SemGuard：面向高斯泼溅的语义解耦水印

Yingying Shi、Zhong Zhou、Bin Zhou

arXiv:2611.28636v1 [cs.CV] 21 Nov 2026

摘要——**高斯泼溅为静态和动态三维场景的高保真渲染提供了高效的表示方式，推动了对高斯资产进行版权保护的需求。在 CLIP引导的高斯水印中，首先基于文本水印表示对通用消息解码器进行预训练，随后将其应用于同时包含水印信息和场景相关语义的视觉特征。然而，纯文本预训练并未明确考虑水印提取过程中遇到的语义干扰。为应对这一挑战，我们提出了 SemGuard，这是一种简单而有效的语义解耦水印框架，旨在通用消息解码器预训练期间学习抑制场景依赖的语义干扰。具体而言，SemGuard 将与消息相关的CLIP文本特征与从无水印场景图像中提取的与消息无关的视觉特征相结合，构建具有可控语义干扰的训练表示，使其与目标高斯资产相关。渐进式语义干扰课程逐步提高视觉特征的贡献，使解码器首先学习可靠的消息表示，随后适应日益增强的场景依赖语义干扰。在针对给定消息长度和语义图像池完成预训练后，解码器被冻结，并可在多个高斯资产之间重复使用；这些资产的视觉语义均已被预训练数据覆盖。在静态和动态高斯泼溅资产上的实验表明，该方法在实现可靠水印提取的同时提高了渲染保真度；此外，在静态资产上的评估进一步表明，其抵抗图像空间失真的鲁棒性得到增强。代码见 https://github.com/shisyy/SemGuard。**

索引词——**高斯泼溅、数字水印、语义解耦、通用消息解码器、CLIP**。

I. 引言

3D 高斯泼溅（3DGS）已成为实现高保真新视角合成和实时渲染的高效表示方法 [1]。将高斯泼溅扩展到动态场景（包括四维高斯泼溅（4DGS））后，进一步能够对时变内容进行建模和渲染 [2]–[4]。随着高斯资产日益易于获取并被广泛共享，保护其知识产权并验证所有权已成为重要问题。数字水印通过将不可感知的二进制消息嵌入数字内容，为此提供了一种实用解决方案 [5]–[7]。与保护单幅图像的水印技术不同，保护高斯资产要求嵌入的消息能够从同一底层表示的不同渲染观测中恢复出来。因此，有效的高斯水印技术框架应在不同视点和时间条件下保持可靠的水印提取能力，同时最大限度地减少水印引起的渲染失真。

近期研究探索了针对高斯泼溅资产的水印技术 [8]–[10]。其中，Guard-Splat [10]通过在 CLIP 的共享视觉-语言嵌入空间中预训练通用消息解码器，引入了一种高效的 CLIP 引导的水印框架 [11]。具体而言，二进制消息会被转换为 CLIP 文本表示，随后由轻量级解码器学习恢复相应的消息位。预训练完成后，解码器被冻结，并被重复用于从渲染的高斯视图的 CLIP 视觉特征中提取水印消息。然而，解码器预训练与实际水印提取之间仍存在差异。在预训练期间，解码器仅根据 CLIP 文本特征学习消息恢复；而在提取期间，它接收的 CLIP 视觉特征同时包含水印相关信息和依赖于场景的语义。尽管两个阶段都在 CLIP 的共享视觉-语言嵌入空间中进行，但纯文本预训练并未让解码器显式接触水印提取过程中遇到的视觉语义干扰。我们

我们将这种以水印为中心的文本训练特征与语义纠缠的视觉提取特征之间的差异称为语义纠缠差距。如图 1 所示，当外观修改受到保持渲染保真度这一条件的限制时，这种差异可能会妨碍水印的可靠恢复。

在保留解码器可复用性的同时解决语义干扰，仍是一项重要挑战。Guard4D [12]通过引入时序建模与语义解耦（TMSD）模块，利用控制片段汇聚时序信息并抑制场景依赖的语义干扰，从而提升动态高斯资产的水印提取效果。尽管其预训练消息解码器保持冻结，TMSD 模块仍需针对资产进行优化，这给完整的水印提取流程带来了额外开销。这促使我们采用一种互补的语义干扰处理方法：与其在资产特定水印优化期间学习额外的语义校正模块，能否让消息解码器在预训练期间学会抑制场景相关视觉语义的影响，并随后在多个高斯资产间复用？这种设计将语义干扰处理转移到共享预训练阶段，使同一个冻结解码器能够为不同资产中的水印嵌入提供指导，而这些资产的视觉语义由训练

Yingying Shi 和 Zhong Zhou 任职于中国北京的北京航空航天大学虚拟现实技术与系统国家重点实验室、计算机科学与工程学院，同时也任职于中国北京的中关村实验室（电子邮箱：yyshi@buaa.edu.cn；zz@buaa.edu.cn）。Bin Zhou 任职于中国北京的北京航空航天大学虚拟现实技术与系统国家重点实验室、计算机科学与工程学院，同时也任职于中国沈阳的和平湾虚拟现实研究院（电子邮箱：zhoubin@buaa.edu.cn）。通讯作者：Bin Zhou（电子邮箱：zhoubin@buaa.edu.cn）。

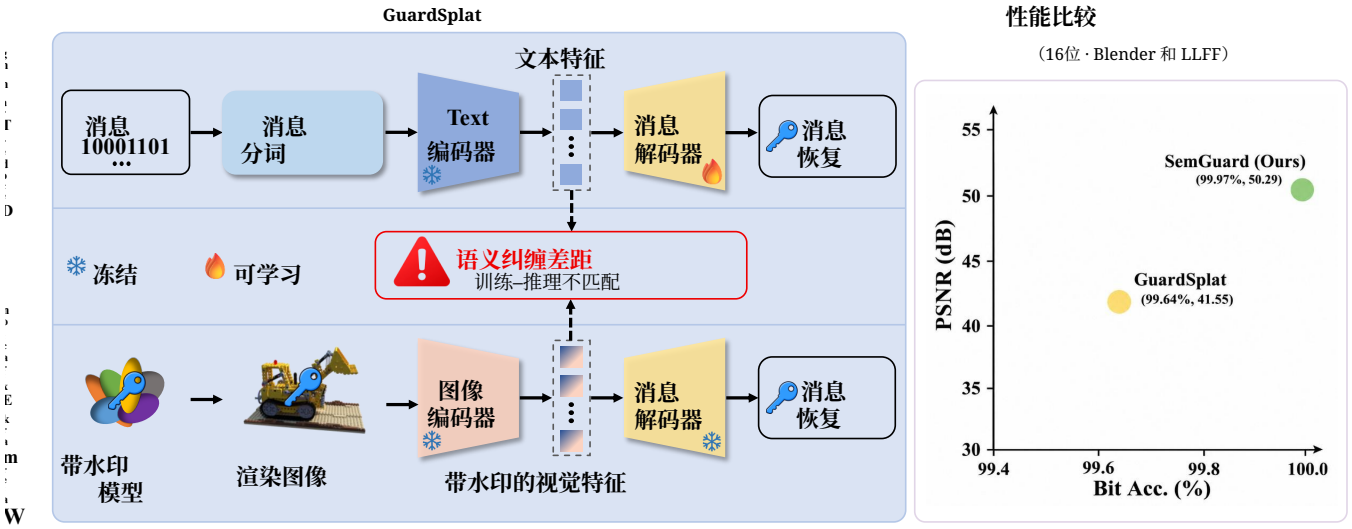

图 1. SemGuard 的研究动机与性能比较。GuardSplat 使用 CLIP 文本特征预训练通用消息解码器，但在提取过程中将其应用于同时包含水印信息和场景相关语义的视觉特征，导致解码器预训练与水印提取之间存在语义差异。右侧面板比较了 Blender 和 LLFF 上的 16位水印提取准确率和峰值信噪比，结果表明，SemGuard 在保持近乎完美的水印提取准确率的同时，显著提升了渲染保真度。

图像池。

为此，我们提出 SemGuard，这是一种简单而有效的语义解耦水印框架，将视觉语义干扰直接融入通用解码器预训练。SemGuard 不再仅从文本水印表示中学习消息恢复，而是将与消息相关的 CLIP 文本特征与从无水印图像中提取的、与消息无关的视觉特征相结合，构造语义纠缠训练特征。辅助视觉特征提供与目标高斯资产相关的视觉语义，而二进制消息仍作为监督目标。通过监督解码器从这些混合表示中恢复消息，促使其依赖水印相关信息，同时降低其对场景依赖语义干扰的敏感性，从而学会在存在预训练中所表示的视觉语义时恢复水印消息。此外，我们引入渐进式语义干扰课程，在预训练期间逐步提高视觉特征的贡献，使解码器先建立可靠的消息表示，再适应日益严峻的语义干扰。对于每种消息长度和语义图像池，SemGuard 训练一个可复用的消息解码器，随后将其冻结并由多个高斯资产共享，无需进一步调整解码器。每个目标资产均通过仅优化其球谐外观偏移来独立嵌入水印，同时保持预训练的 CLIP 编码器和消息解码器固定。如此一来，语义干扰学习在解码器预训练期间基于一系列场景语义联合完成，而水印嵌入仍针对各资产单独进行，并明确受到保留原始渲染外观的约束。

大量实验表明，渲染保真度显著提升，同时仍能可靠地提取水印，并增强了对静态高斯资产中图像空间失真的鲁棒性。具体而言，对于静态高斯资产的16位水印，SemGuard相比GuardSplat将峰值信噪比从41.55 dB提升至50.29 dB，并将学习感知图像块相似度从0.0017降至0.0004，同时保持了99.97%的近乎完美比特准确率。我们的主要贡献总结如下：

- 我们发现，在纯文本解码器预训练过程中缺少场景相关视觉语义，是 CLIP引导的高斯水印面临的一项关键挑战，尤其是在必须满足渲染保真度约束的同时实现可靠的水印恢复时。
- 我们提出 SemGuard，这是一种简单而有效的语义解耦水印框架，在渐进式混合课程下，将与消息相关的CLIP文本特征与场景相关的视觉特征相结合。通过在预训练池所涵盖的视觉语义背景下学习消息恢复，同一个冻结解码器便可在多个高斯资产间重复使用，无需针对特定资产调整解码器。
- 我们在静态和动态高斯泼溅资产上评估了SemGuard，结果表明，在保持可靠水印恢复能力的同时，渲染保真度得到显著提升，并且静态资产对图像空间失真的鲁棒性也有所增强。

II. 相关工作

A. 视觉表示与神经表示中的水印技术

数字水印已在不同的视觉表征中得到广泛研究，旨在保护内容所有权，同时兼顾不可感知性与鲁棒性。经典图像水印方法通常在空间域或变换域中嵌入信息 [5]，[13]，而基于学习的方法则采用可训练的嵌入网络和提取网络，以提高其抵抗图像失真的鲁棒性，并

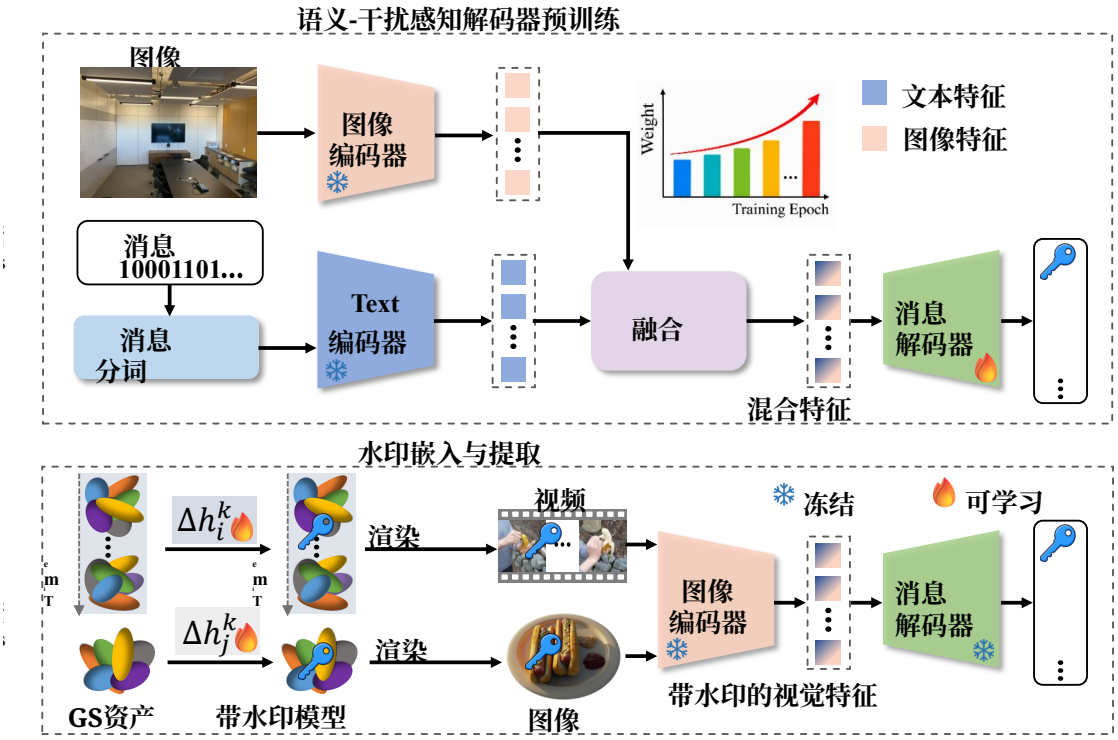

图 2. SemGuard 概览。在阶段 I 中，将与消息相关的 CLIP 文本特征与覆盖目标高斯资产的未加水印场景图像中的视觉特征混合，并采用渐进式混合课程训练一个共享的语义干扰感知消息解码器。在阶段 II 中，冻结预训练解码器和 CLIP 编码器，同时优化资产特定的球谐偏移，以在静态和动态高斯资产中嵌入目标消息，并尽量减少渲染失真。

抵抗后处理操作 [7], [14]–[17]。针对生成式视觉内容 [18], [19] 和几何表征（包括网格和点云），人们也开发了水印技术和信息隐藏技术 [20]–[26]。随着神经渲染的发展 [27]–[29], ，水印技术进一步扩展至隐式场景表示，其中消息被嵌入神经辐射场，并从合成视图中恢 [30], [31] 复。ZeroW-NeRF [32]进一步通过不修改原始模型参数的不确定性驱动的零水印策略，探索神经场景版权保护。这些研究表明，水印技术和信息隐藏技术适用于视觉、几何和神经场景表征，从而推动了对显式高斯资产保护的进一步研究。

B. 静态高斯资产水印

近年来，3D 高斯泼溅技术取得了进展，通过自适应球谐建模和多尺度渲染策略提升了渲染保真度 [33], [34]。3D 高斯泼溅的显式可微表示 [1], [35]使得水印信息能够直接嵌入高斯属性中。然而，修改这些属性可能会引入可见的渲染失真，因此需要谨慎权衡水印可恢复性与视觉保真度。现有方法通过不同的表示感知嵌入策略应对这一挑战。GaussianMarker [8]采用不确定性引导的高斯致密化来嵌入水印信息，同时限制渲染失真。GS-Hider [36] 将原始球谐系数替换为耦合安全特征，并采用独立的场景与消息解码器来恢复渲染内容和隐藏信息。与此同时，3D-GSW [9]结合频率引导的致密化、梯度屏蔽和小波域监督，以提升渲染质量，并增强对图像级和模型级失真的鲁棒性。与这些侧重嵌入的方法相辅相成，GuardSplat [10]利用 CLIP [11] 共享的视觉-语言嵌入空间，通过文本水印表示预训练通用消息解码器，随后将其冻结，以便从高斯渲染结果中提取消息。尽管其解码器设计具有可复用性，GuardSplat仍依赖纯文本预训练，未明确考虑水印提取过程中遇到的场景相关视觉语义，因此语义干扰对保真度-可恢复性权衡的影响尚未得到充分解决。

C. 动态高斯资产水印

与静态高斯资产相比，动态高斯表示因时间形变和外观变化而给水印提取带来了额外挑战 [2]–[4], [37]。这些时间变化进一步增加了在保持渲染保真度的同时可靠恢复水印的难度。Guard4D [12] 提出了时间建模与语义解耦（TMSD）模块，该模块利用从未加水印资产渲染得到的干净及受噪声扰动的控制片段来汇总时序信息并减轻语义干扰。尽管其预训练消息解码器仍保持

冻结状态，TMSD 模块仍需要额外的资产特定优化。Mark4D [38] 通过在潜在视频-文本空间中进行水印解码、采用轨迹对齐偏移量以及运动自适应损失加权，着重提升时间一致性。然而，如何在时间变化下可靠地恢复水印，同时保持渲染保真度并避免额外的资产特定语义处理，仍是一项重要挑战。

III. 方法

如图 2 所示，SemGuard 包含两个阶段：语义干扰感知解码器预训练和球谐偏移高斯水印嵌入。在阶段 I 中，我们将与消息相关的 CLIP 文本特征与从无水印场景图像中提取的、与消息无关的视觉特征相结合，使解码器能够在与目标高斯资产相关的、依赖场景的语义干扰下学习水印恢复。针对给定的消息长度和语义图像池完成预训练后，解码器将被冻结，并在预训练数据涵盖的多个高斯资产上重复使用。在阶段 II 中，我们优化特定于资产的 SH 偏移量来嵌入目标水印，同时保持高斯几何结构、变形参数、CLIP 编码器和消息解码器不变。此设计旨在提升渲染保真度，同时确保可靠的水印提取。

A. 高斯泼溅基础知识

一个高斯泼溅资产表示为一组各向异性高斯基元 [1]:

$$\mathcal{G} = \{(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \alpha_i, \mathbf{h}_i)\}_{i=1}^{N}, \tag{1}$$

其中，N 表示高斯的数量，而 $\boldsymbol{\mu}_i$、$\boldsymbol{\Sigma}_i$、α_i和$\mathbf{h}_i$ 分别表示第 i个高斯的位置信息、协方差、不透明度和球谐（SH）外观系数。球谐系数由直流分量$\mathbf{h}_i^{\mathrm{dc}}$和高阶分量$\mathbf{h}_i^{\mathrm{rest}}$组成，用于描述高斯基元在不同观察方向下的外观。给定相机位姿 π，可微渲染器会生成图像 $\mathbf{I} = \mathcal{R}(\mathcal{G}, \pi)$。对于动态高斯资产，时间形变会随时间变换高斯基元，渲染图像表示为$\mathbf{I}_t = \mathcal{R}(\mathcal{G}, \pi, t)$ [3]。

B. 语义干扰感知解码器预训练

语义干扰下的解码器学习。为缓解解码器预训练与水印提取之间的语义差异，我们提出一种感知语义干扰的预训练策略，将场景相关视觉语义融入解码器学习。给定二进制消息 $\mathbf{m} \in \{0,1\}^L$，其中 L 表示消息长度，我们首先使用预定义的二进制到词元的映射将消息转换为 CLIP 兼容的词元序列$\mathbf{t_m}$。随后，冻结的 CLIP 文本编码器 E_T 提取相应的消息相关表征:

$$\mathbf{F}_T = E_T(\mathbf{t_m}). \tag{2}$$

为引入视觉语义干扰，我们从语义图像池中选取一幅辅助图像$\mathbf{I}_s$，并使用冻结的 CLIP 视觉编码器 E_V对其进行编码:

$$\mathbf{F}_I = E_V(\mathbf{I}_s). \tag{3}$$

构建语义图像池旨在让解码器接触可能干扰第二阶段水印提取的视觉内容。在我们的主要设置中，该图像池包含来自与目标高斯资产对应场景的未加水印图像，从而在 CLIP 的视觉特征空间中为这些资产提供语义覆盖。这些图像在未经过显式语义相似度筛选或特定于消息的选择的情况下被纳入，其内容与二进制水印标签相互独立。我们进一步考虑了一个扩展图像池，其中还加入了来自未参与相应水印嵌入实验场景的图像。这种混合来源设置使我们能够研究，引入额外的非目标场景语义会改善还是削弱保真度-可恢复性权衡。在这两种设置中，所有图像均由水印提取第二阶段所使用的同一个冻结 CLIP 视觉编码器进行处理。

我们在共享的 CLIP 嵌入空间中结合与消息相关的文本特征和辅助视觉特征:

$$\mathbf{F}_{mix} = (1-\rho)\mathbf{F}_T + \rho\mathbf{F}_I, \tag{4}$$

其中，$\rho \in [0,1)$是分配给视觉特征的名义混合系数。当 $\rho = 0$ 时，解码器仅接收文本水印表示，对应于纯文本解码器预训练。随着 ρ 增大，视觉特征被分配的名义权重也随之增加，从而在混合表示中引入更强的场景相关语义干扰。所得特征在传递给轻量级消息解码器 D_ψ之前会进行归一化:

$$\mathbf{o} = D_\psi\left(\frac{\mathbf{F}_{mix}}{\|\mathbf{F}_{mix}\|_2}\right), \tag{5}$$

其中 $\mathbf{o} \in \mathbb{R}^L$ 表示预测的消息逻辑值。解码器使用二元交叉熵损失进行优化:

$$\mathcal{L}_{dec} = \mathrm{BCE}(\mathbf{o}, \mathbf{m}), \tag{6}$$

其中，二元交叉熵（BCE）由预测逻辑值直接计算，并在消息位和训练样本上取平均。在 SemGuard 中，语义解耦是指减少与消息无关的场景语义对水印预测的影响，而非显式地将 CLIP 特征分解为独立的语义子空间和水印子空间。由于辅助视觉特征并不决定目标消息，从混合表示中学习恢复消息会促使解码器依赖与消息相关的信息，同时降低对场景依赖变化的敏感度。

渐进式语义干扰课程。在解码器尚未建立文本表示与二进制消息之间可靠的映射关系之前，直接引入强烈的视觉干扰可能会使消息恢复变得困难。为促进解码器学习，我们采用渐进式语义干扰课程，逐步提高

表 I Blender 和 LLFF 上水印可恢复性与渲染保真度的定量比较。 ⋆表示使用来自 Blender 和 LLFF 的静态场景图像进行 SemGuard 解码器预训练，而 † 还纳入了来自 D-NeRF 和 HyperNeRF 的动态场景图像。粗体表示各项指标的最佳结果。

方法	16 比特				32 比特				48 比特			
	比特准确率 ↑	峰值信噪比 ↑	结构相似性指数 ↑	学习感知图像块相似度 ↓	比特准确率 ↑	峰值信噪比 ↑	结构相似性指数 ↑	学习感知图像块相似度 ↓	比特准确率 ↑	峰值信噪比 ↑	结构相似性指数 ↑	学习感知图像块相似度 ↓
基于 *NeRF* 的水印方法												
CopyRNeRF [30]	91.16	26.29	0.9100	0.0380	78.08	26.13	0.8960	0.0410	60.06	27.56	0.8950	0.0660
WateRF [31]	95.67	32.79	0.9480	0.0330	88.58	31.19	0.9360	0.0400	85.82	30.86	0.9300	0.0400
基于水印图像构建的 *3DGS*												
CIN [17] + 3DGS [1]	56.73	31.95	0.9194	0.1027	53.13	31.74	0.9279	0.0944	55.78	30.25	0.9139	0.1120
自监督学习 [16] + 3DGS [1]	58.94	36.51	0.9737	0.0094	61.85	35.24	0.9706	0.0179	58.79	35.88	0.9710	0.0123
HiDDeN [6] + 3DGS [1]	63.07	31.59	0.9790	0.0171	52.46	34.51	0.9682	0.0209	53.08	33.42	0.9687	0.0299
DwtDctSvd [13] + 3DGS [1]	55.44	34.78	0.9582	0.0399	53.15	32.32	0.9477	0.0547	51.83	31.09	0.9302	0.0704
StegaStamp [7] + 3DGS [1]	79.72	35.52	0.9697	0.0181	82.36	34.04	0.9601	0.0265	83.97	32.54	0.9523	0.0406
基于高斯表示的水印												
3DGS [1]搭配 WateRF [31]	92.89	31.01	0.9678	0.0475	90.15	29.56	0.9611	0.0492	87.30	29.13	0.9562	0.0534
3D-GSW [9]	98.45	36.07	0.9834	0.0226	98.01	36.16	0.9811	0.0422	94.13	33.42	0.9821	0.0431
GaussianMarker [8]	99.36	34.42	0.9822	0.0124	98.85	33.98	0.9788	0.0163	98.25	32.12	0.9723	0.0234
GuardSplat [10]	99.64	41.55	0.9957	0.0017	99.04	39.40	0.9939	0.0022	98.29	38.90	0.9923	0.0028
使用不同预训练语义图像池的 *SemGuard*												
SemGuard（我们的方法†）	99.91	46.97	0.9986	0.0007	**99.52**	**43.72**	**0.9975**	**0.0011**	98.05	**42.58**	**0.9969**	**0.0014**
SemGuard（我们的方法⋆）	**99.97**	**50.29**	**0.9993**	**0.0004**	99.47	43.59	**0.9975**	**0.0011**	**98.72**	42.32	**0.9969**	**0.0014**

预训练过程中的视觉混合系数。具体而言，在第 e个训练轮次，干扰系数定义为：

$$\rho_e = \rho_{\min} + (\rho_{\max} - \rho_{\min})\frac{e}{E}, \tag{7}$$

其中， E 表示训练轮次总数，$e \in \{1, \ldots, E\}$。系数 ρ_e 在公式（4）中替代了 ρ。该调度策略逐步将解码器学习从以文本为主导的表示转向更强的视觉干扰，从而促进在日益严峻的语义变化下进行消息恢复。针对给定的消息长度和语义训练分布完成预训练后，解码器被冻结，并在多个高斯资产之间复用。

C. 球谐偏移高斯水印嵌入

遵循 Guard-Splat 的球谐偏移参数化， [10], SemGuard 通过优化高斯球谐系数的加性扰动来嵌入目标消息，同时保持原始几何结构和变形参数不变。令 $\mathcal{G}$ 和 $\mathcal{G}$ 分别表示原始高斯资产和含水印高斯资产。对于每个高斯基元 i，含水印球谐系数定义为：

$$\widetilde{\mathbf{h}}_i^k = \mathbf{h}_i^k + \Delta\mathbf{h}_i^k, \quad k \in \{\mathrm{dc}, \mathrm{rest}\}, \tag{8}$$

其中， $\Delta\mathbf{h}_i^k$ 表示相应球谐分量的可学习外观偏移。这些偏移量被收集到$\Delta\mathbf{H}^{\mathrm{dc}}$ 和 $\Delta\mathbf{H}^{\mathrm{rest}}$中。在水印嵌入过程中，仅优化这些球谐偏移量，而动态高斯资产的高斯位置、尺度、旋转、不透明度和变形参数保持不变。

从静态高斯资产中提取水印。给定相机位姿 π，我们分别将带水印的资产及其原始对应资产渲染为 $\mathbf{I} = \mathcal{R}(\mathcal{G}, \pi)$ 和 $\mathbf{I} = \mathcal{R}(\mathcal{G}, \pi)$。带水印的渲染结果经冻结的 CLIP 视觉编码器处理，其视觉特征经过归一化后传入冻结的消息解码器：

$$\mathbf{o} = D_\psi\left(\frac{E_V(\widetilde{\mathbf{I}})}{\|E_V(\widetilde{\mathbf{I}})\|_2}\right). \tag{9}$$

消息损失通过可微渲染引导球谐偏移优化，使梯度能够经由冻结的视觉编码器和消息解码器传播至可学习的外观偏移。

从动态高斯资产中提取水印。对于动态高斯资产，水印提取必须考虑由时间变形和观看条件变化引起的外观变化。因此，我们在消息解码之前采用一种简单的时间特征聚合策略。考虑一个采样的视点序列 v，其中包含 J 帧，其相机姿态为 $\{\pi_{v,j}\}_{j=1}^{J}$，时间戳为 $\{t_j\}_{j=1}^{J}$。相应的含水印帧为 ，渲染为 $\mathbf{I}_{v,j} = \mathcal{R}(\mathcal{G}, \pi_{v,j}, t_j)$。在归一化和消息解码之前，对其 CLIP 视觉特征进行平均：

$$\begin{aligned} \overline{\mathbf{F}}_{V,v} &= \frac{1}{J}\sum_{j=1}^{J} E_V\left(\widetilde{\mathbf{I}}_{v,j}\right), \\ \mathbf{o}_v &= D_\psi\left(\frac{\overline{\mathbf{F}}_{V,v}}{\|\overline{\mathbf{F}}_{V,v}\|_2}\right). \end{aligned} \tag{10}$$

此处，$\overline{\mathbf{F}}_{V,v}$ 表示视点序列 v 的时间聚合视觉特征。不同于逐帧解码，该策略针对每个视点序列生成一条消息预测，并在采样序列上使用相同的目标消息进行监督。该聚合直接在冻结的CLIP特征空间中进行。

D. 优化与验证

在水印嵌入过程中，预训练CLIP视觉编码器 E_V 和消息解码器 D_ψ 保持固定，仅优化高斯球谐偏移。优化目标在水印可恢复性、原始渲染外观的保持以及球谐扰动的幅度之间进行权衡。

消息恢复损失。对于包含 B 个解码视图或视点序列的小批量数据，我们最小化预测的消息逻辑值与目标二进制消息之间的二元交叉-

表II D-NeRF 和 HyperNeRF 上水印可恢复性与渲染保真度的定量比较。 △ 表示使用来自 D-NeRF 和 HyperNeRF 的动态场景图像进行解码器预训练，而 † 还额外包括来自 Blender 和 LLFF 的静态场景图像。

方法	16 比特				32 比特				48 比特			
	比特准确率 ↑	峰值信噪比 ↑	结构相似性指数 ↑	学习感知图像块相似度 ↓	比特准确率 ↑	峰值信噪比 ↑	结构相似性指数 ↑	学习感知图像块相似度 ↓	比特准确率 ↑	峰值信噪比 ↑	结构相似性指数 ↑	学习感知图像块相似度 ↓
基于高斯表示的水印方法												
GuardSplat [10]	87.61	43.82	0.9972	0.0014	76.06	40.35	0.9947	0.0029	72.61	38.98	0.9935	0.0039
Mark4D [38]	88.27	43.96	0.9974	0.0013	75.20	41.39	0.9946	0.0030	73.57	39.24	0.9934	0.0038
采用不同预训练语义图像池的 *SemGuard*												
SemGuard（我们的方法†）	**95.35**	**47.39**	**0.9987**	0.0010	**91.91**	**45.41**	**0.9981**	**0.0011**	86.85	43.31	0.9971	**0.0015**
SemGuard（我们的方法△）	92.75	47.23	0.9985	**0.0008**	90.67	44.05	0.9975	0.0014	**87.89**	**43.35**	**0.9972**	**0.0015**

预测的消息逻辑值与目标二进制消息之间的熵：

$$\mathcal{L}_{\text{msg}} = \frac{1}{B}\sum_{b=1}^{B}\text{BCE}(\mathbf{o}_b, \mathbf{m}), \tag{11}$$

其中，每个$\mathbf{o}_b$ 对应静态资产的单个渲染视图，或动态资产经过时间聚合的视点序列。所有采样观测均以相同的目标消息$\mathbf{m}$ 作为监督信号。

渲染保真度损失。为保持渲染保真度，我们约束带水印的渲染结果 $\widetilde{\mathbf{I}}$ 与原始未加水印高斯资产在相同视点和时间戳下对应的渲染结果$\mathbf{I}$ 之间的差异。我们结合逐像素重建、结构相似性和感知一致性：

$$\begin{aligned}\mathcal{L}_{\text{recon}} = &(1-\lambda_{\text{ssim}})\,\text{MAE}(\widetilde{\mathbf{I}}, \mathbf{I}) \\ &+ \lambda_{\text{ssim}}\left[1 - \text{SSIM}(\widetilde{\mathbf{I}}, \mathbf{I})\right] \\ &+ \lambda_{\text{lpips}}\,\text{LPIPS}(\widetilde{\mathbf{I}}, \mathbf{I}),\end{aligned} \tag{12}$$

其中，MAE 表示平均绝对像素误差。对于动态资产，渲染保真度损失在各个帧上分别评估，而消息恢复则通过时间聚合特征进行监督。

SH偏移正则化。为抑制过度的外观扰动，我们对优化后的 SH 偏移量的幅度进行惩罚：

$$\mathcal{L}_{\text{off}} = \text{MSE}(\Delta\mathbf{H}^{\text{dc}}, \mathbf{0}) + \text{MSE}(\Delta\mathbf{H}^{\text{rest}}, \mathbf{0}), \tag{13}$$

其中，MSE 表示逐元素均方误差。结合这三个损失项，最终的嵌入目标为：

$$\mathcal{L} = \lambda_{\text{msg}}\mathcal{L}_{\text{msg}} + \lambda_{\text{recon}}\mathcal{L}_{\text{recon}} + \lambda_{\text{off}}\mathcal{L}_{\text{off}}. \tag{14}$$

我们的目标不是不惜任何视觉代价最大化比特准确率，而是在尽可能减小水印引起的渲染失真的同时，实现可靠的消息提取。消息恢复项用于确保水印可恢复性，而重建项和偏移正则化项则用于避免对原始高斯外观进行不必要的修改。

水印提取评估。 嵌入水印后，将目标消息$\mathbf{m}$ 与从渲染视图中提取的预测消息进行比较。令$\mathbf{o}_v$ 表示使用公式（9）或公式（10）从第 v 个评估视图或视点序列中获得的消息逻辑值。第 q 个消息比特的预测值为 $\hat{m}_{v,\,q} = \mathbb{I}(o_{v,\,q} > 0)$，其中 $\mathbb{I}$ 表示指示函数。给定 K 个评估视图或视点序列，比特准确率的平均值为：

$$\mathcal{A} = \frac{100}{KL}\sum_{v=1}^{K}\sum_{q=1}^{L}\left(1 - |\hat{m}_{v,q} - m_q|\right). \tag{15}$$

此处， $\mathcal{A}$表示在评估视图或视点序列上取平均后，正确恢复的消息位所占的百分比。该指标用于衡量静态和动态高斯资产的水印提取准确率。

IV. 实验

A. 数据集与评估协议

我们在静态和动态高斯资产上对 SemGuard 进行评估。对于 3DGS水印，我们使用 Blender [40] 和 LLFF [41]。对于 4DGS水印，我们使用 D-NeRF [42]和 HyperNeRF [43]。在第一阶段预训练期间，我们从无水印场景图像中构建语义图像池，以提供后续在第二阶段处理的高斯资产的视觉语义覆盖。我们研究三种配置：包含 Blender 和 LLFF 图像（ ⋆ ）的静态池、包含 D-NeRF 和 HyperNeRF 图像（ △ ）的动态池，以及包含全部四个数据集图像（†）的混合池。对于静态水印，静态池提供目标场景的语义覆盖，而混合池则额外引入了动态场景中的图像，这些图像在相应实验中未被嵌入水印。对于动态水印，二者的作用相反：动态池覆盖目标场景，而混合池则额外纳入静态场景图像。这些配置旨在考察额外的非目标场景语义所产生的影响，而非建立场景不相交泛化。对于每种消息长度和图像池，我们预训练并冻结一个解码器，以便在所有对应的高斯资产之间复用，同时针对每个资产独立优化 SH 偏移量。在所评估的协议中，目标水印消息从解码器预训练所使用的消息集合中采样，而水印嵌入则针对每个高斯资产独立执行。

B. 评估指标

我们通过综合考量渲染保真度和水印提取可靠性来评估高斯水印的性能，尤其关注在保留原始

表 III 使用来自静态资产的场景图像，在 Blender 和 LLFF 上对语义干扰强度及渐进式训练策略进行消融研究。对比固定混合系数 $\rho \in \{0, 0.1, 0.3, 0.5, 0.7, 0.9\}$ 与渐进式干扰调度，以评估其对渲染保真度和水印可恢复性的影响。

策略	16 比特				32 比特				48 比特			
	比特准确率 ↑	峰值信噪比 ↑	结构相似性指数 ↑	学习感知图像块相似度 ↓	比特准确率 ↑	峰值信噪比 ↑	结构相似性指数 ↑	学习感知图像块相似度 ↓	比特准确率 ↑	峰值信噪比 ↑	结构相似性指数 ↑	学习感知图像块相似度 ↓
固定干扰权重												
0	99.64	41.55	0.9957	0.0017	99.04	39.40	0.9939	0.0022	98.29	38.90	0.9923	0.0028
0.1	99.95	47.10	0.9987	**0.0004**	98.17	39.91	0.9942	0.0028	95.81	38.25	0.9923	0.0036
0.3	99.93	46.27	0.9984	0.0008	98.02	40.58	0.9951	0.0023	98.12	40.25	0.9950	0.0024
0.5	99.96	48.23	0.9990	0.0005	98.80	41.14	0.9958	0.0020	98.59	40.66	0.9955	0.0021
0.7	99.96	50.10	**0.9993**	**0.0004**	99.21	42.59	0.9967	0.0015	98.70	41.88	0.9963	0.0017
0.9	99.89	45.45	0.9981	0.0010	93.11	43.16	0.9972	0.0012	86.42	41.84	0.9967	0.0018
渐进式干扰调度												
Step	99.96	48.10	0.9989	0.0005	99.18	42.59	0.9969	0.0014	98.37	42.23	0.9968	**0.0014**
连续	**99.97**	**50.29**	**0.9993**	**0.0004**	**99.47**	**43.59**	**0.9975**	**0.0011**	**98.72**	**42.32**	**0.9969**	**0.0014**

表 IV 图像空间失真下 16 位 3DGS 水印的比特准确率（%）。⋆ 表示使用来自 Blender 和 LLFF 的图像进行解码器预训练，而 † 还包括来自 D-NeRF 和 HyperNeRF 的图像。

方法	比特准确率（%）↑									
	None	噪声 (σ=0.1)	旋转 ($\pm\pi/6$)	缩放 ($\leq$ 25%)	Blur (σ=0.1)	Crop (40%)	亮度 (0.5∼1.5)	JPEG (10% 质量)	VAE 攻击 [39]	综合
CopyRNeRF [30]	91.16	90.04	88.13	89.33	90.06	–	–	–	–	–
WateRF [31]	95.67	95.36	93.13	93.29	95.25	95.40	90.91	86.99	51.73	84.12
3DGS [1] 搭配 WateRF [31]	92.89	87.35	88.28	90.33	91.92	89.07	88.71	88.49	55.48	86.37
GaussianMarker [8]	99.36	99.13	70.84	97.89	94.40	98.52	95.78	86.22	52.00	83.49
3D-GSW [9]	98.36	84.95	88.14	94.98	96.43	93.18	94.68	93.71	67.23	91.13
GuardSplat [10]	99.64	99.60	94.56	98.75	99.27	98.71	97.46	94.70	82.35	93.38
SemGuard（我们的方法†）	99.91	99.79	96.86	99.56	99.90	99.24	98.39	96.48	95.04	96.65
SemGuard（我们的方法⋆）	99.97	99.81	97.01	99.98	99.94	99.15	99.40	96.53	95.26	97.24

渲染外观的同时确保可靠的消息恢复。为量化水印造成的视觉失真，而非相对于数据集真实图像的重建误差，我们报告在相同相机视点和时间戳下，带水印与未加水印的高斯渲染结果之间的 PSNR、SSIM [44], 和 LPIPS [45]。PSNR 和 SSIM 越高，且 LPIPS 越低，表示对原始渲染外观的保留效果越好。水印提取可靠性以比特准确率（Bit Acc.，%）衡量，表示正确恢复的消息位所占比例。对于 3DGS 资产，每个渲染视点均独立解码；对于 4DGS 资产，则先对每个视点序列中的 CLIP 视觉特征进行时间平均，再进行消息解码，随后对所评估的序列取比特准确率平均值。我们还评估了静态资产在噪声、旋转、缩放、模糊、裁剪、亮度变化、JPEG 压缩、基于 VAE 的攻击以及组合失真等图像空间失真下的水印提取鲁棒性。

C. 实现细节

我们使用 CLIP ViT-B/32 [11] 作为冻结的视觉和文本编码器。在阶段一的解码器学习中，仅优化轻量级消息解码器。在阶段二的水印嵌入过程中，解码器和 CLIP 图像编码器均保持冻结，仅优化 SH 偏移量，同时固定高斯几何结构、不透明度和变形参数。对于不同消息长度，我们分别训练解码器，并使用 Adam 优化 SH 偏移量。详细的训练配置和超参数见补充材料。

D. 与其他方法的比较

3DGS 资产上的结果。表 I 比较了在 Blender 和 LLFF 上使用 16、32 和 48 位消息时的水印提取效果与渲染保真度。在采用基于高斯表示的水印技术的方法中，GuardSplat 尤为相关，因为它同样使用预训练的 CLIP 引导解码器和基于 SH 的水印嵌入。SemGuard 的最大优势在于，在保持水印提取高度可靠的同时保留渲染保真度。对于 Blender 和 LLFF 上的 16 位水印，GuardSplat 的比特准确率已达到近乎饱和的 99.64%，因此仅提升提取准确率的空间有限。尽管如此，SemGuard(⋆) 将 PSNR 从 41.55 dB 提高至 50.29 dB，将 SSIM 从 0.9957 提高至 0.9993，并将 LPIPS 从 0.0017 降至 0.0004，同时比特准确率达到 99.97%。这些结果表明，语义干扰感知解码器使基于 SH 偏移的嵌入过程能够在不牺牲水印可恢复性的前提下，更好地保留原始渲染外观。在 32 位设置下，与 GuardSplat 相比，SemGuard(†) 将 PSNR 从 39.40 dB 提高至 43.72 dB，比特准确率也从 99.04% 提高至 99.52%。在 48 位设置下，SemGuard(⋆) 的 PSNR 达到 42.32 dB，比特准确率为 98.72%；GuardSplat 的相应数值则为 38.90 dB 和 98.29%。语义图像池也会影响结果。尽管混合来源的变体取得了最高的 48 位 PSNR（42.58 dB），

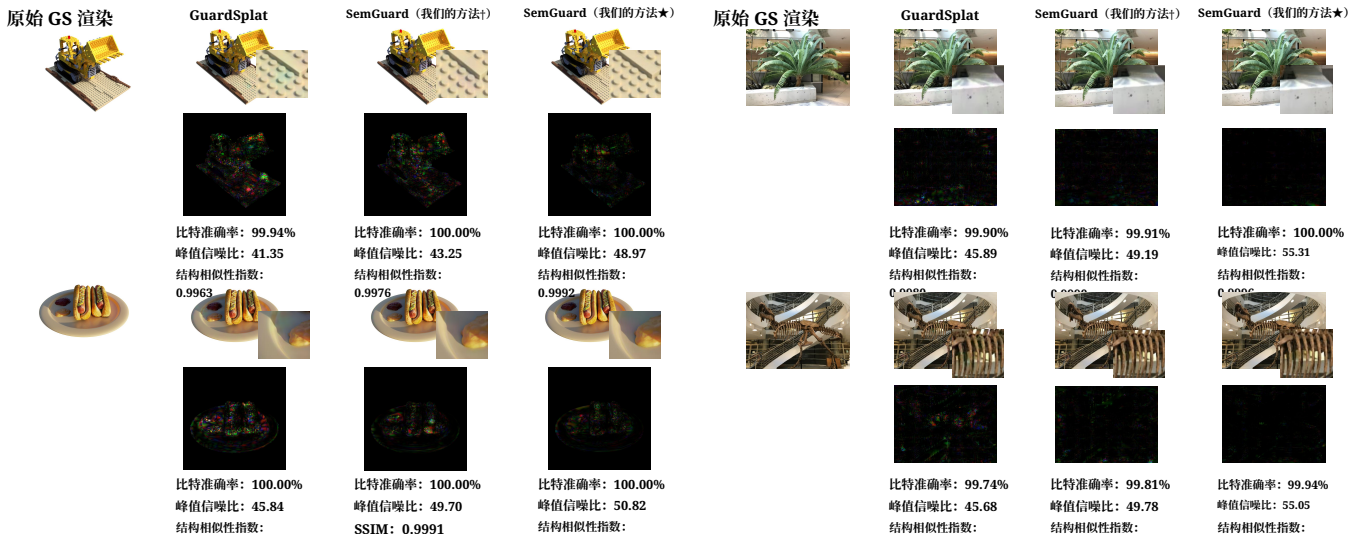

图3. Blender 和 LLFF 上16位水印的定性比较。误差图展示了带水印与原始高斯渲染结果之间的绝对差异，并将其放大10倍以便于观察；颜色越深表示偏差越小。

但其比特准确率（98.05%）低于 GuardSplat（98.29%）和 SemGuard(⋆)（98.72%）。这些结果表明，语义图像池的效果可能取决于其视觉内容对后续水印提取过程中遇到的语义干扰的表征程度。从非目标场景中纳入图像，并不一定能改善渲染保真度与水印可恢复性之间的平衡。

4DGS 资产上的结果。表II使用视点序列预测评估了 D-NeRF 和 HyperNeRF 上的动态水印，报告的聚合结果对两个基准数据集赋予相同权重。该设置尤其具有代表性，因为运动和时序外观变化会改变解码器可利用的视觉证据。在32位设置下，SemGuard(†)显著提升了渲染保真度，达到了45.41 dB峰值信噪比、0.9981结构相似性指数和0.0011学习感知图像块相似度；相比之下，GuardSplat分别为40.35 dB、0.9947和0.0029。同时，比特准确率从76.06%提升至91.91%，表明视觉保真度的提高伴随着更可靠的水印恢复。与 Mark4D 相比，在所评估的协议下，SemGuard(†)在渲染保真度和提取准确率方面也取得了更好的结果。在48位设置下，SemGuard(△) đạt到了87.89%的比特准确率和43.35 dB峰值信噪比，而SemGuard(†)分别为86.85%和43.31 dB。这些结果表明，在解码器预训练期间加入额外的静态场景图像，并不能始终改善动态资产上的消息恢复效果，而两个变体仍保持相近的渲染保真度。由于对比方法采用了不同的解码和时序处理设计，该表反映的是其完整水印处理流程的性能，而非单独评估解码器预训练的影响。

E. 消融研究

语义干扰和渐进式连续课程的影响。表 III 研究了场景相关视觉干扰及其加权策略对 3DGS 水印的影响。$\rho = 0$ 设置表示纯文本解码器预训练，并作为评估引入视觉语义干扰效果的参考基线。与渐进式训练的比较展示了由此产生的渲染保真度和水印可恢复性差异。与该基线相比，连续课程在 16 比特时将峰值信噪比从 41.55 dB 提升至 50.29 dB，在 32 比特时从 39.40 dB 提升至 43.59 dB，在 48 比特时从 38.90 dB 提升至 42.32 dB，同时保持或提升水印提取准确率。这些结果表明，在解码器预训练期间融入场景相关视觉语义，可以在所评估的 SH偏移嵌入协议下改善保真度-可恢复性权衡，但最终性能取决于干扰强度及其训练计划。固定干扰系数的影响呈非单调性。在 32 比特时，将 ρ 从 0.7 增加至 0.9 可使峰值信噪比从 42.59 dB 提升至 43.16 dB，但会使比特准确率从 99.21% 降至 93.11%，凸显了平衡渲染保真度与水印可恢复性的重要性。在 48 比特时，连续课程达到 98.72% 的比特准确率和 42.32 dB 的峰值信噪比，而固定$\rho = 0.9$下则为 86.42% 和 41.84 dB。尽管固定 $\rho = 0.7$ 在 48 比特时仍具有竞争力，达到 98.70% 的比特准确率和 41.88 dB 的峰值信噪比，但在所评估的消息长度范围内，连续课程提供了更有利的保真度-可恢复性权衡。这些结果证明，在解码器预训练期间融入场景相关视觉语义具有明显优势，而渐进加权则可进一步改善渲染保真度与可靠水印恢复之间的平衡。

F. 对失真的鲁棒性

表 IV 比较了 Blender 和 LLFF 上图像空间失真条件下的16位水印提取准确率。这些实验评估的是对渲染图像修改的鲁棒性，而非对高斯模型参数的攻击。在

基于VAE的攻击下，SemGuard(⋆) 的比特准确率仍达到 95.26%，高于 GuardSplat 的 82.35%；相较于无攻击时的提取准确率，两者分别下降了 4.71 和 17.29 个百分点。SemGuard(⋆) 在旋转和 JPEG 压缩下的准确率也分别达到 97.01% 和 96.53%，而 GuardSplat 分别为 94.56% 和 94.70%。这些结果表明，在所评估的图像空间失真下，SemGuard 的水印提取比 GuardSplat 更可靠。此外，对于静态资产，SemGuard(⋆) 的表现通常优于 SemGuard(†)，包括在组合攻击下（97.24% 对比 96.65%），这表明引入额外的语义来源未必能提升目标高斯场景的鲁棒性。

G. 定性比较

图 3 展示了在具有代表性的 Blender 和 LLFF 场景上的定性比较。GuardSplat 和 SemGuard 的两种变体均实现了近乎饱和的 16位水印提取准确率，而 SemGuard(⋆) 更好地保留了原始渲染外观，这一点体现在误差图中的差异减小，并在放大插图中更加明显。这些视觉观察与其更高的峰值信噪比和结构相似性指数相一致，表明其在不影响水印恢复的情况下提升了渲染保真度。与混合源变体相比，SemGuard(⋆) 在所示场景中也产生了更小的渲染偏差，这表明在解码器预训练期间融入额外的动态场景语义，并不一定能提升目标静态高斯资产的渲染保真度。

V. 结论

我们提出了 SemGuard，这是一种简单而有效的语义解耦水印框架，它将无水印图像中与场景相关的视觉语义融入通用消息解码器预训练中。通过逐步提高场景相关视觉特征的贡献，SemGuard 促使解码器降低对语义干扰的敏感度，同时保持可靠的水印恢复能力。在 SH 偏移水印嵌入过程中，预训练解码器保持冻结状态，并可在多个视觉语义涵盖于预训练图像池中的高斯资产之间重复使用。针对静态和动态高斯泼溅资产的实验表明，渲染保真度显著提升，具体体现在峰值信噪比、结构相似性指数和学习感知图像块相似度等指标上，同时仍能可靠地提取水印。对静态高斯资产的其他评估进一步表明，该方法对图像空间失真具有更强的鲁棒性。结果还表明，水印性能取决于预训练图像池的语义覆盖范围及其构成，以及干扰训练计划。未来的工作将探索如何将解码器预训练扩展到涵盖多样化语义分布的大规模图像数据集，以提升其在更广泛的高斯资产上的泛化能力。

致谢

本研究部分得到了中国国家自然科学基金（U25A20441）、中国国家科技重大专项（2025ZD1700704）、中国国家重点研发计划（2024YFB3311703）和中关村实验室的资助。

参考文献

[1] B. Kerbl、G. Kopanas、T. Leimk¨uhler、G. Drettakis 等，“用于实时辐射场渲染的 3D 高斯泼溅”。*ACM*图形学汇刊，第 42 卷，第 4 期，第 1–14 页，2023年。[2] J. Luiten、G. Kopanas、B. Leibe 和 D. Ramanan，“动态 3D 高斯：通过持久动态视图合成进行跟踪”，载于*2024*年国际 *3D* 视觉会议（*3DV*）。IEEE，2024年，第 800–809 页。[3] G. Wu、T. Yi、J. Fang、L. Xie、X. Zhang、W. Wei、W. Liu、Q. Tian 和 X. Wang，“用于实时动态场景渲染的四维高斯泼溅”，载于*IEEE/CVF*计算机视觉与模式识别会议论文集，2024年，第 20 310–20 320 页。[4] Z. Yang、X. Gao、W. Zhou、S. Jiao、Y. Zhang 和 X. Jin，“用于高保真单目动态场景重建的可变形 3D 高斯”，载于*IEEE/CVF*计算机视觉与模式识别会议论文集，2024年，第 20 331–20 341 页。[5] I. J. Cox、J. Kilian、F. T. Leighton 和 T. Shamoon，“用于多媒体的安全扩频水印”，*IEEE*图像处理汇刊，第 6 卷，第 12 期，第 1673–1687 页，1997年。[6] J. Zhu、R. Kaplan、J. Johnson 和 L. Fei-Fei，“HiDDeN：利用深度网络隐藏数据”，载于欧洲计算机视觉会议论文集（*ECCV*），2018年，第 657–672 页。[7] M. Tancik、B. Mildenhall 和 R. Ng，“StegaStamp：实体照片中的不可见超链接”，载于*IEEE/CVF*计算机视觉与模式识别会议论文集，2020年，第 2117–2126 页。[8] X. Huang、R. Li、Y.-m. Cheung、K. C. Cheung、S. See 和 R. Wan，“GaussianMarker：基于不确定性的 3D 高斯泼溅版权保护”，神经信息处理系统进展，第 37 卷，第 33 037–33 060 页，2024年。[9] Y. Jang、H. Park、F. Yang、H. Ko、E. Choo 和 S. Kim，“3D-GSW：用于鲁棒水印的 3D 高斯泼溅”，载于计算机视觉与模式识别会议论文集，2025年，第 5938–5948 页。[10] Z. Chen、G. Wang、J. Zhu、J. Lai 和 X. Xie，“GuardSplat：面向 3D 高斯泼溅的高效鲁棒水印技术”，载于计算机视觉与模式识别会议论文集，2025年，第 16 325–16 335 页。[11] A. Radford、J. W. Kim、C. Hallacy、A. Ramesh、G. Goh、S. Agarwal、G. Sastry、A. Askell、P. Mishkin、J. Clark 等，“从自然语言监督中学习可迁移的视觉模型”，载于国际机器学习会议。PmLR，2021年，第 8748–8763 页。[12] Y. Shi、Z. Zhou 和 B. Zhou，“Guard4D：通过解耦解码实现面向四维高斯泼溅的鲁棒水印”，载于第三十五届国际人工智能联合会议（*IJCAI*）论文集，2026年，第 1595–1603 页。[13] K. Navas、M. C. Ajay、M. Lekshmi、T. S. Archana 和 M. Sasikumar，“基于 DWT-DCT-SVD 的水印”，载于*2008*年第 *3* 届通信系统、软件与中间件及研讨会国际会议（*COMSWARE'08*）。IEEE，2008年，第 271–274 页。[14] H. Fang、Z. Jia、Y. Qiu、J. Zhang、W. Zhang 和 E.-C. Chang，“De-End：解码器驱动的水印网络”，*IEEE*多媒体汇刊，第 25 卷，第 7571–7581 页，2022年。[15] S. Wu、W. Lu 和 X. Luo，“基于多层水印特征融合的鲁棒水印”，*IEEE*多媒体汇刊，第 27 卷，第 5156–5169 页，2025年。[16] P. Fernandez、A. Sablayrolles、T. Furon、H. J´egou 和 M. Douze，“在自监督潜在空间中对图像进行水印处理”，载于*ICASSP 2022-2022 IEEE*声学、语音与信号处理国际会议（*ICASSP*）。IEEE，2022年，第 3054–3058 页。[17] R. Ma、M. Guo、Y. Hou、F. Yang、Y. Li、H. Jia 和 X. Xie，“迈向盲水印：结合可逆与不可逆机制”，载于第 *30* 届 *ACM* 国际多媒体会议论文集，2022年，第 1532–1542 页。

[18] P. Fernandez、G. Couairon、H. J´egou、M. Douze 和 T. F uron，“稳定签名：将水印植入潜在扩散模型”，载于 IEEE/CVF国际计算机视觉大会论文集，2023年，第 22 466–22 477 页。[19] Y. Wen、J. Kirchenbauer、J. Geiping 和 T. Goldstein，“树轮水印：扩散图像中的不可见指纹”，神经信息处理系统进展，第 36 卷，58 063，2023年 58 047页。– 。[20] E. Praun、H. Hoppe 和 A. Finkelstein，“鲁棒网格水印”，载于第 26 届计算机图形学与交互技术年度会议论文集，1999年，第 49–56 页。[21] R. Ohbuchi、A. Mukaiyama 和 S. Takahashi，“三维形状水印的频域方法”，载于 Computer Graphics Forum，第 21 卷，第 3 期。Wiley Online Library，2002年，第 373–382 页。[22] J. Zhu、Y. Zhang、X. Zhang 和 X. Cao，“用于三维网格隐写的高斯模型”，IEEE信号处理快报，第 28 卷，1733，2021年 1729页。– 。[23] Y. Zhang、J. Zhu、M. Xue、X. Zhang 和 X. Cao，“基于特征保持失真的自适应三维网格隐写”，IEEE图形学与可视化汇刊，第 30 卷，第 8 期，第 5299–5312 页，2023年。[24] X. Zhu、G. Ye、X. Luo 和 X. Wei，“重新思考网格水印：迈向高度鲁棒且自适应的深度三维网格水印”，载于 AAAI人工智能会议论文集，第 38 卷，第 7 期，2024年，第 7784–7792 页。[25] I. Yoo、H. Chang、X. Luo、O. Stava、C. Liu、P. Milanfar 和 F. Yang，“深度三维到二维水印：在三维网格中嵌入消息并从二维渲染图中提取消息”，载于 IEEE/CVF计算机视觉与模式识别会议论文集，2022年，第 10 031–10 040 页。[26] K. A. U. Zaman、M. Z. Alam、M. N. Ali 和 M. H. Miraz，“用于三维点云鲁棒版权保护的深度神经水印”，arXiv 预印本 arXiv:2510.27533，2025年。[27] J. T. Barron、B. Mildenhall、M. Tancik、P. Hedman、R. Martin-Brualla 和 P. P. Srinivasan，“Mip-NeRF：用于抗混叠神经辐射场的多尺度表示”，载于 IEEE/CVF国际计算机视觉大会论文集，2021年，第 5855–5864 页。[28] J. T. Barron、B. Mildenhall、D. Verbin、P. P. Srinivasan 和 P. Hedman，“Mip-NeRF 360：无界抗混叠神经辐射场”，载于 IEEE/CVF计算机视觉与模式识别会议论文集，2022年，第 5470–5479 页。[29] T. M¨uller、A. Evans、C. Schied 和 A. Keller，“借助多分辨率哈希编码实现即时神经图形基元”，ACM图形学汇刊（TOG），第 41 卷，第 4 期，第 1–15 页，2022年。[30] Z. Luo、Q. Guo、K. C. Cheung、S. See 和 R. Wan，“CopyRNeRF：保护神经辐射场的版权”，载于 IEEE/CVF国际计算机视觉大会论文集，2023年，第 22 401–22 411 页。[31] Y. Jang、D. I. Lee、M. Jang、J. W. Kim、F. Yang 和 S. Kim，“WateRF：用于版权保护的辐射场鲁棒水印”，载于 IEEE/CVF计算机视觉与模式识别会议论文集，2024年，第 12 087–12 097 页。[32] X. Zhu、Z. Wang、L. You、X. Yang、Z. Sheng、Y. Gu、K. Zhao 和 D. Zeng，“ZeroW-NeRF：不确定性驱动的神经辐射场零水印版权保护”，IEEE多媒体汇刊，2026年。[33] Z. Huang、M. Xu 和 S. Perry，“StructGS：用于卓越 3D 高斯泼溅的自适应球谐函数与渲染增强”，IEEE多媒体汇刊，2025年。[34] Y. Zhao、G. Chen、B. Wu、Q. Jin 和 T. Zeng，“MSA-Splatting：用于高保真视图合成的多尺度自适应高斯泼溅”，IEEE多媒体汇刊，2026年。[35] Z. Yu、A. Chen、B. Huang、T. Sattler 和 A. Geiger，“Mip-Splatting：无混叠三维高斯泼溅”，载于 IEEE/CVF计算机视觉与模式识别会议论文集，2024年，第 19 447–19 456 页。[36] X. Zhang、J. Meng、R. Li、Z. Xu、Y. Zhang 和 J. Zhang，“GS-Hider：将消息隐藏到 3D 高斯泼溅中”，神经信息处理系统进展，第 37 卷，第 49 780–49 805 页，2024年。[37] Z. Li、Z. Chen、Z. Li 和 Y. Xu，“用于实时动态视图合成的时空高斯特征泼溅”，载于 IEEE/CVF计算机视觉与模式识别会议论文集，2024年，第 8508–8520 页。[38] J. Lee、M. Jeong、J. Park、Y. Hwang、S. Baek 和 W. H. Kim，“Mark4D：用于四维高斯泼溅的时间一致性水印”，载于 IEEE/CVF计算机视觉与模式识别会议论文集

，2026，第39页，393–39402。

[39] X. Zhao、K. Zhang、Z. Su、S. Vasan、I. Grishchenko、C. Kruegel、G. Vigna、Y.-X. Wang 和 L. Li，《生成式人工智能可证明地移除不可见图像水印》，神经信息处理系统进展，第37卷，第8643–8672页，2024年。[40] B. Mildenhall、P. P. Srinivasan、M. Tancik、J. T. Barron、R. Ramamoorthi 和 R. Ng，《NeRF：将场景表示为用于视图合成的神经辐射场》，*ACM*通讯，第65卷，106，2021年1，99期，第 – 页。[41] B. Mildenhall、P. P. Srinivasan、R. Ortiz-Cayon、N. K. Kalantari、R. Ramamoorthi、R. Ng 和 A. Kar，《局部光场融合：采用指导性采样准则的实用视图合成》，*ACM*图形学汇刊 *(ToG)*，第38卷，第4期，第1–14页，2019年。[42] A. Pumarola、E. Corona、G. Pons-Moll 和 F. Moreno-Noguer，《D-NeRF：动态场景的神经辐射场》，载于*IEEE/CVF*计算机视觉与模式识别会议论文集，2021年，第10 318–10 327页。[43] K. Park、U. Sinha、P. Hedman、J. T. Barron、S. Bouaziz、D. B. Goldman、R. Martin-Brualla 和 S. M. Seitz，《HyperNeRF：拓扑变化神经辐射场的高维表示》，*arXiv*预印本 *arXiv:2106.13228*，2021年。[44] Z. Wang、A. C. Bovik、H. R. Sheikh 和 E. P. Simoncelli，《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，第13卷，第4期，第600–612页，2004年。[45] R. Zhang、P. Isola、A. A. Efros、E. Shechtman 和 O. Wang，《深层特征作为感知度量的非凡有效性》，载于*IEEE*计算机视觉与模式识别会议论文集，2018年，第586–595页。