使用概率普罗克鲁斯特映射的非定位 3DGS 重建

Chong Cheng* **Zijian Wang*** **Sicheng Yu** **Yu Hu** **Nanjie Yao** **Hao Wang†**
The Hong Kong University of Science and Technology (Guangzhou)
{ ccheng735, zwang886, yhu847 } @connect.hkust-gz.edu.cn
yusch@mail2.sysu.edu.cn nanjiey@uci.edu haowang@hkust-gz.edu.cn

Abstract

3D 高斯撒点（3DGS）已经成为 3D 表示的核心技术。它的有效性很大程度上依赖于精确的相机位姿和准确的点云初始化，这些通常是从预训练的多视图立体（MVS）模型中得出的。然而，在无姿态重构任务中，对于数百张户外图像，现有的 MVS 模型可能会因内存限制而表现不佳，并且随着输入图像数量的增加准确性下降。为了解决这一限制，我们提出了一种新颖的无姿态 3DGS 重构框架，将预训练的 MVS 先验与概率的普鲁克鲁斯特映射策略相结合。该方法将输入图像划分为子集，将子映射映射到全局空间，并利用 3DGS 对几何和位姿进行联合优化。从技术上讲，我们将数千万点云的映射表述为一个概率的普鲁克鲁斯特问题，并解决了一个封闭形式的对齐问题。通过使用概率耦合以及软垃圾箱机制以拒绝不确定的对应关系，我们的方法在几分钟内在数百张图像中全局对齐点云和位姿。此外，我们还提出了一种 3DGS 和相机位姿的联合优化框架。该框架从置信度感知的锚点构建高斯，并将 3DGS 可微渲染与解析雅可比相结合，以共同优化场景和位姿，能够实现精确的重构和位姿估计。在 Waymo 和 KITTI 数据集上的实验表明，我们的方法在无姿态图像序列中实现了精确重构，为无姿态 3DGS 重构设定了新的技术标准。

1 介绍

3D 高斯分布（3DGS）由于其出色的渲染质量和实时性能，已经成为一种革命性的三维表示和新视图合成技术 (**??**) 。通过优化一组 3D 高斯参数来表示场景，3DGS 实现了高保真和高效的视觉效果，并迅速成为研究的焦点。

然而，将 3DGS 应用于真实场景，尤其是从数百个未经校准的户外图像进行重建，仍然非常具有挑战性。传统的 3DGS 流程严重依赖于精确预计算的相机姿态和初始点云 (**??**) 。这些前提条件在复杂的户外环境中通常难以获得，这极大地限制了 3DGS 的广泛适用性。

一些研究尝试以端到端的方式从图像中联合优化相机位姿和高斯参数，从而实现无拍摄姿态的 3DGS 重建。然而，由于尺度歧义、稀疏监督和对噪声初始化的敏感性，它们在室外场景中表现困难，通常导致精度有限。另一种常见策略是将运动结构（SfM）与 3DGS 结合，但 SfM 阶段通常计算量大，经常需要几个小时的处理时间，并且在具有挑战性的室外条件下容易失败。预训练多视图立体（MVS）模型 (**??**) 长期以来一直作为一种结构化的方法来直接从图像推断密集点云和相机姿态，并仍然是无姿态 3D 重建的一个有前途的基础。相比之下，前馈 3DGS 方法直接从图像中预测高斯分布，具有更高的效率，但通常仅支持十几个视图，并容易出现内存不足（OOM）问题 **????** 。虽然现代 MVS 模型可以处理更大的输入批次，但随着视图数量的增加，尤其是在室外场景中，它们仍然面临准确性下降和内存瓶颈的问题 (**??**) 。这些挑战激发了一种分而治之的策略：将大型图像集分解为较小的子集，单独处理它们，并将其合并成一个全局一致的重建。然而，由于每个子图是其自身局部框架中推断出的，结果通常会因尺度不确定性和几何不一致而受影响。现有的配准方法 (**????**) 通

*Equal contribution.
†Corresponding author.

Preprint. Under review.

常在尺度不明确、几何偏差以及对齐数千万点的计算挑战下失败。因此，一个关键的挑战是如何高效地将这些子图对齐到一个统一的坐标系统，以实现高质量的 3DGS 重建。为了解决这些挑战，我们提出了一个整合预训练 MVS 与分而治之策略的协同框架，用于无姿态 3DGS 重建。利用前馈先验和跨图像组的重叠视图，我们逐步从局部子图中恢复出全局一致的点云和相机姿态，从而实现高质量的 3DGS 重建。具体来说，我们通过在像素级设计重叠帧的对应关系，将原始的数千万点的对齐重构为一个概率 Procrustes 问题。我们首先使用 Kabsch-Umeyama 算法获得一个闭形式的 $Sim(3)$ 解，然后通过与一个柔性的垃圾箱机制进行概率耦合，细化该解，以拒绝不确定的匹配。这种方法有效地解决了子地图之间的尺度模糊和局部几何差异，在几分钟内实现了稳健的全局对齐。

此外，我们提出了一种用于 3DGS 和相机姿态的联合优化框架，其中高斯从通过置信度感知的对应关系过滤获得的下采样锚点初始化。相机姿态通过可微分的 3DGS 渲染进行优化，并通过解析四元数雅可比矩阵传播梯度，从而提高了姿态准确性和视图合成质量。

Our main contributions are as follows:

1. 我们提出了一种将子地图映射作为概率性 Procrustes 问题来处理的对齐方法。它结合了闭式 $Sim(3)$ 估计与概率性和异常值拒绝，使得可以在几分钟内从数百张图像中恢复全局姿态和点云。
2. 我们提出了一个 3DGS 和姿态联合优化模块，该模块从置信度引导的锚点构建高斯，通过具有解析雅可比的 3DGS 可微渲染方式优化场景和姿态，提高了姿态准确性和重建质量。
3. 在 Waymo 和 KITTI 数据集上的实验表明，我们的方法能从未标定的图像中实现高效且准确的全局重建，为未标定的 3DGS 重建设定了新的技术水平。

2 相关工作

2.1 非姿态三维高斯点溅射

传统的 3D 高斯铺洒 (3DGS) (**?**) 依赖于由 COLMAP (**?**) 通常提供的精确相机姿态和稀疏点云。由于 COLMAP 的高计算成本和在复杂条件下的有限鲁棒性，近期的研究工作旨在从多视角图像直接恢复相机参数并重建高斯场景。CF-3DGS (**?**) 通过单目深度初始化高斯场，并逐步优化相机参数和高斯以支持无定位重建。COGS (**?**) 通过二维对应关系注册相机逐步重建场景，而 Rob-GS (**?**) 则引入一种鲁棒的成对姿态估计策略。NoParameters (**?**) 通过联合优化内参、外参和高斯，消除了对先验相机校准的需求。InstantSplat (**?**) 利用预训练的点图模型 DUSt3R (**?**) 进行初始化，并通过并行网格分区加速优化，但仍局限于含有相对较少图像的稀疏视图情境。另一领域的工作 **???** 利用预训练使前馈网络能够从成对图像直接预测高质量的高斯场景。近期的扩展支持更多输入并提高质量，但通常仅适用于十几个视图。随着场景大小和视图数量的增长，这些方法面临显著的内存和运行时间需求或鲁棒性下降的问题。为了使得在拥有数百张图像的户外场景上能够进行可扩展的无定位 3DGS，我们引入了预训练的多视角立体模型，并采用分而治之的策略进行 3DGS 重建。

2.2 多视图三维重建

传统的多视图重建流程 (**?**) 包括特征匹配、三角测量和束调整等手工制作的阶段。像 COLMAP (**???**) 这样的系统在静态场景中表现良好，但在累积误差、高计算成本和复杂场景中失败。

基于学习的方法 (**????**) 利用端到端网络从校准图像中恢复高质量的几何结构。更近期，端到端可微分的 SfM 框架 (**????**) 旨在直接从图像集合中联合估计相机参数和场景结构。

DUSt3R (**?**) 和 MASt3R (**?**) 从配对图像中回归密集点云和相机参数，用预训练的骨干网替代了手工设计的组件。这种前馈范式已经通过使用存储编码器 (**???**) 和子图融合网络 (**?**) 扩展到多图像设置。VGGT (**?**) 和 Fast3R (**?**) 进一步采用全局注意机制来跨多个视图进行推理。MV-DUSt3R+ (**?**) 和 FLARE (**?**) 可实现从稀疏视图输入到端到端的三维高斯撒点重建，类似的策略已应用于动态场景建模 (**??**) 。然而，这些方法在视图数量增加时遇到了挑战，面临内存瓶颈和重建鲁棒性下降的问题。在独立处理的子图间结构和尺度不一致使得全局对齐变得复杂，限制了大规模室外场景中的保真度。为应对这些挑战，我们采用了一种分而治之的策略，将图像分成局部子图，通过对齐和联合优化重建出全局一致的 3DGS 场景。

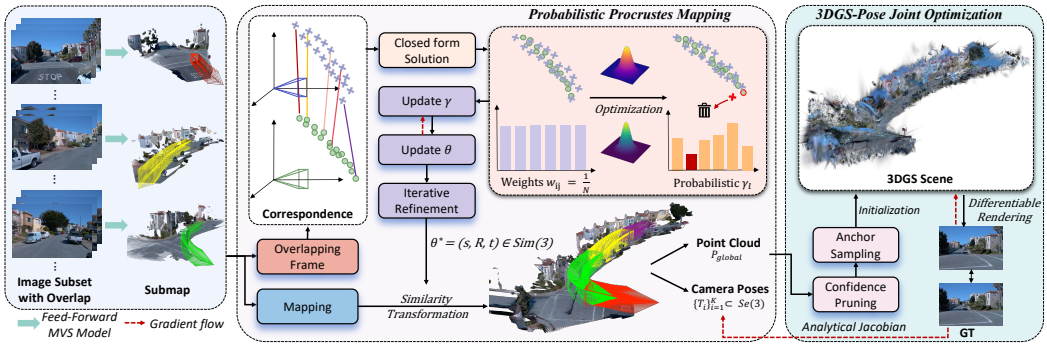

Figure 1: 我们首先将未处理的图像序列分成多个子集，并应用一个预训练的 MVS 模型来推断局部点云和相机位姿。通过构建重叠帧对应，将大规模子图对齐重新表述为一个概率 Procrustes 问题。这是通过一个封闭形式的 $Sim(3)$ 估计器来解决的，随后进行概率精炼和软异常值拒绝。最终的 3DGS 和位姿通过基于锚点的初始化和可微渲染共同优化，梯度通过解析雅可比矩阵传播。

3 方法

我们的目标是从数百张没有姿态的室外图像中重建高质量的三维高斯场景。如图 1 所示，图像集被划分为重叠的子集，每个子集由预训练的多视图立体模型独立处理，以估计局部点云和相机姿态。这些子图通过概率 Procrustes 映射进行全局对齐，然后对三维高斯场景和相机姿态进行联合优化，从而获得高保真且全球一致的重建结果。

给定 K 图像，$\{I_k\}_{k=1}^K$ 被分成固定大小的子集 G，每个子集被输入到一个预训练的 MVS 网络中，以生成一个局部子图 $\mathcal{M}_g = (\boldsymbol{P}_g, \{T_i^{(g)}\}_{i\in\mathcal{S}_g})$，包含一个稠密点云及其相机位姿。我们的目标是将这些融合成一个全局一致的场景 $(\boldsymbol{P}_{\text{global}}, \{T_i\}_{i=1}^K)$。

为了实现子图之间的全局一致对齐，我们旨在估计子图之间的最佳相似变换 $\theta = (s, R, \boldsymbol{t}) \in \text{Sim}(3)$，其中 $s \in \mathbb{R}^+$ 是比例因子，$R \in SO(3)$ 是旋转矩阵，$\boldsymbol{t} \in \mathbb{R}^3$ 是平移向量。然而，前馈式多视图立体（MVS）子图遭受尺度歧义和几何畸变的困扰。室外场景的结构复杂性进一步导致标准配准方法的失效。此外，每个子图通常包含数千万个点，这使得全局对齐成为一个高维度和计算密集的任务，挑战着准确性和效率。

为了解决这些挑战，我们在每对相邻子集之间定义了 k 个重叠帧，记为 $\mathcal{O}_{ab}$。这使我们能够将多子图对齐重新表述为一个经典的 Procrustes 问题。然后，我们在 $\mathcal{O}_{ab}$ 内的子图 a 和 b 之间提取每个像素的 3D 对应关系：

$$\mathcal{C}_{ab} = \{(\boldsymbol{p}_i^a, \boldsymbol{q}_j^b) \mid \pi(T_i^a \boldsymbol{p}_i^a) = \pi(T_j^b \boldsymbol{q}_j^b)\}, \tag{1}$$

其中 $\pi : \mathbb{R}^3 \to \mathbb{R}^2$ 是标准的相机投影模型。然后，我们通过优化来解决经典的 Procrustes 问题，该优化最小化变换后的点对之间的距离：

$$\theta^* = \underset{s>0, R\in SO(3), \boldsymbol{t}\in\mathbb{R}^3}{\arg\min} \sum_{(i,j)\in\mathcal{C}_{ab}} \left\| s\, R\, \boldsymbol{p}_i^a + \boldsymbol{t} - \boldsymbol{q}_j^b \right\|^2. \tag{2}$$

3.1 概率普氏映射

3.1.1 Procrustes 闭式解

为了有效估计方程 (2) 中的最优相似变换，我们采用 Kabsch-Umeyama 算法 (**??**) 来基于对应集 $\mathcal{C}_{ab}$ 计算一个闭合形式的解。首先，我们计算每个点集的质心：

$$\bar{\boldsymbol{p}} = \frac{1}{N} \sum\nolimits_{(i,j)\in\mathcal{C}_{ab}} \boldsymbol{p}_i^a, \qquad \bar{\boldsymbol{q}} = \frac{1}{N} \sum\nolimits_{(i,j)\in\mathcal{C}_{ab}} \boldsymbol{q}_j^b, \tag{3}$$

其中 $N = |\mathcal{C}_{ab}|$ 表示点对的数量。这些质心反映了两个点云的全局偏移，并将用于计算平移向量。接下来，我们构造两个点云之间的互协方差矩阵 Σ，以捕捉它们的空间相关结构：

$$\Sigma = \frac{1}{N} \sum_{(i,j)\in\mathcal{C}_{ab}} (\boldsymbol{p}_i^a - \bar{\boldsymbol{p}})(\boldsymbol{q}_j^b - \bar{\boldsymbol{q}})^\top. \tag{4}$$

通过进行奇异值分解（SVD）$\Sigma = U\Lambda V^\top$ ，我们获得了两个点集的主要方向。这使我们能够计算 θ^* 的闭合形式解：

$$R_0 = U\ \mathrm{diag}\big(1,\,1,\,\det(UV^\top)\big)\,V^\top, \quad s_0 = \frac{\mathrm{tr}(\Lambda)}{\mathrm{tr}(\Sigma_p)}, \quad \boldsymbol{t}_0 = \bar{\boldsymbol{q}} - s_0 R_0 \bar{\boldsymbol{p}}, \tag{5}$$

，其中 R_0 是确保右手坐标系的适当旋转，s_0 由奇异值与源点云方差的比率给出，$\boldsymbol{t}_0$ 用于对齐质心。这个闭式步骤为细化子图对齐提供了高效的初始化。

虽然闭式解在理论上是最优的，但它依赖于两个关键假设：

1. 两个点集的空间分布必须是相同的。
2. 对应关系必须无噪声，例如，$\left\|s_0\,R_0\,\boldsymbol{p}_i^a + \boldsymbol{t}_0 - \boldsymbol{q}_j^b\right\| = 0,\ \forall (i,j) \in \mathcal{C}_{ab}$ 。

这些条件在点云完全准确且几何一致的理想情况下成立。然而，在实际单目重建中，即使有重叠帧提供像素级对应关系，3D 点的空间分布仍可能有显著差异。这违反了闭式解的假设，并且在直接使用时常导致系统偏差，使映射结果的鲁棒性降低。

3.1.2 概率映射

我们观察到，由前馈多视图立体（MVS）模型预测的点云由于学习到的先验而表现出结构性偏差，这导致了在闭式对齐中出现系统误差。为了解决这个问题，我们将点云配准表述为一个带有储尘箱机制的概率普罗克路斯特问题。具体来说，我们将子图间的每个候选对应 $(\boldsymbol{p}_l, \boldsymbol{q}_l)$ 与一个概率匹配权重 $\gamma_l \in [0,1]$ 联系起来。

为了处理异常值，我们引入了一种基于概率的垃圾箱机制：一个控制参数 $\eta \in [0,1]$ 指定了允许从对齐中排除的最大对应关系的比例。为实现这一点，我们在目标集合中增加一个虚拟垃圾箱点 $\boldsymbol{q}_{\text{dustbin}}$ ，并为其分配一个固定的边际权重 $b_{\text{dustbin}} = \delta$ 。

我们的目标是共同优化相似变换 $\theta = (s, R, \boldsymbol{t}) \in \mathrm{Sim}(3)$ 和对应的概率 $\{\gamma_l\}_{l=1}^N$ 。每个权重 γ_l 在当前变换下编码了源点 $\boldsymbol{p}_l$ 与其目标 $\boldsymbol{q}_l$ 之间的软关联强度。目标被表述为：

$$\min_{s,R,\boldsymbol{t},\gamma} \sum_l \gamma_l \|sR\boldsymbol{p}_l + \boldsymbol{t} - \boldsymbol{q}_l\|^2 + \epsilon \sum_l \gamma_l \ln \gamma_l, \quad \text{subject to} \quad \sum_l \gamma_l = 1, \tag{6}$$

其中 $\gamma_l \in [0,1]$ 表示源点 $\boldsymbol{p}_l$ 和目标点 $\boldsymbol{q}_l$ 之间的软匹配概率。

我们使用闭式 Kabsch–Umeyama 方法初始化变换参数 $\theta^{(0)} = (s^{(0)}, R^{(0)}, \boldsymbol{t}^{(0)})$ 。给定一个固定的变换，关联权重 γ 通过熵正则化优化进行更新：

$$\gamma_l \propto \exp\left(-\frac{\|sR\boldsymbol{p}_l + \boldsymbol{t} - \boldsymbol{q}_l\|^2}{\epsilon}\right), \tag{7}$$

，在比例关系之后，通过一个归一化步骤，利用逐步迭代优化来满足边际约束。

变换更新。 固定 γ ，我们计算目标 $\mathcal{L}_\theta$ 对变换参数 $\theta = (s, R, \boldsymbol{t})$ 的梯度。设 $\boldsymbol{p}_l' = R\boldsymbol{p}_l$ ，则：

$$\nabla_{\boldsymbol{t}} \mathcal{L}_\theta = 2 \sum_{l=1}^N \gamma_l^{(k)} \left(s\boldsymbol{p}_l' + \boldsymbol{t} - \boldsymbol{q}_l\right). \tag{8}$$

$$\nabla_s \mathcal{L}_\theta = 2 \sum_{l=1}^N \gamma_l^{(k)} \left(s\boldsymbol{p}_l' + \boldsymbol{t} - \boldsymbol{q}_l\right)^\top \boldsymbol{p}_l'. \tag{9}$$

为了更新旋转 R ，我们使用单位四元数 $q = (w, v^\top)^\top$, where $v = (x,y,z)^\top$ 对其进行参数化，并使用链式法则来计算：

$$\nabla_q \mathcal{L}_\theta = 2 \sum_{l=1}^N \gamma_l^{(k)} \left(sR(q)\boldsymbol{p}_l + \boldsymbol{t} - \boldsymbol{q}_l\right)^\top s \frac{\partial (R(q)\boldsymbol{p}_l)}{\partial q}. \tag{10}$$

使用梯度下降法更新转换参数：

$$\theta^{(k+1)} = \theta^{(k)} - \eta_\theta \nabla_\theta \mathcal{L}_\theta(\theta^{(k)}), \tag{11}$$

其中 η_θ 是学习率。当姿态收敛或达到最大迭代次数时，优化终止。在实际应用中，精确的闭式初始化通常在几次迭代内即可收敛。

然后，将得到的最优变换 $\theta^\star = (s_g, R_g, \boldsymbol{t}_g)$ 应用于子图 $\mathcal{S}_g$ 中的所有 3D 点和相机姿态，将它们转换到全局坐标框架中并更新相应的姿态。对所有子图迭代地应用此过程，将产生一个全局一致的点云和一个统一的相机轨迹。

3.2 3DGS 与姿态联合优化

子地图对齐后，我们在统一的全局坐标系中获得初始相机位姿和稠密点云。然而，由于单目重建中固有的比例不确定性、深度噪声及残留位姿漂移，因此需要进一步优化。为此，我们使用一种可微分的 3DGS 渲染管线共同优化三维高斯参数和相机位姿，从而提高位姿准确性和重建质量。

我们将场景建模为一组三维高斯分布：$\mathcal{G} = \{\mathcal{G}_i : (\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \boldsymbol{c}_i, \Lambda_i) | i = 1, ..., N\}$ 。每个高斯点由位置 μ_i 、三维协方差矩阵 $\Sigma_i \in \mathbb{R}^{3\times3}$ 、不透明度 Λ_i 和由球谐函数获得的颜色 $\boldsymbol{c}_i$ 定义。

对于一个特定的视图，给定相机位姿 $T = (R, t)$ 和相机内参 $\boldsymbol{K} \in \mathbb{R}^{3\times3}$ ，我们可以通过光栅化管道渲染出 RGB 图像 $\hat{I}$ 。首先，将我们的 3D 高斯投影到 2D 图像平面：

$$\mu' = \pi(T \cdot \mu), \qquad \Sigma' = JW\Sigma W^\top J^\top, \tag{12}$$

，其中 π 是投影操作，W 是 T 的旋转分量，J 是投影变换的仿射近似的雅可比矩阵。然后，像素的颜色可以表示为与该像素重叠的 N 个有序点的 alpha 混合：

$$C = \sum_{i\in N} c_i \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j), \tag{13}$$

，其中 α_i 是通过评价协方差为 Σ' 的二维高斯获得的密度。

联合优化 我们首先从全局点云中提取一个高置信度子集，并应用下采样来获得初始锚集 $\mathcal{A} = \{\boldsymbol{x}_i\}$ ，该锚集用于初始化 3D 高斯 $\mathcal{G}_i = (\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \boldsymbol{c}_i, \Lambda_i)$ 。这些高斯 $\mathcal{G} = \{\mathcal{G}_i\}$ 共同构成初始全局场景。

基于此，我们定义了一个闭环优化框架，该框架通过以下目标联合优化相机姿态 $T = (R, \boldsymbol{t})$ 和高斯参数 $\{\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \boldsymbol{c}_i, \Lambda_i\}_{i=1}^N$ ：

$$\mathcal{L}_{\text{total}} = \alpha \left\| \hat{I}_k - I_k \right\|_1 + (1 - \alpha)\text{SSIM}(\hat{I}_k, I_k), \tag{14}$$

，其中 $\hat{I}_k$ 表示在当前视图 k 下渲染的图像，I_k 是真实值，α 是平衡 L1 和 SSIM 项的权重。

从 3DGS 渲染管道可以看出，摄像机位姿 T 的梯度依赖于两个中间量：Σ' 和每个高斯 $\mathcal{G}_i$ 的投影坐标 μ_i' 。通过应用链式法则，我们可以推导出 $\frac{\partial \mathcal{L}}{\partial T}$ 的完整解析表达式，从而避免自动微分的运行时开销，并在四元数归一化过程中确保数值稳定。得到的解析梯度具有以下形式：

$$\frac{\partial \mathcal{L}}{\partial T} = \frac{\partial \mathcal{L}}{\partial \hat{I}_k} \frac{\partial \hat{I}_k}{\partial T} = \frac{\partial \mathcal{L}}{\partial \hat{I}_k} \frac{\partial \hat{I}_k}{\partial \alpha_i} \left(\frac{\partial \alpha_i}{\partial \Sigma'} \frac{\partial \Sigma'}{\partial T} + \frac{\partial \alpha_i}{\partial \mu'} \frac{\partial \mu'}{\partial T} \right), \tag{15}$$

$$\frac{\partial \Sigma'}{\partial T} = \frac{\partial \Sigma'}{\partial J} \frac{\partial J}{\partial \mu_c} \frac{\partial \mu_c}{\partial T} + \frac{\partial \Sigma'}{\partial W} \frac{\partial W}{\partial T}, \tag{16}$$

$$\frac{\partial \mu'}{\partial T} = \frac{\partial \mu'}{\partial \mu_c} \frac{\partial \mu_c}{\partial T}, \tag{17}$$

其中 μ_c 表示世界坐标中的点 μ 通过位姿 T 变换到摄像机坐标系中。我们通过一个单位四元数 $q \in \mathbb{R}^4$ 和一个平移向量 $t \in \mathbb{R}^3$ 参数化摄像机位姿 $T = (R, t)$ ，并在附录 **??** 中提供了关于 q 和 t 的解析梯度。为了保持 q 的单位长度，我们应用投影梯度更新：

$$q \leftarrow \frac{q - \eta \nabla_q \mathcal{L}}{\|q - \eta \nabla_q \mathcal{L}\|}. \tag{18}$$

通过联合优化相机位姿、三维高斯参数和图像重投影，我们获得了一个全局一致的三维高斯场景，其具有准确的位姿估计和高保真渲染。

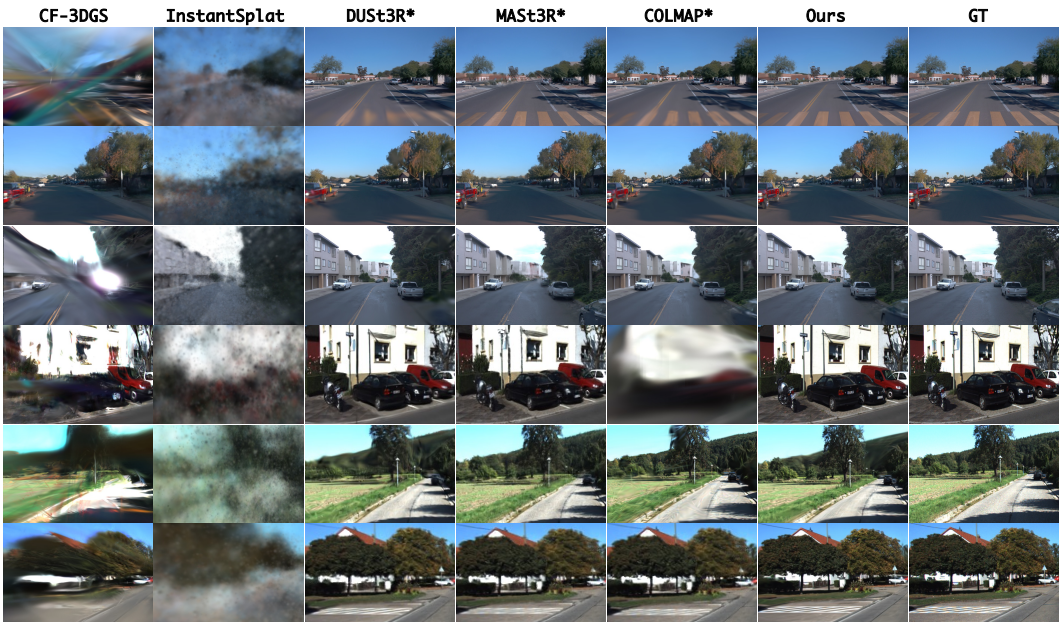

Figure 2: Waymo（前三行）和 KITTI（后三行）上的定性比较。标有星号（*）的方法使用 3DGS 重建。由于内存限制，InstantSplat 仅在 80 张图像上进行训练。我们的方法实现了高保真图像重建，具有更清晰的纹理和更精细的细节。

Table 1: Waymo 和 KITTI 数据集的定量结果。T_m 表示匹配时间，T_t 表示训练时间。用星号 (*) 标记的方法表示使用 3DGS 重建的方法。Flare 由于内存不足 (OOM) 而失败。ATE 衡量姿态精度；PSNR、SSIM 和 LPIPS 评估图像重建质量。最佳结果以粗体显示。我们的方法在准确性和重建保真度方面达到了最佳。

Method	Waymo (?)							KITTI ?						
	ATE ↓	PSNR ↑	SSIM ↑	LPIPS ↓	T_m	T_t	GPU	ATE ↓	PSNR ↑	SSIM ↑	LPIPS ↓	T_m	T_t	GPU
COLMAP+SPSG*	3.68	30.17	0.893	0.314	45min	58min	13GB	12.1	19.52	0.647	0.438	41 min	35 min	10GB
CF-3DGS	5.46	22.69	0.736	0.316	-	67min	12GB	5.99	15.91	0.490	0.486	-	195min	11GB
DUSt3R*	5.59	29.39	0.871	0.310	38min	42min	46GB	3.10	23.87	0.767	0.311	32min	51min	45GB
MASt3R*	6.12	27.49	0.867	0.308	82min	46min	40GB	5.71	23.63	0.778	0.274	94min	57min	37GB
Fast3R*	43.9	20.66	0.764	0.468	30s	46min	14GB	47.3	16.73	0.533	0.581	30s	28min	10GB
Flare	-	-	-	-	-	-	-	-	-	-	-	-	-	-
InstantSplat	2.55	19.22	0.739	0.515	58min	8min	41GB	2.23	13.09	0.414	0.680	97min	7min	40GB
Ours	1.41	31.53	0.915	0.245	1min	63min	14GB	1.64	24.83	0.780	0.272	8min	31min	12GB

4 实验

4.1 实验设置

实现细节。

我们的实验是使用 PyTorch 框架在单个 NVIDIA RTX A6000 GPU 和 AMD EPYC 7542 CPU 上进行的。所有结果都是使用表现最好的预训练 MVS 模型 VGGT (?) 报告的。

我们将组的大小设置为 60，组间重叠设置为 $K = 1$，通过经验观察得出这是最佳性能。相机姿态的优化起始学习率为 10^{-5}，随着迭代收敛逐渐衰减至 10^{-7}。尘箱容量设置为 20 %。为了实现高效的优化，我们首先剪除置信度最低的 3 % 的点，然后应用基于体素的下采样保留 0.05 % 的点作为锚点。更多的实现细节在补充材料中提供。

数据集。我们在两个室外数据集上进行评估，从 Waymo ? 中选择了 9 个场景组，从 KITTI ? 中选择了 8 个场景组，每个场景组由 200 张在不同条件下拍摄的前视图图像组成。所有图像都用于评估。我们评估整个序列中的图像重建质量和估计相机姿态的准确性。

指标。我们评估相机姿态估计和场景重建（基于图像渲染质量）。对于姿态，我们通过绝对轨迹误差（ATE）报告平移误差，以米（m）为单位测量。对于重建，我们使用 PSNR、SSIM 和 LPIPS。此外，我们记录训练时间和峰值内存，以评估效率和可扩展性。

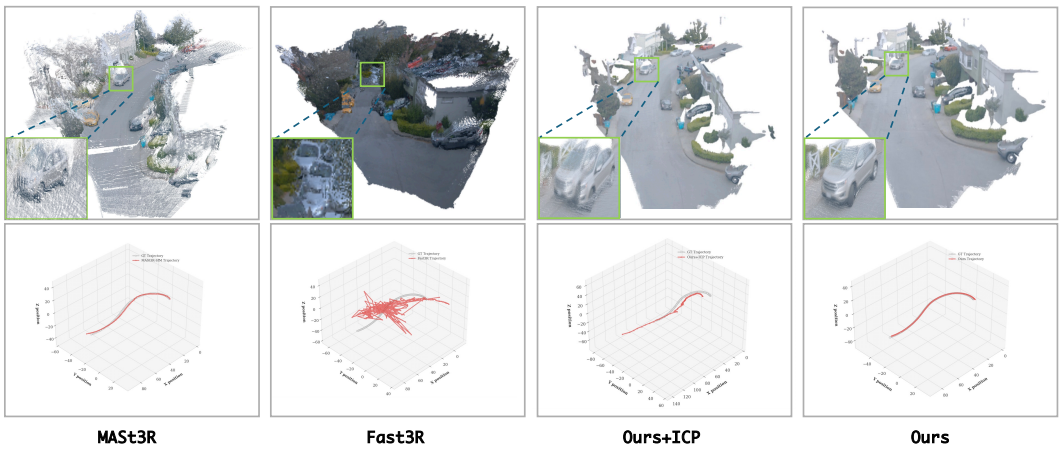

Figure 3: 重建点云的定性比较。底部行显示了估计的相机轨迹。Fast3R 表现出显著的漂移，而 Ours+ICP 仍然存在不对齐的问题。我们的方法实现了精确的子图融合和全局一致的位姿估计。

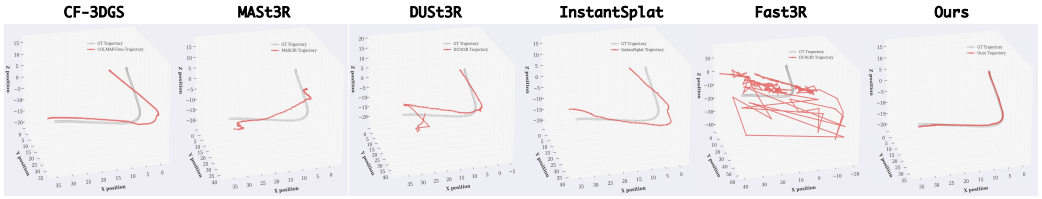

Figure 4: 相机姿态估计的定性比较。红色表示估计的姿态，而灰色表示真实值。与其他方法相比，我们的方法在姿态准确性上表现优越。

基线。我们将我们的方法与七个基线进行比较，包括 COLMAP+SPSG (?) 、CF-3DGS (?) 、DUSt3R (?) 、MASt3R (?) 、Fast3R (?) 、Flare (?) 和 InstantSplat (?) 。由于 COLMAP+SPSG、DUSt3R、MASt3R 和 Fast3R 仅从图像中估计点云和相机位姿而不直接生成高斯分布，我们使用原始的 3DGS (?) 训练管道进行场景重建，用星号（*）表示。

4.2 实验结果分析。

我们在 Waymo 和 KITTI 数据集上评估了我们的方法，结果汇总在表格 1 中。InstantSplat 是为稀疏视图设置设计的，由于内存限制无法扩展到大输入，即使限制在 80 个视图时表现也很差。Fast3R 效率高，但因姿态估计严重不准确而受到影响。Flare 仅支持有限数量的输入视图大小。COLMAP 在多个场景中由于发散而表现不佳，导致较大的错误影响了平均 ATE。

相比之下，我们的方法结合了预训练的 VGGT 模型、PPM 映射模块和联合姿态优化，以达到卓越的重建质量和轨迹准确性。图 2 展示了在 6 个场景中的渲染结果对比，包括道路布局、建筑结构、车辆和植被。图 3 突出了子地图边界处点云的一致性。与 Fast3R、Mast3R 和基于 ICP 的配准相比，我们的方法在组间实现了无缝对齐，在重叠区域的漂移最小。相应的轨迹图证实了我们的姿态优化的有效性。此外，我们的方法能够在短短几分钟内生成全球一致且准确的点云和相机姿态。

图 4 进一步展示了我们的估计轨迹明显比竞争方法更准确和稳定。这些结果证明了我们的框架在从数百张户外图像中进行无位姿重建的有效性。

4.3 消融研究

我们在 Waymo 数据集上进行消融实验，以评估所提出的概率 Procrustes 映射（PPM）和联合 3DGS 优化模块的有效性。如表格 2 所示，我们首先比较不同的对齐策略。使用 ICP 或 COLMAP 预测的相对位姿进行子图注册会导致显著的位姿误差以及在最终重建中出现明显的错位。相比之下，我们的概率 Procrustes 映射模块实现了显著更高的注

Table 2: Waymo 数据集上的消融结果。上图：根据我们的方法比较不同的子图对齐策略。“我们的 + ICP” 和 “我们的 + COLMAP” 表示使用来自 ICP 或 COLMAP 的相对位姿进行子图映射。下图：我们的概率 Procrustes 映射 (PPM) 和 3DGS 位姿联合优化 (联合优化) 模块的消融实验。

Method	ATE	PSNR	SSIM	LPIPS
Ours w/ ICP	3.24	27.63	0.852	0.341
Ours w/ COLMAP	3.77	28.13	0.835	0.336
w/o PPM	5.97	27.54	0.824	0.363
w/o Joint Opt.	0.68	30.47	0.903	0.242
Ours	0.56	32.72	0.935	0.211

册准确性和重建保真度，表明结合闭合形式对齐和概率精化的优势。

我们进一步剖析每个核心模块以评估其单独影响。禁用 PPM 模块并用 VGGT 相对位姿估计替代会导致全球一致性的退化和较低质量的新视图合成。同样，在 3DGS 训练阶段固定相机位姿会导致图像质量和几何一致性下降。这些结果证实了联合优化相机位姿和场景表示对于准确和稳健的重建是必不可少的。总体而言，PPM 映射模块和联合位姿优化在确保准确的全球对齐和高质量的 3DGS 重建中发挥了关键作用。

4.4 局限性

我们的方法依赖于预训练的 MVS 预测在初始位姿和几何结构上的质量。虽然联合优化阶段可以纠正中等错误，但初始化中的严重不准确可能会降低最终重建的质量。随着输入帧数量的增加，累积漂移和更高的优化成本可能会限制大规模或长序列的可扩展性。此外，在具有频繁运动或遮挡的高度动态场景中，视图间缺乏一致的对应关系可能会阻碍稳定优化并降低重建的保真度。

我们提出了一个可扩展且稳健的框架用于无姿态的 3D 高斯斑点重建。通过将预训练的多视图立体模型与分而治之策略相结合，我们的方法能够高效处理数百个未校准视角的户外场景。我们引入了一个概率的 Procrustes 映射模块用于全局配准，随后是一个 3DGS 和姿态联合优化模块，用于共同优化相机姿态和 3D 高斯。我们的方法达到了最先进的性能，并为真实世界场景中的无姿态 3D 重建提供了实用价值。

在本附录中，我们推导了旋转点 $R(q)\,\boldsymbol{p}$ 关于单位四元数 $q = [w,\, v^\top]^\top$ 的雅可比矩阵，其中 $v = (x, y, z)^\top$ 。

我们从旋转的等效表达式开始：

$$R(q)\,\boldsymbol{p} = \left(w^2 - \|v\|^2\right)\boldsymbol{p} + 2\,v\,(v^\top \boldsymbol{p}) + 2\,w\,(v \times \boldsymbol{p}). \tag{19}$$

定义：

$$\Delta(q) = R(q)\,\boldsymbol{p} = A + B + C, \tag{20}$$

具有：

$$A = (w^2 - v^\top v)\,\boldsymbol{p}, \quad B = 2\,v\,(v^\top \boldsymbol{p}), \quad C = 2\,w\,(v \times \boldsymbol{p}). \tag{21}$$

4.5

对 w 的导数

只有 A 和 C 依赖于 w 。我们有：

$$\frac{\partial A}{\partial w} = 2w\,\boldsymbol{p}, \qquad \frac{\partial C}{\partial w} = 2\,(v \times \boldsymbol{p}). \tag{22}$$

因此：

$$\frac{\partial\,(R(q)\,\boldsymbol{p})}{\partial w} = 2\,w\,\boldsymbol{p} + 2\,(v \times \boldsymbol{p}). \tag{23}$$

4.6

2. 对 v 的导数

令 $[\boldsymbol{p}]_\times$ 表示满足 $[\boldsymbol{p}]_\times u = \boldsymbol{p} \times u$ 的斜对称矩阵。我们计算：

$$\frac{\partial A}{\partial v} = -2\,(v^\top \boldsymbol{p})\,I_3 = -2\,\boldsymbol{p}\,v^\top, \tag{24}$$

$$\frac{\partial B}{\partial v} = 2\,(v^\top \boldsymbol{p})\,I_3 + 2\,v\,\boldsymbol{p}^\top, \tag{25}$$

$$\frac{\partial C}{\partial v} = 2\,w\,[\boldsymbol{p}]_{\times}. \tag{26}$$

结合这些项得出：

$$\frac{\partial\,(R(q)\,\boldsymbol{p})}{\partial v} = -2\,\boldsymbol{p}\,v^{\top} + 2\,(v^{\top}\boldsymbol{p})\,I_3 + 2\,v\,\boldsymbol{p}^{\top} + 2\,w\,[\boldsymbol{p}]_{\times}, \tag{27}$$

4.7

3. 装配 3×4 雅可比矩阵

将关于 w 和 v 的偏导数堆叠起来就得到了完整的雅可比矩阵：

$$\frac{\partial\,(R(q)\,\boldsymbol{p})}{\partial q} = \Big[\underbrace{2w\,\boldsymbol{p} + 2\,(v \times \boldsymbol{p})}_{3\times 1} \;\Big|\; \underbrace{-2\,\boldsymbol{p}\,v^{\top} + 2\,(v^{\top}\boldsymbol{p})\,I_3 + 2\,v\,\boldsymbol{p}^{\top} + 2\,w\,[\boldsymbol{p}]_{\times}}_{3\times 3}\Big]_{3\times 4}. \tag{28}$$

，其中第一列对应于 $\partial/\partial w$ ，剩下的三列对应于 $\partial/\partial x, \partial/\partial y, \partial/\partial z$ 。这个雅可比矩阵可以直接用于基于梯度的优化中。

在本附录中，我们推导了关于相机姿态 $T = (R(q), t)$ 的投影坐标 $\mu' = \pi(\mu_c)$ 和投影协方差 $\Sigma' = J\,R(q)\,\Sigma R(q)^\top J^\top$ 的解析梯度，其中

$$\mu_c = R(q)\,\mu + t, \quad q = [q_r, q_i, q_j, q_k]^\top, \quad t \in \mathbb{R}^3,$$

是一个三维点，$J = \frac{\partial \pi(\mu_c)}{\partial \mu_c}$ 是 2×3 投影雅可比矩阵。

4.8

1. 投影坐标 μ' 的梯度

根据链式法则，相对于 平移的导数为

$$\frac{\partial \mu'}{\partial t} = J\,\frac{\partial \mu_c}{\partial t} = J \;=\; \begin{bmatrix} \frac{f_x}{z_c} & 0 & -\frac{f_x\,x_c}{z_c^2} \\ [8pt]0 & \frac{f_y}{z_c} & -\frac{f_y\,y_c}{z_c^2} \end{bmatrix}. \tag{29}$$

。相对于 四元数的导数为

$$\frac{\partial \mu'}{\partial q} = J\,\frac{\partial \mu_c}{\partial q} = J\begin{bmatrix} \frac{\partial \mu_c}{\partial q_r} & \frac{\partial \mu_c}{\partial q_i} & \frac{\partial \mu_c}{\partial q_j} & \frac{\partial \mu_c}{\partial q_k} \end{bmatrix}. \tag{30}$$

。

这里的 3×1 块 $\partial \mu_c / \partial q_\alpha$ 为：

$$\frac{\partial \mu_c}{\partial q_r} = 2\begin{bmatrix} 0 & -q_k & q_j \\ q_k & 0 & -q_i \\ -q_j & q_i & 0 \end{bmatrix}\mu, \tag{31}$$

$$\frac{\partial \mu_c}{\partial q_i} = 2\begin{bmatrix} 0 & q_j & q_k \\ q_j & -2q_i & -q_r \\ q_k & q_r & -2q_i \end{bmatrix}\mu, \tag{32}$$

$$\frac{\partial \mu_c}{\partial q_j} = 2\begin{bmatrix} -2q_j & q_i & q_r \\ q_i & 0 & q_k \\ q_r & q_k & -2q_j \end{bmatrix}\mu, \tag{33}$$

$$\frac{\partial \mu_c}{\partial q_k} = 2\begin{bmatrix} -2q_k & -q_r & q_i \\ q_r & -2q_k & q_j \\ q_i & q_j & 0 \end{bmatrix}\mu. \tag{34}$$

4.9

2. 投影协方差的梯度 Σ'

由于平移不影响协方差：

$$\frac{\partial \Sigma'}{\partial t} = 0. \tag{35}$$

对于四元数：

$$\frac{\partial \Sigma'}{\partial q} = J\,\frac{\partial\big(R\,\Sigma_{\mathrm{w}}\,R^\top\big)}{\partial q}\,J^\top = J\Big(\frac{\partial R}{\partial q}\,\Sigma_{\mathrm{w}}\,R^\top + R\,\Sigma_{\mathrm{w}}\,\frac{\partial R^\top}{\partial q}\Big)J^\top, \tag{36}$$

，其中 $\partial R / \partial q$ 是旋转矩阵关于四元数的经典梯度，$\partial R^\top / \partial q$ 是其转置。

这些闭式导数使得可以通过可微的 3DGS 渲染管道高效地对 μ' 和 Σ' 进行反向传播。

5

NeurIPS 论文检查清单

1. **Claims**

   问题：摘要和引言中提出的主要论点是否准确反映了论文的贡献和范围？

   答案：[Yes]

   说明：摘要和引言准确地总结了论文的贡献，包括提出的方法、关键技术见解和经验上的改进。这些论点有理论分析和实验结果的强力支持。

   指南：

   - 答案 NA 表示摘要和介绍部分不包括论文中提出的主张。
   - 摘要和/或引言应明确说明所提出的主张，包括论文中的贡献以及重要的假设和限制条件。对这个问题给出“否”或“无”的回答不会被审稿人很好地接受。
   - 做出的声明应与理论和实验结果相符，并反映结果在多大程度上可以推广到其他环境。
   - 只要清楚地说明这些目标并未通过本文实现，将理想目标作为动机是可以的。

2. **Limitations**

   问题：论文是否讨论了作者进行的工作的局限性？

   答案：[Yes]

   说明：论文包含一个专门的局限性部分，讨论假设、潜在的失败情况和普遍性问题。

   指南：

   - 答案 NA 意味着论文没有限制，而答案 No 意味着论文有限制，但这些限制没有在论文中讨论。
   - 我们鼓励作者在他们的论文中创建一个单独的“局限性”部分。
   - 论文应指出任何强假设以及结果对这些假设的违反有多稳健（例如，独立性假设、无噪声环境、模型的良好规范、渐近逼近仅在局部成立）。作者应反思这些假设在实践中可能如何被违反以及其影响会是什么。
   - 作者应该反思所提出主张的范围，例如，如果该方法仅在少数数据集或实验中进行了测试。通常，实证结果常常依赖于隐含的假设，这些假设应该明确表达。
   - 作者应该反思影响该方法性能的因素。例如，当图像分辨率低或者图像是在光线较暗的环境下拍摄时，人脸识别算法的性能可能会很差。又或者，由于无法处理技术术语，语音转文字系统可能无法可靠地为在线讲座提供闭字幕。
   - 作者应讨论所提出算法的计算效率及其如何随数据集规模的变化而变化。
   - 如果适用，作者应讨论其方法在解决隐私和公平性问题上可能存在的局限性。
   - 虽然作者可能担心对局限的完全坦诚可能会被审稿人用作拒稿的理由，但更糟糕的结果可能是审稿人发现了论文中未被承认的局限。作者应该运用他们最好的判断力，并认识到支持透明度的个人行为在发展保持社区完整性的规范中起到了重要作用。审稿人将被特别指示不要因为诚实对待局限而进行惩罚。

3. **Theory assumptions and proofs**

   问题：对于每个理论结果，论文是否提供了一整套假设和完整的（且正确的）证明？

   答案：[NA]

   说明：本文不包含正式的理论结果或证明。

   指南：

   - 答案 NA 表示论文不包括理论结果。
   - 论文中的所有定理、公式和证明都应编号和交叉引用。
   - 在任何定理的陈述中，所有假设都应明确说明或引用。
   - 证明可以出现在论文正文中或补充材料中，但如果它们出现在补充材料中，建议作者提供一个简短的证明概要以提供直观理解。
   - 相反，论文核心部分提供的任何非正式证明都应由附录或补充材料中提供的正式证明来补充。
   - 证明所依赖的定理和引理应被正确引用。

4. **Experimental result reproducibility**

   问题：论文是否充分披露了再现论文主要实验结果所需的所有信息，以至于影响论文的主要论点和/或结论（无论代码和数据是否提供）?

   答案：[Yes]

   说明：本文详细介绍了实验设置、数据集使用、模型架构、训练过程和评估协议。包含了充足的信息，以便能够重现所有关键结果。

   指南：

   - 答案 NA 表示论文不包含实验。
   - 如果论文包含实验，这个问题的“否”回答可能会被审稿人不认可：无论是否提供代码和数据，使论文可重复性是很重要的。
   - 如果贡献是一组数据和/或一个模型，作者应描述为了使结果可重复或可验证所采取的步骤。
   - 根据贡献的不同，可通过多种方式实现可重复性。例如，如果贡献是一个新颖的架构，完全描述该架构可能就足够了，或者如果贡献是一个特定的模型和实证评估，可能需要使他人能够使用相同的数据集复制该模型，或者提供对该模型的访问。通常，发布代码和数据是实现这一点的一个好方法，但可重复性也可以通过详细的复制结果说明、访问托管模型（例如，在大语言模型的情况下）、发布模型检查点或其他适合所进行研究的方法来提供。
   - 虽然 NeurIPS 不要求发布代码，但会议确实要求所有提交作品提供某种合理的可复现性途径，这可能取决于贡献的性质。例如
     (a) 如果文章的主要贡献是一个新的算法，那么应该明确如何复现该算法。
     (b) 如果贡献主要是一个新的模型架构，论文应清楚且完整地描述该架构。
     (c) 如果该贡献是一个新模型（例如，一个大型语言模型），那么应该有一种方式可以访问这个模型以重现结果，或者一种重现该模型的方法（例如，使用开源数据集或构建数据集的说明）。
     (d) 我们认识到，在某些情况下，可重复性可能会比较棘手，在这种情况下，作者可以描述他们提供可重复性的特定方式。对于闭源模型，可能访问模型的方式会有某种限制（例如，仅限于注册用户），但其他研究人员应该有某种途径来复现或验证结果。

5. **Open access to data and code**

   问题：论文是否提供对数据和代码的开放访问，并附有足够的说明以忠实再现主要实验结果，正如附录材料中所描述的?

   答案：[Yes]

   理由：代码和数据将在接收后发布，并附有重现主要结果的完整说明。

   指南：

   - 答案 NA 表示该论文不包括需要代码的实验。
   - 有关更多详细信息，请参阅 NeurIPS 代码和数据提交指南（`https://nips.cc/public/guides/CodeSubmissionPolicy`）。
   - 尽管我们鼓励发布代码和数据，但我们理解这可能无法实现，因此“不”是可以接受的回答。论文不能仅仅因为没有包含代码而被拒稿，除非代码对于论文贡献是核心部分（例如，用于一个新的开源基准测试）。
   - 说明应该包含用于重现结果的确切命令和环境。有关更多细节，请参见 NeurIPS 代码和数据提交指南（`https://nips.cc/public/guides/CodeSubmissionPolicy`）。
   - 作者应提供关于数据访问和准备的说明，包括如何访问原始数据、预处理数据、中间数据和生成的数据等。
   - 作者应该提供脚本，以重现新提出的方法和基线的所有实验结果。如果只有一部分实验是可重复的，他们应该说明哪些实验被脚本省略以及原因。
   - 在提交时，为了保持匿名性，作者应发布匿名化版本（如果适用）。
   - 建议在补充材料（附加到论文中）中提供尽可能多的信息，但允许包含数据和代码的 URL。

6. **Experimental setting/details**

问题：论文是否详细说明了所有必要的训练和测试细节（例如，数据划分，超参数，它们是如何选择的，优化器类型等）以理解结果？

答案: [Yes]

理由：我们在主文中报告了实验细节。

指南：

- 答案 NA 意味着论文不包含实验。
- 实验设置应在论文的核心部分进行详细展示，以便读者欣赏和理解结果。
- 完整的细节可以通过代码、附录或补充材料提供。

7. **Experiment statistical significance**
8. 问题：这篇论文是否适当且正确地定义了误差条，或提供了实验统计显著性的其他适当信息？

   答案: [No]

   理由：本文报告了 PSNR、SSIM、ATE、LPIPS，这些通常用作图像处理实验中性能的衡量标准。这种方法在该领域是标准的，足以传达所研究方法的性能。

   指导原则：

   - 答案 NA 表示论文不包含实验。
   - 如果结果附有误差棒、置信区间或显著性检验，至少是对于支持论文主要结论的实验，作者应该回答“是”。
   - 误差条所捕捉的可变性因素应清晰陈述（例如，训练/测试划分、初始化、某个参数的随机抽取或在给定实验条件下的整体运行）。
   - 应解释误差棒的计算方法（闭式公式、调用库函数、自举法等）。
   - 应说明所作的假设（例如，正态分布的误差）。
   - 应该明确误差棒是标准差还是均值的标准误差。
   - 可以报告 1 西格玛误差带，但应说明这一点。如果误差的正态性假设未得到验证，作者应优先报告 2 西格玛误差带，而非声称他们有 96 % 置信区间。
   - 对于不对称分布，作者应注意不要在表格或图中显示对称误差条，这样会导致结果超出范围（例如，负的错误率）。
   - 如果在表格或图表中报告了误差条，作者应该在文中解释它们是如何计算的，并在文中引用相应的图形或表格。
9. **Experiments compute resources**

   问题：对于每个实验，论文是否提供了足够的信息关于重复实验所需的计算机资源（计算工作者的类型、内存、执行时间）？
10. 答案：[Yes]

    证明：我们描述了在实验中使用的计算环境，包括 GPU 类型、内存大小、训练小时数。

    指南：

    - 答案 NA 表示论文不包括实验。
    - 论文应指出计算工作节点的类型，例如 CPU 或 GPU，内部集群或云服务提供商，包括相关的内存和存储。
    - 论文应提供每次单独实验运行所需的计算量，并估计总计算量。
    - 论文应披露整个研究项目是否需要比论文中报告的实验更多的计算（例如，未能进入论文的初步实验或失败实验）。
11. **Code of ethics**

    问题：论文中进行的研究是否在各个方面都符合 NeurIPS 伦理守则 `https://neurips.cc/public/EthicsGuidelines`？

    答案：[Yes]

    证明：我们没有预见到与数据使用、环境影响或公平性相关的任何伦理问题。

    指南：

    - 答案 NA 表明作者尚未审阅 NeurIPS 伦理规范。

- 如果作者回答“否”，他们应解释需要偏离《伦理规范》的特殊情况。
- 作者应确保保密性（例如，如果由于其司法管辖区的法律或法规而有特别考虑）。

12. **Broader impacts**

13. 问题：论文是否讨论了所做工作的潜在积极社会影响和消极社会影响?

    答案：[Yes]

    理论依据：论文包含一个更广泛影响部分，讨论我们 3D 场景重建框架的潜在社会应用。

    指南：

    - 答案 NA 表示所进行的工作没有社会影响。
    - 如果作者回答 NA 或否，他们应该解释为什么他们的工作没有社会影响或为什么论文没有涉及社会影响。
    - 负面社会影响的例子包括潜在的恶意或意外用途（例如，虚假信息、生成虚假档案、监控）、公平性考量（例如，可能做出对特定群体造成不公平影响的技术部署）、隐私考量和安全考量。
    - 会议预计，许多论文将是基础性研究，而不与特定应用挂钩，更不用说实际部署了。然而，如果有直接通向任何负面应用的路径，作者应当指出。例如，可以指出生成模型质量的提高可能被用于生成用于虚假信息的深度伪造。另一方面，没有必要指出优化神经网络的通用算法可以使人们更快地训练生成深度伪造的模型。
    - 作者应该考虑在技术按预期使用并正常运行时可能产生的危害，在技术按预期使用但给出错误结果时可能产生的危害，以及因（有意或无意）误用技术而导致的危害。
    - 如果存在负面的社会影响，作者们还可以讨论可能的缓解策略（例如，分阶段发布模型，除了攻击之外还提供防御，监控滥用行为的机制，监控系统如何随着时间的推移从反馈中学习的机制，提高机器学习的效率和可访问性）。

14. **Safeguards**

    问题：论文中是否描述了对可能被滥用的数据或模型（例如，预训练语言模型、图像生成器或抓取的数据集）负责任发布所采取的保障措施?

    答案：[NA]

    正当性：论文没有发布具有高误用风险的模型或数据

    指导方针：

    - 答案 NA 意味着论文不存在这样的风险。
    - 对于那些具有被误用或双重用途高风险的发布模型，应当配备必要的防护措施以允许对模型的受控使用，例如通过要求用户遵守使用准则或访问模型的限制，或实施安全过滤器。
    - 从互联网抓取的数据集可能带来安全风险。作者应描述他们如何避免发布不安全的图像。
    - 我们认识到提供有效的保障措施是具有挑战性的，且许多论文并不需要这一点，但我们鼓励作者将其考虑在内，并尽最大努力。

15. **Licenses for existing assets**

16. 问题：论文中使用的资产（例如，代码、数据、模型）的创作者或原始所有者是否得到了适当的认可，并且许可和使用条款是否被明确提及并得到适当的尊重?

    答案：[Yes]

    理由：本文中使用的所有第三方数据集和工具均已正确引用，并在适用情况下说明了许可证。

    指南：

    - 答案 NA 表示论文没有使用现有资产。
    - 作者应该引用生成该代码包或数据集的原始论文。
    - 作者应说明使用的是哪个版本的资产，并且在可能的情况下，提供一个 URL。
    - 每个资源都应该包括许可证的名称（例如，CC-BY 4.0）。

- 对于来自特定来源（例如，网站）的抓取数据，该来源的版权和服务条款应被提供。
- 如果发布了资产，则应该提供包中的许可、版权信息和使用条款。对于流行的数据集，`paperswithcode.com/datasets` 已为某些数据集整理了许可。他们的许可指南可以帮助确定数据集的许可。
- 对于重新打包的现有数据集，既要提供原始许可，又要提供衍生资产的许可（如果它已更改）。
- 如果这些信息在网上不可用，建议作者联系资产的创建者。

17. **New assets**

18. 问题：论文中引入的新资产是否有详细记录，并且这些记录是否与资产一起提供？

    答案：[NA]

    理由：本文未引入需要文档记录的新数据集或模型。

    指南：

    - 答案 NA 表示该论文没有发布新的资产。
    - 研究人员应通过结构化模板在提交时传达数据集/代码/模型的详细信息。这包括关于训练、许可、限制等的详细信息。
    - 论文应该讨论是否以及如何获得使用资产者的同意。
    - 在提交时，请记得对您的资产进行匿名处理（如适用）。您可以创建一个匿名链接或包含一个匿名的压缩文件。

19. **Crowdsourcing and research with human subjects**

20. 问题：对于众包实验和涉及人类受试者的研究，论文是否包括提供给参与者的完整指示文本和截图（如适用），以及有关补偿（如果有）的详细信息？

21. 答案：[NA]

    证明理由：该研究不涉及众包或以人为对象的实验。

    指南:

    - 答案“NA”意味着该论文不涉及众包或涉及人类受试者的研究。
    - 将这些信息包含在补充材料中是可以的，但如果论文的主要贡献涉及到人类受试者，那么在主论文中应尽可能详细地包括这些细节。
    - 根据 NeurIPS 伦理准则，参与数据收集、策划或其他劳动的工作人员应至少获得数据收集国的最低工资。

22. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    问题：论文是否描述了研究参与者可能面临的风险，这些风险是否向参与者披露，以及是否获得了机构审查委员会（IRB）审批（或根据您的国家或机构要求获得的相应批准/审查）？

    回答：[NA]

    理由：研究中没有涉及人体实验对象，因此不适用 IRB 批准。

23. 指南:

    - 答案 NA 表示该论文不涉及众包或以人为研究对象的研究。
    - 根据研究进行的国家的不同，任何涉及人类被试的研究可能需要获得 IRB 批准（或同等机构的批准）。如果您已获得 IRB 批准，您应在论文中清楚地说明这一点。
    - 我们认识到，不同机构和地点在这方面的程序可能会有显著差异，我们期望作者遵守 NeurIPS 的道德规范以及其所在机构的指导方针。
    - 对于初次提交，请不要包含任何会破坏匿名性的信息（如果适用），例如进行评审的机构。

24. **Declaration of LLM usage**

    问题：如果 LLM 是此研究核心方法中一个重要的、原创的或非标准的组成部分，论文是否描述了其使用？请注意，如果 LLM 仅用于写作、编辑或格式化目的且不影响研究的核心方法、科学严谨性或原创性，则无需声明。

回答: [NA]

理由: 在本文的核心方法设计或实现中没有使用 LLMs。它们仅用于一些小的编辑支持。

指南:

- 答案 NA 表示本研究中的核心方法开发不涉及 LLMs 作为任何重要、原创或非标准的组成部分。
- 请参考我们的 LLM 政策 (`https://neurips.cc/Conferences/2025/LLM`) 以了解应该或不应该描述的内容。