

额外 GS：几何感知的轨迹外推与不确定性引导的生成先验

Kaiyuan Tan^{1, 2} Yingying Shen² Haohui Zhu² Zhiwei Zhan² Shan Zhao²
Mingfei Tu² Hongcheng Luo² Haiyang Sun^{2†} Bing Wang²✉
Guang Chen² Hangjun Ye²
¹UIUC ²Xiaomi EV

Project Page: <https://xiaomi-research.github.io/extrags/>

Abstract

从记录的驾驶日志中合成外推视图对于模拟自动驾驶车辆的驾驶场景至关重要，然而这仍然是一个具有挑战性的任务。最近的方法利用生成先验作为伪真实值，但常常导致几何一致性差和过度平滑的渲染效果。为了解决这些限制，我们提出了 ExtraGS，这是一个将几何先验和生成先验整合起来的整体框架来实现轨迹外推。ExtraGS 的核心是一种基于混合高斯-符号距离函数(SDF)设计的新颖道路表面高斯(RSG)表示法，以及远场高斯(FFG)，使用可学习的缩放因子有效处理远处物体。此外，我们开发了一个基于球谐函数的自监督不确定性估计框架，该框架允许只在发生外推伪影的地方选择性地整合生成先验。在多个数据集、多样化的多摄像头设置和各种生成先验上的广泛实验表明，ExtraGS 显著提升了外推视图的真实感和几何一致性，同时保持了沿原始轨迹的高保真度。

1 介绍

随着端到端(E2E)自动驾驶方法的迅速发展，对可扩展、可控且无领域差距的闭环仿真环境的需求日益增长。

基于重建的方法(Yan et al. 2024; Chen et al. 2024; Yang et al. 2023a; Chen et al. 2023; Zhou et al. 2024b)，建立在体积渲染技术的最新进展之上(Mildenhall et al. 2020; Kerbl et al. 2023)，已经成功生成了具有极小领域差距的高保真驾驶场景模拟环境。然而，这些方法主要针对插值的新视角，并在外推轨迹(例如变道)上渲染时面临重大挑战。

最近，几种方法(Yu et al. 2024b; Zhao et al. 2024; Ni et al. 2024; Yan et al. 2025; Yang et al. 2024; Fan et al. 2024)提出了通过伪真实值来增强驾驶日志的外推视图，这些伪真实值是利用预训练的生成先验并由各种控制信号(例如，彩色激光雷达投影、噪声渲染图像、

参考图像和对象边界框)引导生成的。虽然这些方法显示了有希望的结果，但生成的伪真实值本质上仍然存在噪声，并且经常与原始轨迹数据不一致，导致模糊或扭曲的渲染效果。此外，这些生成方法需要对外推轨迹的先验知识，在实践中对于所有可能的未来路径而言计算上不可行。

这激励了两个关键问题：

如何通过更好的场景表示设计来提高外推性能？ 与生成方法不同，数据独立的方法更具通用性，并避免了昂贵的扩散模型训练。然而，这种方法仍然处于探索不足的状态。在场景元素中，我们确定了两个在外推时严重失真的关键元素。即路面和远处物体。

路面呈现独特的挑战，由于平行视角和稀疏纹理的存在，使得传统的体积渲染难以对其进行重建。在插值过程中，它们还经历了显著的视图方向变化。为了解决这些挑战，我们提出了道路表面高斯(RSG)模型：一种结合了维度减少的距离场用于几何形状以及扁平的3D高斯分布用于外观的混合模型。

另一方面，远处的物体跨越了极大的深度范围并且缺乏激光雷达覆盖，这使得直接优化效率低下且容易过拟合。为了解决这个问题，我们引入远场高斯(FFG)，通过可学习的比例因子共同调节位置和尺度，从而使远处物体能够快速收敛。

我们如何更好地利用生成先验？ 虽然生成模型可以提高外推质量，但它们无法区分由视角外推引起噪声与场景固有特征(例如复杂图案)之间的差异。这一根本限制导致要么对外推视图中的有意义场景组件进行过度平滑(图1)，要么保留不需要的伪影(图5)。核心挑战在于识别哪些像素是“不确定”的——这意味着它们包含由外推引起的噪声的可能性很高，生成模型应对此进行处理。为了解决这个问题，我们提出了一种自监督不确定性估计框架。我们的方法通过在现有视点相关颜色表示旁边存储一组额外的SH系数，无缝集成到3DGS

[†] 项目负责人，✉ 通讯作者。

框架中。我们的贡献可以总结如下：

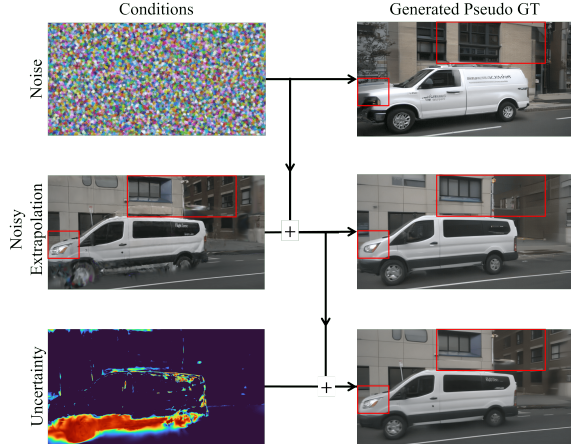


图 1: 不确定性引导生成的演示。顶部：纯噪声生成产生语义上不一致的结果。中间：条件设置改善了对齐，但由于统一的像素处理，失去了细节。底部：我们的不确定性引导方法区分了由外推引起的噪声和场景内在噪声，实现了有针对性的生成修正。

1. 我们介绍了两种新颖的几何表示形式以实现稳健的外推。道路表面高斯 (RSG) 结合了道路表面的先验知识，而远场高斯 (FFG) 则能够高效地优化远处的对象。
2. 我们提出了一种基于球谐函数的自监督不确定性估计框架，该框架明确区分了由外推引起的噪声和场景固有特征，从而能够精确且选择性地集成生成先验。
3. 通过在多个数据集、多摄像头配置和生成先验选择上的全面实验，我们证明了 ExtraGS 显著提升了外推视图的真实感和结构一致性。

2 相关工作

驾驶场景下的新颖视图合成 基于体积渲染的进步 (Mildenhall et al. 2020; Kerbl et al. 2023)，街景合成近年来受到了广泛关注。诸如 SUDS (Turki et al. 2023)、EmerNeRF (Yang et al. 2023a) 和 PVG (Chen et al. 2023) 等方法通过统一的表示隐式处理动态对象，其中静态和动态分离在训练过程中出现。尽管优雅，它们将动态对象视为未分化的整体，限制了模拟应用中的控制能力。显式的分解方法使用物体级姿态分别建模动态演员，从而实现操作和编辑。早期基于神经场景图的工作 (Xie et al. 2023; Wu et al. 2023; Yang et al. 2023b; Tonderski et al. 2023) 为每个物体和背景训练单独的 NeRFs。最近的方法 (Zhou et al. 2024b; Yan et al. 2024; Zhou et al. 2024a; Yu et al. 2024b; Chen et al.

2024; Hess et al. 2024) 利用 3D 高斯散射的真实能力形成高斯场景图，在实时渲染中实现了细粒度控制。虽然这些方法在插值视点处实现越来越高的渲染质量，但在超出原始训练轨迹外进行外推时常难以保持保真度，这对于下游应用如闭环模拟至关重要。

视图外推与生成先验 最近，基于扩散的生成模型在视图外推任务中增强了视觉质量 (Liu, Zhou, and Huang 2024; Liu, Shao, and Lu 2024; Yu et al. 2024a)。就驾驶场景而言，SGD (Yu et al. 2024b) 领先于通过使用生成图像来增强稀疏视图输入。DriveDreamer4D (Zhao et al. 2024) 采用预训练的世界模型和亲属数据策略提供伪地面真相。ReconDreamer (Ni et al. 2024; Zhao et al. 2025) 进一步将世界模型微调为在线恢复网络。FreeVS (Wang et al. 2024)、StreetCrafter 和 DriveX (Yan et al. 2025; Yang et al. 2024) 使用彩色激光雷达投影作为基于扩散生成的条件。

相比之下，我们的方法采取了不同的视角。我们不专注于设计特定的生成先验或条件信号，而是探索场景表示的设计空间以实现更强大的外推能力——无论是否使用生成先验。此外，尽管先前的工作主要集中在前视图或单摄像头外推上，我们的方法明确针对更具挑战性的多视图外推设置。

3 方法

问题表述 给定一个包含图像 $\{I_i\}_{i=1}^N$ 的驾驶日志、自我姿态 $\{P\}_{i=1}^N$ 以及一组相机 $\{K_j\}_{j=1}^C$ ，我们的目标是学习一种 4D 表示形式，该表示将时间、姿态和相机映射到图像 $\mathcal{S} : \mathbb{R} \times SE(3) \times C \rightarrow \mathbb{R}^{H \times W \times 3}$ 。理想情况下，这种表示方法能够从任意视角和时间点进行渲染——包括超出原始轨迹的那些。

3.1 几何感知动态场景建模

遵循 NSG (Ost et al. 2021)，我们将动态驾驶场景表示为一个场景图，其中每个节点建模具有不同几何属性的元素（例如，无限远的天空、刚性的车辆）。为了从给定视点渲染一个场景，根据每个节点的姿态独立地进行渲染，并通过 alpha 混合将结果组合在一起。

在本节中，我们介绍了两种新型节点，称为路面高斯分布 (RSG) 和远场高斯函数 (FFG)，进入场景图库存。完整的总体场景图结构和其他节点类型的描述见补充材料。

3.2 道路表面高斯分布

重建道路表面面临着独特的挑战，包括低纹理区域和视图方向的剧烈变化，这些通常会在外推视角下导致

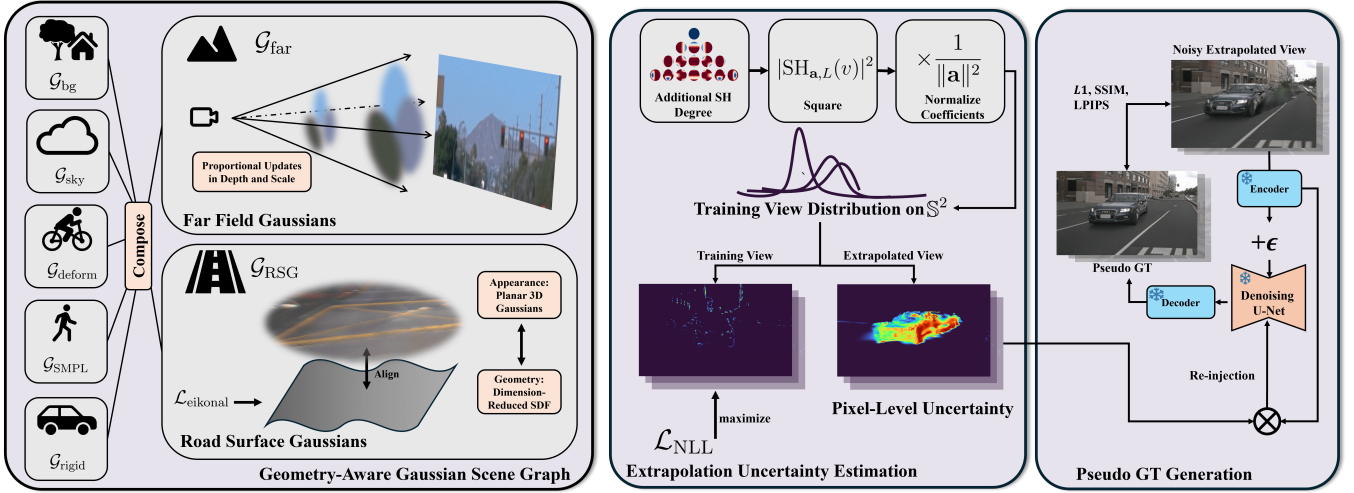


图 2: ExtraGS 概览: 左: ExtraGS 引入了两个额外的节点到场景图结构中: 道路表面高斯 (RSG), 它结合了 3D 高斯与降维 SDF 用于道路表面建模, 以及远场高斯 (FFG) 通过调整对齐的深度尺度实现远处物体的快速收敛。中间: ExtraGS 区分由外推引起的噪声和场景固有的噪声, 从而能够选择性地生成不可靠区域的改进。正确: 我们利用现成的生成模型在扩展视图中生成伪真实值。估计的不确定性为生成先验提供了像素级控制。

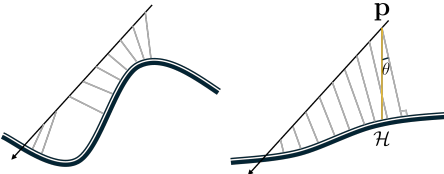


图 3: 几何动机维度缩减 SDF: 灰色线段显示了光线行进过程中到表面的垂直距离。左边: 一般表面。正确: 具有局部平面结构的道路表面, 产生几乎平行的距离向量。这使我们的二维公式 $d(\mathbf{p}) = |\cos \theta| (p_z - \mathcal{H}(p_x, p_y))$ 成为可能, 其中 $|\cos \theta|$ 将垂直距离转换为垂直距离。

严重的失真。为了解决这些问题, 我们提出引入强大的几何先验知识。具体来说, 我们认为道路表面是**局部平面的**并且表现出**全局平滑高程过渡**。这些假设与人类感知预期相符, 并推动了我们的降维 SDF 公式化。为了保持快速渲染速度, 我们将一组 3D 高斯分布与学习到的 SDF 对齐, 以高效表示外观。

平面先验的 SDF 我们提出了一种降维 SDF, 明确利用了路面局部平面结构。我们的关键见解是, 尽管道路几何形状可能在全球范围内弯曲, 但在光线的局部邻域内可以很好地近似为平面。这使得使用仅二维水平坐标来参数化有符号距离场成为可能 (图 3)。设 $\mathbf{p} = (p_x, p_y, p_z)$ 为一个三维点, 在该坐标系中, z 轴向上。在局部平面假设下, 有符号的距离定义为:

$$d(\mathbf{p}) = |\cos \theta| (p_z - \mathcal{H}(p_x, p_y)) \quad (1)$$

其中 $\mathcal{H} : \mathbb{R}^2 \rightarrow \mathbb{R}$ 表示水平坐标 (p_x, p_y) 处的局部表面高度, 而 θ 是表面法线与垂直轴之间的角度。 $\mathcal{H}$ 和 $|\cos \theta|$ 仅依赖于 (p_x, p_y) , 并使用神经网络进行参数化。这导致了以下 2D SDF 公式:

$$\begin{cases} d(\mathbf{p}) = \text{MLP}_{\text{slope}}(\mathbf{f}) [p_z - \text{MLP}_{\text{elevation}}(\mathbf{f})] \\ \mathbf{c}(\mathbf{p}, \mathbf{v}) = \text{MLP}_{\text{color}}(\mathbf{f}, \mathcal{F}(\mathbf{v})) \end{cases} \quad (2)$$

其中, $\mathbf{f} = \gamma(p_x, p_y)$ 是从多级哈希网格 (Li et al. 2023) 中提取的已学习特征, 而 $\mathcal{F}$ 表示视角方向 $\mathbf{v} \in S^2$ 的傅里叶嵌入 (Tancik et al. 2020)。函数 $\text{MLP}_{\text{slope}}$ 、 $\text{MLP}_{\text{elevation}}$ 和 $\text{MLP}_{\text{color}}$ 分别预测表面坡度、高程和视点相关颜色。我们的 2D SDF 公式可以无缝集成到从 (Wang et al. 2021) 的体积渲染框架中, 以生成渲染图像 $\hat{\mathbf{C}}$ 。

道路表面高斯分布 为了利用 3D 高斯渲染的高速度, 我们引入一组与学习到的 SDF 对齐的高斯 $\mathcal{G}_{\text{RSG}}$ 。每个高斯位于 $\boldsymbol{\mu} = (\mu_x, \mu_y, \mathcal{H}(\mu_x, \mu_y))$, 其中高度直接从 SDF 查询。由于假设道路表面局部是平面的, 因此每个高斯仅用两个比例维度定义。在训练期间, 高斯位置在每次迭代时重新计算, 无需梯度。仅优化这些高斯的 RGB 外观属性。

优化 在每次训练迭代中, 我们随机从驾驶日志中采样一张图像 $\mathbf{I} \in \mathbb{R}^{H \times W \times 3}$, 并使用预先计算的语义分割掩码从 $\mathbf{I}$ 中选择一批对应于路面的像素 $\mathbf{P} \in \mathbb{R}^{B \times 3}$ 。SDF 使用结合了 Eikonal 正则化 (Gropp et al. 2020) 的一个

$\mathcal{L}_1$ 重建损失进行优化，以强制执行平滑表面：

$$\mathcal{L}_{\text{sdf}} = \|\hat{\mathbf{P}} - \mathbf{P}\|_1 + \sum_{\mathbf{r} \in \mathbf{P} \sum_{i=1}^N (\|\nabla d(\mathbf{r}_i)\|_2 - 1)^2} \quad (3)$$

值得注意的是，在训练过程中 SDF 完全作为路面高斯分布的几何先验。在测试时，仅渲染高斯分布 $\mathcal{G}_{\text{RSG}}$ 。此设计使 RSG 能够从 SDF 表示的结构保真度中获益，同时保留 3D 高斯渲染的速度和灵活性。

3.3 远场高斯函数

当重建驾驶场景时，远处的物体跨越了非常不同的深度范围——从数百米外的建筑到数千米外的云层。这些远处的结构通常缺乏激光雷达覆盖，使得直接通过位置参数 μ 上的梯度下降进行 3D 高斯优化既低效又容易陷入局部最小值。虽然这样的次优解在训练轨迹上可能会产生可接受的结果，但在从外推视角渲染时会出现显著的伪影。

为了解决这一挑战，我们引入了一种新的高斯节点，称为远场高斯 $\mathcal{G}_{\text{FFG}}$ ，它同时调节每个高斯基元的位置和尺度（图 2）。具体来说，我们在每个高斯函数中增加了一个可学习的缩放因子 f ，该因子同时转换位置 μ 和尺度 s 。

$$\begin{cases} \mu' = \exp(f) \cdot \mu \\ s' = \exp(f) \cdot s \end{cases} \quad (4)$$

缩放因子 f 在训练期间与所有其他高斯参数联合优化。

直观上，这种表述将相对比例调整与绝对定位解耦，使得 3D 高斯函数能够快速适应表示远处的结构，而无需进行可能病态的位置更新。

3.4 外推不确定性估计

近期的研究表明，将生成先验作为伪地面真相可以显著提高视图外推性能。然而，一个关键的限制是生成模型无法区分由视点外推引起的噪声和场景固有特征之间的区别。这导致了有意义的组件被过度平滑（图 1），或者在外推视图中保留冗余伪影（图 4）。这促使需要量化哪些像素是“不确定”：意味着由外推引起的噪声概率很高。我们通过一个自我监督的不确定性估计框架来解决这个问题，该框架能够识别出可能被外推噪声损坏的像素。

对于外推视图中的每个像素，我们定义不确定性为其观察光线 $\mathbf{r}$ 与驾驶日志中训练观察光线对齐的**非可能性**。这可以形式化为一个核密度估计 (KDE) 问题。对于每个高斯分布 g ，我们希望量化训练视图 $v \in \mathbb{S}^2$ 的分

布 $p_g(v)$ 。我们定义此分布为：

$$p_g(v) := \frac{1}{N} \sum_{i=1}^N K(v, v_i) \approx \frac{|\text{SH}_{\mathbf{a}, L}(v)|^2}{|\mathbf{a}|^2} \quad (5)$$

其中 v_i 是与 g 相关的训练视图方向，而 K 是一种方向相似性测量（例如余弦相似度）。

为了避免显式存储所有训练视图方向，我们使用球谐函数 (SH) 近似 $p_g(v)$ 。因此，方程 5 被进一步简化为 $\frac{|\text{SH}_{\mathbf{a}, L}(v)|^2}{|\mathbf{a}|^2}$ ，其中 $\mathbf{a} \in \mathbb{R}^{(L+1)^2}$ 是可学习的 SH 系数， L 是 SH 阶数。这源于球谐函数在表示球面上平方可积函数方面的完备性。正式证明见补充材料。

在对每高斯训练视图分布进行建模后，不确定性图被渲染为： $U \in \mathbb{R}^{1 \times H \times W} = 1 - \sum_{i=1}^N \prod_{j=1}^{i-1} (1 - \alpha_i) \alpha_i p_g$ ，并通过对数似然损失进行优化：

$$\mathcal{L}_{\text{NLL}} = -\frac{1}{HW} \sum_{i,j} \log(1 - U_{i,j}) \quad (6)$$

该过程如图 2 所示。由于 3DGS 管道已经使用球谐函数来处理视图相关的颜色，我们的不确定性可以通过存储一组可学习的 SH 系数无缝集成到高斯散射框架中。

3.5 伪彩色编码

在实际操作中，我们观察到伪真实图像经常表现出过度饱和的颜色或视觉伪影。这些不一致导致了伪真实图像与渲染图像之间的颜色明显不符。我们引入了一个伪颜色编码层，该层学习颜色空间中的校正变换： $C_{\text{out}} = C_{\text{in}} + \text{MLP}_{\text{pseudo}}(C_{\text{in}})$ ，其中 $\text{MLP}_{\text{pseudo}}$ 预测残差颜色调整。这一层仅在对推断视图进行训练时应用，以减少对原始轨迹渲染的干扰。

3.6 损失函数

在实践中，为了保持每个高斯分布仅代表其期望的类别，我们采用与先前工作 (Chen et al. 2024; Yan et al. 2024) 类似的策略，利用预训练分割模型预测的语义道路和天空掩码，并应用交叉熵损失项 $\mathcal{L}_{\text{mask}}$ 。总损失定义为：

$$\mathcal{L} = (1 - \mathbb{1}_{\text{extrapolate}})(\mathcal{L}_1 + \mathcal{L}_{\text{SSIM}} + \mathcal{L}_{\text{LPIPS}} + \mathcal{L}_{\text{mask}} + \mathcal{L}_{\text{sdf}} + \mathcal{L}_{\text{NLL}}) + \mathbb{1}_{\text{extrapolate}} \mathcal{L}_{\text{ex}}$$

4 实现细节

为了展示 ExtraGS 的通用性，我们将两个现成的生成先验整合到我们的场景表示中。

损失项之间的相对权重在此为了简洁起见被省略。

额外 GS-M 首先采用 MagicDriveV2 (Gao et al. 2025) 作为生成先验。MagicDriveV2 是一个基于流匹配的视频生成模型。我们利用其流匹配模型的去噪过程来生成伪真实值，其中低不确定性像素逐渐被“重新注入”以保留高保真细节。这使得能够有选择性地仅在最有益的地方使用生成先验。生成伪代码显示在补充材料中。

额外 GS-D 我们使用基于 Stable Diffusion Turbo 的 Difx3D (Wu et al. 2025) 作为我们的生成先验。由于 Difx3D 使用单步去噪，我们通过根据 $\alpha' = u \cdot \alpha$ 缩放高斯不透明度来提供像素级不确定性控制，有效地移除被遮挡的对象和不需要的伪影。

实验设置 我们遵循 OmniRe(Chen et al. 2024) 进行数据处理和高斯场景图的实现。在我们的所有实验中，首先预训练模型 30,000 次迭代，然后在第 30,000 次和第 35,000 次迭代时分别通过横向移动 1.5 米和 3 米引入伪真实值，总计 40,000 次迭代。损失系数 $\lambda_1, \lambda_{SSIM}, \lambda_{LPIPS}, \lambda_{mask}, \lambda_{sdf}, \lambda_{NLL}, \lambda_{ex}$ 分别设置为 0.8, 0.2, 0.05, 0.5, 0.5, 1.0。训练在单个 NVIDIA H20 GPU 上大约需要 3 小时。

5 实验

数据集 我们在两个大规模的自动驾驶基准数据集上评估了我们方法的外推性能：nuScenes 和 Waymo 开放数据集 (WOD) (Caesar et al. 2020; Ettinger et al. 2021)。对于 nuScenes，所有实验均在数据集原生分辨率 900×1600 像素下进行，使用全部六个环绕视角摄像头对 10 秒片段进行了处理。对于 WOD，我们遵循了 (Yan et al. 2025) 中的协议，将图像缩小到 1066×1600 分辨率。我们选择大约 4 秒钟长的场景片段以便与之前的工作进行比较 (Zhao et al. 2024)。额外的实验细节在补充材料中。

基线 我们将 ExtraGS-D 与包含生成式外推的最先进的方法进行了比较：DriveDreamer4D (Zhao et al. 2024)，ReconDreamer (Ni et al. 2024)，ReconDreamer++ (Zhao et al. 2025)，FreeVS (Wang et al. 2024)，DriveX (Yang et al. 2024)，StreetCrafter (Yan et al. 2025) 和 Difx3D+。为了进一步展示 ExtraGS 在生成先验选择和多摄像头设置中的通用性，我们将 ExtraGS-D 与 EmerNeRF (Yang et al. 2023a)、PVG (Chen et al. 2023)、Street Gaussian (Yan et al. 2024) 和 OmniRe (Chen et al. 2024) 在 nuScenes 数据集上进行比较。

Method	Gen. Condition	NTA-IoU $\uparrow$	NTL-IoU $\uparrow$
DriveDreamer4D	3D box + HD Map	0.457	53.30
ReconDreamer	3D box + HD Map	0.539	54.58
ReconDreamer++	3D box + HD Map	0.572	57.06
FreeVS	LiDAR Projection	0.505	56.84
DriveX	LiDAR Projection	0.567	<u>58.29</u>
StreetCrafter	LiDAR Projection	<u>0.582</u>	57.15
Difix3D+OmniRe	无	0.572	53.86
ExtraGS-D	无	0.592	58.49

表 1: 在 Waymo 开放数据集上的下游任务性能。ExtraGS-D 在不需要额外的条件输入的情况下实现了最先进的结果。

5.1 Waymo 开放数据集上的下游任务

我们通过 Waymo 开放数据集上的下游感知任务评估 ExtraGS-D 的实际效用，并将其与最先进的轨迹外推方法进行比较。结果如表 1 和图 4 所示。

评估协议 我们采用 DriveDreamer4D 中定义的 NTA-IoU 和 NTL-IoU 分数及其官方实现。NTA-IoU 通过计算 YOLO11 (Jocher and Qiu 2024) 预测与地面真实 3D 边界框投影之间的 IoU 来衡量目标检测性能。NTL-IoU 则通过 TwinLiteNet (Che et al. 2023) 预测与地面真实高清地图之间的 IoU 来评估车道检测性能。所有外推轨迹都向侧方移动了 3 米，以模拟车道变换场景。

结果 我们对比了七种最先进的方法，这些方法具有不同的条件要求。DriveDreamer4D、ReconDreamer、FreeVS 和 DriveX 的结果直接取自 (Ni et al. 2024; Zhao et al. 2025; Yang et al. 2024)。我们使用 StreetCrafter 的官方实现，并将 Difx3D 与 OmniRe 集成以处理动态对象。

ExtraGS-D 达到了最先进的性能，NTA-IoU 为 0.592，NTL-IoU 为 58.49，超越了所有竞争方法。值得注意的是，ExtraGS-D 不需要额外的条件输入。与 Difx3D+OmniRe (NTA-IoU: 0.572, NTL-IoU: 53.86) 进行比较进一步验证了 ExtraGS 的有效性，因为使用相同的生成先验 (Difix3D) 但不同的场景表示产生了明显较差的结果。



图 4: 不同方法在 Waymo 开放数据集上的定性比较. DriveDreamer4D、ReconDreamer 和 DriveX 的结果分别来自它们的官方可视化或根据所提供的代码和权重渲染而来。

		外推法			原始轨迹		
	FID @ 3m↓	mAP _{obj} ↑	mAP _{map} ↑	PSNR↑	SSIM↑	LPIPS↓	FID↓
EmerNerf	104.49	0.137	0.217	28.82	0.826	0.390	51.75
PVG	138.34	0.319	0.164	28.95	0.846	0.332	60.04
StreetGS	95.29	0.450	0.321	29.64	0.878	0.240	26.85
OmniRe	93.60	0.419	0.308	29.70	0.879	0.236	26.74
ExtraGS-M	77.19	0.513	0.384	29.72	0.874	0.242	26.30

表 2: 系统级比较在 nuScenes 数据集上。结果以所有 6 个摄像头的 900x1600 分辨率报告。我们的方法在保持原始轨迹的视觉质量的同时实现了增强的外推结果。

图 4展示了 ExtraGS-D 在多种场景下的优越视觉质量。ExtraGS-D 一致生成清晰的车道标记，同时保持与原始轨迹的一致性，而竞争方法通常会产生模糊或过饱和的车道标记。

5.2 系统级比较在 nuScenes 数据集上

为进一步证明 ExtraGS 的通用性，我们展示了使用预训练视频生成模型 MagicDriveDiT 作为生成先验 (ExtraGS-M) 的结果。结果见表 2。

评估协议 我们从两个类别评估性能：使用 Fréchet Inception Distance(FID) 进行外推质量评估，通过多摄像头 3D 物体检测 (mAP_{obj}) 和 HD 地图预测 (mAP_{map}) 评估下游任务准确性，并使用标准指标 (PSNR, SSIM 和 LPIPS) 报告原始轨迹的保真度结果。对于下游评估，

我们采用预训练的 BEVFormer V2 (Yang et al. 2022) 和 MapTR V2 (Liao et al. 2024) 来预测外推视图。这些预测被移回原始轨迹进行评估。我们报告 3D 物体检测（专注于“汽车”类别）的平均精度 (AP) 以及在 0.5 米、1.0 米和 1.5 米的 chamfer 距离阈值下的地图预测的平均精度 (mAP)。

分析 表 2显示 ExtraGS-M 在外推性能方面表现出色，具有最低的 FID@3m (77.19) 和最高的下游任务得分 (mAP_{obj} : 0.513, mAP_{map} : 0.384)，显著优于所有基线。至关重要的是，这些改进并未损害原始轨迹的质量—ExtraGS-M 保持了有竞争力的 PSNR (29.72) 和 FID (26.30) 分数。

5.3 消融研究



图 5: 关键组件的 ExtraGS 可视化消融研究

Method	PSNR $\uparrow$	SSIM $\uparrow$	FID $\downarrow$ @ 3m
3DGS(Kerbl et al. 2023)	29.22	0.862	90.67
2DGS(Huang et al. 2024)	24.67	0.758	<u>75.51</u>
SDF(Guo et al. 2023)	27.81	0.741	77.08
Ours	<u>27.92</u>	<u>0.812</u>	67.72

表 3: 道路表面表示的消融研究

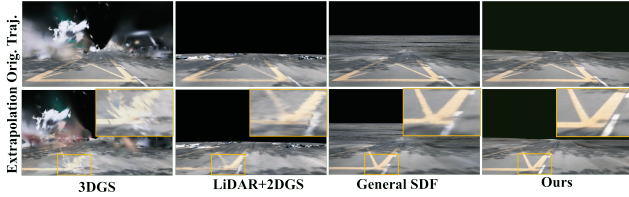


图 6: 不同路面表示的定性消融研究

路面表示方法 我们研究了不同的道路表面建模策略没有生成先验 (表 3和图 6)。我们将 3DGS、2DGS 与 LiDAR 初始化 (Huang et al. 2024)、基于 StreetSurf 的 SDF (Guo et al. 2023) 以及我们的 RSG 表示进行了比较。我们使用在预处理语义道路掩码指示的可行驶区域上计算得到的 PSNR 和 SSIM 来评估重建质量, 并通过在 3 米横向偏移下的 FID 来评估外推时的感知一致性。

结果揭示了重建保真度与外推质量之间的基本权衡。虽然 3DGS 达到了最高的 PSNR (29.22) 和 SSIM (0.862), 但它未能捕捉到准确的路面几何形状, 导致外推视图质量较差 (FID: 90.67)。这一限制源于 3DGS 无法对路面施加几何约束。相比之下, 我们提出的 RSG 模型实现了最优平衡, 在保持具有竞争力的重建质量的同时, 达到了最低的外推 FID (67.72), 展示了有效的路面几何建模。

	PSNR $\uparrow$	FID (before) $\downarrow$	FID (after) $\downarrow$
Ours w/o FFG	31.15	147.85	87.19
Ours w/o RSG	31.95	138.92	84.74
Ours w/o Uncertainty	31.94	—	87.47
Ours w/o Pseudo Color Encod.	30.86	—	85.20
Ours Full	32.31	128.36	81.03

表 4: 关键组件的消融研究我们报告了原始轨迹中的 PSNR 以及应用伪真实值前后的外推 FID。

成分分析 为了解每个组件如何影响最终的外推质量, 我们在表 4和图 5中提供了全面的消融研究。

- **路面高斯分布 (RSG)**: 移除 RSG 会导致明显的性能下降, 即使在与生成先验的情况下 (FID 从 81.03 增加到 84.74)。图 5(第 4 列) 中的视觉结果揭示了车道线标记的显著扭曲, 证明了几何先验和生成先验在轨迹外推中的互补作用。
- **不确定性估计**: 缺乏不确定性指导导致原始轨迹质量和外推一致性 (FID: 87.47) 均下降。图 5 (第 2 列) 显示, 没有不确定性时, 在原轨迹被遮挡的区域会出现伪影。
- **远场高斯函数 (FFG)**: 移除 FFG 节点显著影响原始轨迹 PSNR 和外推 FID。此外, 这种退化问题无法通过生成先验来解决, 如表 4 (第 3 列) 和图 5 所示。
- **伪彩色编码**: 去除伪彩色编码导致原始轨迹的 PSNR 显著下降 (从 32.31 降至 30.86)。我们认为这种退化现象是由于生成图像中的色彩偏差造成的。

6 结论

我们提出了 ExtraGS, 一个集成几何和生成先验的整体框架, 用于外推视图合成。在 NSG 范式内构建的

ExtraGS 引入了两个高斯节点：道路表面高斯 (RSG) 结合 3D 高斯与维度减少的 SDF 进行道路表面建模，以及远场高斯 (FFG) 通过对齐深度尺度调整实现远处物体的快速收敛。此外，ExtraGS 区分由外推引起的噪声和场景固有的噪声，使不可靠区域的选择性生成细化成为可能。广泛的实验表明，在外推视图中具有优越的感知质量和结构一致性。

参考文献

- Caesar, H.; Bankiti, V.; Lang, A. H.; Vora, S.; Liong, V. E.; Xu, Q.; Krishnan, A.; Pan, Y.; Baldan, G.; and Beijbom, O. 2020. nuScenes: A multimodal dataset for autonomous driving. In CVPR.
- Che, Q.-H.; Nguyen, D.-P.; Pham, M.-Q.; and Lam, D.-K. 2023. TwinLiteNet: An Efficient and Lightweight Model for Driveable Area and Lane Segmentation in Self-Driving Cars. In 2023 International Conference on Multimedia Analysis and Pattern Recognition (MAPR), 1–6.
- Chen, Y.; Gu, C.; Jiang, J.; Zhu, X.; and Zhang, L. 2023. Periodic Vibration Gaussian: Dynamic Urban Scene Reconstruction and Real-time Rendering. arXiv:2311.18561.
- Chen, Z.; Yang, J.; Huang, J.; de Lutio, R.; Esturo, J. M.; Ivanovic, B.; Litany, O.; Gojcic, Z.; Fidler, S.; Pavone, M.; Song, L.; and Wang, Y. 2024. OmniRe: Omni Urban Scene Reconstruction. arXiv preprint arXiv:2408.16760.
- Ettinger, S.; Cheng, S.; Caine, B.; Liu, C.; Zhao, H.; Pradhan, S.; Chai, Y.; Sapp, B.; Qi, C. R.; Zhou, Y.; Yang, Z.; Chouard, A.; Sun, P.; Ngiam, J.; Vasudevan, V.; McCauley, A.; Shlens, J.; and Anguelov, D. 2021. Large Scale Interactive Motion Forecasting for Autonomous Driving: The Waymo Open Motion Dataset. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 9710–9719.
- Fan, L.; Zhang, H.; Wang, Q.; Li, H.; and Zhang, Z. 2024. FreeSim: Toward Free-viewpoint Camera Simulation in Driving Scenes. arXiv:2412.03566.
- Gao, R.; Chen, K.; Xiao, B.; Hong, L.; Li, Z.; and Xu, Q. 2025. MagicDrive-V2: High-Resolution Long Video Generation for Autonomous Driving with Adaptive Control. In Proceedings of the IEEE/CVF International Conference on Computer Vision.
- Gropp, A.; Yariv, L.; Haim, N.; Atzmon, M.; and Lipman, Y. 2020. Implicit Geometric Regularization for Learning Shapes. In Proceedings of Machine Learning and Systems 2020, 3569–3579.
- Guo, J.; Deng, N.; Li, X.; Bai, Y.; Shi, B.; Wang, C.; Ding, C.; Wang, D.; and Li, Y. 2023. StreetSurf: Extending Multi-view Implicit Surface Reconstruction to Street Views. arXiv preprint arXiv:2306.04988.
- Hess, G.; Lindström, C.; Fatemi, M.; Petersson, C.; and Svensson, L. 2024. SplatAD: Real-Time Lidar and Camera Rendering with 3D Gaussian Splatting for Autonomous Driving. arXiv preprint arXiv:2411.16816.
- Huang, B.; Yu, Z.; Chen, A.; Geiger, A.; and Gao, S. 2024. 2D Gaussian Splatting for Geometrically Accurate Radiance Fields. In SIGGRAPH 2024 Conference Papers. Association for Computing Machinery.
- Jocher, G.; and Qiu, J. 2024. Ultralytics YOLO11.
- Kerbl, B.; Kopanas, G.; Leimkühler, T.; and Drettakis, G. 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. ACM Transactions on Graphics, 42(4).
- Li, Z.; Müller, T.; Evans, A.; Taylor, R. H.; Unberath, M.; Liu, M.-Y.; and Lin, C.-H. 2023. Neuralangelo: High-Fidelity Neural Surface Reconstruction. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
- Liao, B.; Chen, S.; Zhang, Y.; Jiang, B.; Zhang, Q.; Liu, W.; Huang, C.; and Wang, X. 2024. Maptrv2: An end-to-end framework for online vectorized hd map construction. International Journal of Computer Vision, 1–23.
- Liu, K.; Shao, L.; and Lu, S. 2024. Novel View Extrapolation with Video Diffusion Priors. arXiv preprint arXiv:2411.14208.
- Liu, X.; Zhou, C.; and Huang, S. 2024. 3DGS-Enhancer: Enhancing Unbounded 3D Gaussian Splatting with View-consistent 2D Diffusion Priors. In Globerson, A.; Mackey, L.; Belgrave, D.; Fan, A.; Paquet, U.; Tomczak, J.; and Zhang, C., eds., Advances in Neural Information Processing Systems, volume 37, 133305–133327. Curran Associates, Inc.
- Mildenhall, B.; Srinivasan, P. P.; Tancik, M.; Barron, J. T.; Ramamoorthi, R.; and Ng, R. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In ECCV.
- Ni, C.; Zhao, G.; Wang, X.; Zhu, Z.; Qin, W.; Huang, G.; Liu, C.; Chen, Y.; Wang, Y.; Zhang, X.; Zhan, Y.; Zhan, K.; Jia, P.; Lang, X.; Wang, X.; and Mei, W. 2024. ReconstructDreamer: Crafting World Models for Driving Scene Reconstruction via Online Restoration.
- Ost, J.; Mannan, F.; Thuerey, N.; Knodt, J.; and Heide, F. 2021. Neural Scene Graphs for Dynamic Scenes. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2856–2865.
- Tancik, M.; Srinivasan, P. P.; Mildenhall, B.; Fridovich-Keil, S.; Raghavan, N.; Singhal, U.; Ramamoorthi, R.; Barron, J. T.; and Ng, R. 2020. Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains. arXiv:2006.10739.
- Tonderski, A.; Lindström, C.; Hess, G.; Ljungbergh, W.; Svensson, L.; and Petersson, C. 2023. NeuRAD: Neural Rendering for Autonomous Driving. arXiv preprint arXiv:2311.15260.
- Turki, H.; Zhang, J. Y.; Ferroni, F.; and Ramanan, D. 2023. SUDS: Scalable Urban Dynamic Scenes. In Computer Vision and Pattern Recognition (CVPR).

- Wang, P.; Liu, L.; Liu, Y.; Theobalt, C.; Komura, T.; and Wang, W. 2021. NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction. *arXiv preprint arXiv:2106.10689*.
- Wang, Q.; Fan, L.; Wang, Y.; Chen, Y.; and Zhang, Z. 2024. FreeVS: Generative View Synthesis on Free Driving Trajectory. *arXiv:2410.18079*.
- Wu, J. Z.; Zhang, Y.; Turki, H.; Ren, X.; Gao, J.; Shou, M. Z.; Fidler, S.; Gojcic, Z.; and Ling, H. 2025. DIFIX3D+: Improving 3D Reconstructions with Single-Step Diffusion Models. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 26024–26035.
- Wu, Z.; Liu, T.; Luo, L.; Zhong, Z.; Chen, J.; Xiao, H.; Hou, C.; Lou, H.; Chen, Y.; Yang, R.; Huang, Y.; Ye, X.; Yan, Z.; Shi, Y.; Liao, Y.; and Zhao, H. 2023. MARS: An Instance-aware, Modular and Realistic Simulator for Autonomous Driving. *CICAI*.
- Xie, Z.; Zhang, J.; Li, W.; Zhang, F.; and Zhang, L. 2023. S-NeRF: Neural Radiance Fields for Street Views. In *International Conference on Learning Representations (ICLR)*.
- Yan, Y.; Lin, H.; Zhou, C.; Wang, W.; Sun, H.; Zhan, K.; Lang, X.; Zhou, X.; and Peng, S. 2024. Street Gaussians: Modeling Dynamic Urban Scenes with Gaussian Splatting. In *ECCV*.
- Yan, Y.; Xu, Z.; Lin, H.; Jin, H.; Guo, H.; Wang, Y.; Zhan, K.; Lang, X.; Bao, H.; Zhou, X.; and Peng, S. 2025. StreetCrafter: Street View Synthesis with Controllable Video Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Yang, C.; Chen, Y.; Tian, H.; Tao, C.; Zhu, X.; Zhang, Z.; Huang, G.; Li, H.; Qiao, Y.; Lu, L.; Zhou, J.; and Dai, J. 2022. BEVFormer v2: Adapting Modern Image Backbones to Bird’s-Eye-View Recognition via Perspective Supervision. *ArXiv*.
- Yang, J.; Ivanovic, B.; Litany, O.; Weng, X.; Kim, S. W.; Li, B.; Che, T.; Xu, D.; Fidler, S.; Pavone, M.; and Wang, Y. 2023a. EmerNeRF: Emergent Spatial-Temporal Scene Decomposition via Self-Supervision. *arXiv:2311.02077*.
- Yang, Z.; Chen, Y.; Wang, J.; Manivasagam, S.; Ma, W.-C.; Yang, A. J.; and Urtasun, R. 2023b. UniSim: A Neural Closed-Loop Sensor Simulator. In *CVPR*.
- Yang, Z.; Pan, Z.; Yang, Y.; Zhu, X.; and Zhang, L. 2024. Driving Scene Synthesis on Free-form Trajectories with Generative Prior. *arXiv preprint arXiv:2412.01717*.
- Yu, W.; Xing, J.; Yuan, L.; Hu, W.; Li, X.; Huang, Z.; Gao, X.; Wong, T.-T.; Shan, Y.; and Tian, Y. 2024a. ViewCrafter: Taming Video Diffusion Models for High-fidelity Novel View Synthesis. *arXiv preprint arXiv:2409.02048*.
- Yu, Z.; Wang, H.; Yang, J.; Wang, H.; Xie, Z.; Cai, Y.; Cao, J.; Ji, Z.; and Sun, M. 2024b. SGD: Street View Synthesis with Gaussian Splatting and Diffusion Prior. *arXiv:2403.20079*.
- Zhao, G.; Ni, C.; Wang, X.; Zhu, Z.; Zhang, X.; Wang, Y.; Huang, G.; Chen, X.; Wang, B.; Zhang, Y.; Mei, W.; and Wang, X. 2024. DriveDreamer4D: World Models Are Effective Data Machines for 4D Driving Scene Representation.
- Zhao, G.; Wang, X.; Ni, C.; Zhu, Z.; Qin, W.; Huang, G.; and Wang, X. 2025. ReconDreamer++: Harmonizing Generative and Reconstructive Models for Driving Scene Representation.
- Zhou, H.; Shao, J.; Xu, L.; Bai, D.; Qiu, W.; Liu, B.; Wang, Y.; Geiger, A.; and Liao, Y. 2024a. HUGS: Holistic Urban 3D Scene Understanding via Gaussian Splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 21336–21345.
- Zhou, X.; Lin, Z.; Shan, X.; Wang, Y.; Sun, D.; and Yang, M.-H. 2024b. Drivinggaussian: Composite gaussian splatting for surrounding dynamic autonomous driving scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 21634–21643.