTIRE 2025 光场图像超分辨率挑战赛：方法与结果。载于计算机视觉与模式识别会议论文集，第 1227–1246 页，2025b。

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. 图像质量评估：从误差可见性到结构相似性。*IEEE* 图像处理汇刊，2004。

Wu, A., Peng, L., Di, X., Dai, X., Wu, C., Wang, Y., Fu, X., Cao, Y., and Zha, Z.-J. Robustgs：低质量条件下前馈式 3D 高斯泼溅的统一增强。*arXiv* 预印本 *arXiv:2508.03077*，2025a。Wu, B., Zou, C., Li, C., Huang, D., Yang, F., Tan, H., Peng, J., Wu, J., Xiong, J., Jiang, J., 等。混元视频 1.5 技术报告。*arXiv* 预印本 *arXiv:2511.18870*，2025b。Wu, C., Wang, L., Peng, L., Lu, D., and Zheng, Z. 摒弃高倍率下采样：超高清图像复原的一种新型设计范式。发表于*2025 IEEE/CVF* 冬季计算机视觉应用会议 *(WACV)*，第 2390–2399 页。IEEE，2025c。Wu, H., Yang, Y., Chen, H., Ren, J., and Zhu, L. 用于视频中雨滴与雨线联合去除的掩码引导渐进式网络。发表于第 *31* 届 *ACM*国际多媒体会议论文集，第 7216–7225 页，2023年。Wu, H., Yang, Y., Aviles-Rivero, A. I., Ren, J., Chen, S., Chen, H., and Zhu, L. 通过时间解耦专家和分布驱动的对比正则化实现的半监督视频去雪网络。发表于欧洲计算机视觉会议，第 70–89 页。施普林格，2024年。Wu, H., Yang, Y., Xu, H., Wang, W., Zhou, J., and Zhu, L. Rainmamba：利用状态空间模型增强视频去雨的局部性学习。发表于第 *32* 届 *ACM*国际多媒体会议论文集，第 7881–7890 页，2024b。Wu, H., Wu, Y., Jiang, J., Wu, C., Wang, H., and Zheng, Y. Samvsr：利用语义先验实现面向区域聚焦 Mamba 的视频去雪。发表于第 *33* 届 *ACM*国际多媒体会议论文集，第 7376–7385 页，2025d。Wu, X., Qiu, X., Gao, H., Hu, J., Yang, B., and Guo, C. K^{2}VAE：一种用于概率时间序列预测的 Koopman-Kalman 增强型变分自编码器。发表于国际机器学习大会，2025e。

Wu, X., Qiu, X., Li, Z., Wang, Y., Hu, J., Guo, C., Xiong, H., and Yang, B. CATCH：通过频率分块实现通道感知的多变量时间序列异常检测。发表于*ICLR*，2025f。Xia, P., Peng, L., Di, X., Pei, R., Wang, Y., Cao, Y., and Zha, Z.-J. S3mamba：通过可扩展状态空间模型实现任意尺度超分辨率。*arXiv*预印本 *arXiv:2411.11906*，6，2024。Xiao, F., Hu, S., Shen, Y., Fang, C., Huang, J., He, C., Tang, L., Yang, Z., and Li, X. 伪装目标检测及其拓展综述。*CAAIAIR*，2024。Xie, Z. Conquer：用于基于文本的人物搜索的上下文感知表征与查询增强。*arXiv*预印本 *arXiv:2601.18625*，2026。Xie, Z., Wang, C., Wang, Y., Cai, S., Wang, S., and Jin, T. 面向人物检索的聊天驱动文本生成与交互。发表于*2025*年自然语言处理实证方法会议论文集，第5259–5270页，2025年。Xie, Z., Liu, X., Zhang, B., Lin, Y., Cai, S., and Jin, T. HVD：面向文本-视频检索的人类视觉驱动视频表征学习。*arXiv*预印本 *arXiv:2601.16155*，2026a。Xie, Z., Zhang, B., Lin, Y., and Jin, T. 深入探索：面向鲁棒视频-文本检索的层次化视觉感知。*arXiv*预印本 *arXiv:2601.12768*，2026b。Xu, H., Peng, L., Song, S., Liu, X., Jun, M., Li, S., Yu, J., and Mao, X. Camel：资源受限设备上节能的LLM推理。*arXiv*预印本 *arXiv:2508.09173*，2025a。Xu, J., Li, W., Sun, H., Li, F., Wang, Z., Peng, L., Ren, J., Yang, H., Hu, X., Pei, R., et al. 基于一致性校正流的快速图像超分辨率。发表于*IEEE/CVF*国际计算机视觉会议论文集，第11755–11765页，2025b。Yakovenko, A., Chakvetadze, G., Khrapov, I., Zhelezov, M., Vatolin, D., Timofte, R., Oh, Y., Kwon, J., Park, J., Cho, N. I., et al. AIM 2025低光照RAW视频去噪挑战赛：数据集、方法与结果。*arXiv*预印本 *arXiv:2508.16830*，2025年。Yan, Q., Jiang, A., Chen, K., Peng, L., Yi, Q., and Zhang, C. 文本提示引导的图像复原。《人工智能工程应用》，155：110981，2025年。Yu, X., Chen, Z., Zhang, Y., Lu, S., Shen, R., Zhang, J., Hu, X., Fu, Y., and Yan, S. 视觉文档理解与问答：一种结合测试时扩展的多智能体协作框架。*arXiv*预印本 *arXiv:2508.03404*，2025年。

Zeng, Z., Huang, J., Zheng, L., Han, W., Zhong, Y., Chen, L., Yang, L., Chu, Y., He, Y., and Ma, L. Uitron：具备先进感知与规划能力的基础GUI智能体。*arXiv*预印本 *arXiv:2508.21767*，2025年。Zhang, G., Bi, J., Gu, J., Chen, Y., and Tresp, V. Spot!：重新审视用于事件理解的视频语言模型。*arXiv*预印本 *arXiv:2311.12919*，2023年。Zhang, L., Liu, X., Martin, A. V., Bearfield, C. X., Brun, Y., and Guan, H. 利用稳定扩散实现抗攻击的图像水印。*arXiv*预印本 *arXiv:2401.04247*，2024。Zhang, R., Isola, P., Efros, A. A., Shechtman, E., and Wang, O. 深层特征作为感知度量的惊人效力。发表于 *IEEE/CVF*计算机视觉与模式识别会议论文集（*CVPR*），2018年。Zhang, S., Guo, Y., Peng, L., Wang, Z., Chen, Y., Li, W., Zhang, X., Zhang, Y., and Chen, J. VividFace：用于视频人脸增强的高质量、高效率单步扩散。*arXiv*预印本 *arXiv:2509.23584*，2025年。Zhao, X., Zhang, K., Su, Z., Vasan, S., Grishchenko, I., Kruegel, C., Vigna, G., Wang, Y.-X., and Li, L. 生成式AI可被证明能够移除不可见图像水印。发表于神经信息处理系统进展（*NeurIPS*），2024年。Zheng, Y., Zhong, B., Liang, Q., Tang, Z., Ji, R., and Li, X. 利用并行交互网络，通过全局线索与局部线索实现视觉跟踪。*IEEE* 视频技术电路与系统汇刊，33(4)：1671–1683，2022年。Zheng, Y., Zhong, B., Liang, Q., Li, G., Ji, R., and Li, X. 面向视觉语言跟踪的统一令牌学习。*IEEE* 视频技术电路与系统汇刊，34(4)：2125–2135，2023年。Zheng, Y., Zhong, B., Liang, Q., Mo, Z., Zhang, S., and Li, X. OdTrack：用于视觉跟踪的在线密集时序令牌学习。发表于人工智能促进协会人工智能会议论文集，第38卷，第7588–7596页，2024年。Zheng, Y., Zhong, B., Liang, Q., Li, N., and Song, S. 面向自监督跟踪的解耦时空一致性学习。发表于人工智能促进协会人工智能会议论文集，第39卷，第10635–10643页，2025a。Zheng, Y., Zhong, B., Liang, Q., Zhang, S., Li, G., Li, X., and Ji, R. 面向通用模态跟踪的在线密集时序令牌学习。*IEEE* 模式分析与机器智能汇刊，2025b。

Zhou, S.、Li, L.、Zhang, X.、Zhang, B.、Bai, S.、Sun, M.、Zhao, Z.、Lu, X.与Chu, X. LiDAR-PTQ：面向点云3D目标检测的训练后量化。2024。

Zhou, S.、Wang, S.、Yuan, Z.、Shi, M.、Shang, Y.与Yang, D. GSQ-tuning：面向大语言模型设备端微调的全量化训练中的组共享指数整数。载于《计算语言学协会发现：*ACL 2025*》，第22971–22988页，奥地利维也纳，2025a年7月。计算语言学协会。ISBN 979-8-89176-256-5。

Zhou, S.、Yuan, Z.、Yang, D.、Hu, X.、Qian, J.与Zhao, Z. Pillarhist：基于高度感知直方图的量化感知柱状特征编码器。载于 《计算机视觉与模式识别会议论文集》，第27336–27345页，2025b。

Zhou, Z., Lu, S., Leng, S., Zhang, S., Lian, Z., Yu, X., 和 Kong, A. W.-K. Dragflow：利用基于区域的监督释放 DiT先验，以进行拖拽编辑。*arXiv* 预印本 *arXiv:2510.02253*，2025c。

Zhu, J., Kaplan, R., Johnson, J., 和 Fei-Fei, L. HiDDeN：利用深度网络隐藏数据。发表于欧洲计算机视觉会议（*ECCV*）论文集，2018。