

硕 士 研 究 生 读 书 报 告



题目 实时辐射场渲染的3D高斯喷溅方法

作者姓名 陈正和

作者学号 22351050

指导教师 李启雷

学科专业 电子信息

所在学院 软件学院

提交日期 二○二四年一月八日

3D Gaussian Splatting for Real-Time Radiance Field Rendering

A Dissertation Submitted to

Zhejiang University

in partial fulfillment of the requirements for

the degree of

Master of Engineering

Major Subject: Electronic Information

Advisor: Li Qilei

By

Zhenghe Chen

Zhejiang University, P.R. China

2024.1.8

摘要

这篇论文介绍了一种用于实时辐射场渲染的创新方法。该方法通过新颖的3D高斯场景表示和实时可微分渲染器，大幅提高了场景优化和新视角合成的速度。与传统的神经辐射场（NeRF）方法相比，这种方法能够在显著减少训练和渲染时间的同时，保持或提升视觉质量。论文中提出的解决方案依赖于三个核心组件：各向异性的3D高斯表示、3D高斯属性的优化方法，以及用于GPU的快速可微分渲染方法。这些方法的结合使得实时渲染成为可能，同时保持了高质量的视觉效果。

**关键词**：实时渲染； 辐射场； 3D高斯表示； 视觉合成

Abstract

This paper presents an innovative approach for real-time rendering of radiance fields. The method significantly accelerates scene optimization and novel view synthesis by employing a novel 3D Gaussian scene representation coupled with a real-time differentiable renderer. Compared to traditional Neural Radiance Field (NeRF) methods, this approach maintains or enhances visual quality while drastically reducing training and rendering times. The solution proposed in the paper is built on three main components: anisotropic 3D Gaussian representation, an optimization method for these 3D Gaussian properties, and a fast differentiable rendering approach for the GPU. The combination of these methodologies enables real-time rendering while preserving high visual quality.

**Keywords：**Real-Time Rendering; Radiance Fields; Novel View Synthesis

1引言

在这篇论文中，作者介绍了一种用于实时渲染辐射场的新方法。这项研究领域的传统方法依赖于使用网格和点来表示三维场景，这些方法能够适应快速的GPU/CUDA基础栅格化。然而，最近的神经辐射场（NeRF）方法则是建立在连续场景表示的基础上，通常利用多层感知器（MLP）和体积光线行进来进行新视角的合成。尽管这些连续表示方法有助于优化，但所需的随机采样过程在渲染时既昂贵又可能产生噪声​​。

本文的目标是实现对用多张照片捕获的场景进行实时渲染，并以与当前最高效方法相当的优化时间创建表示。近期的方法虽然可以实现快速训练，但在实现当前最先进的NeRF方法所取得的视觉质量方面存在困难。这些快速但质量较低的辐射场方法可以实现交互式渲染时间（每秒10-15帧），但在高分辨率下无法实现实时渲染​。

**2 主旨**

文章提出一种新方法实现了辐射场的实时渲染，能够在较少的训练时间中，实现SOTA级别的视觉效果，针对用多张照片和视频拍摄的场景，允许以 1080p 分辨率进行高质量的实时（≥ 30 fps）新视图合成。

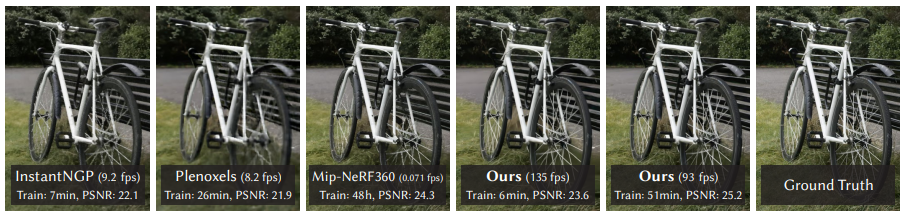


图1 辐射场的实时渲染

具体而言，即从已有点云模型出发，以每个点为中心，建立可学习的3D高斯表达，用Splatting的方法进行渲染，实现了高分辨率的实时渲染，其中包含三个关键步骤：

首先，从相机校准过程中产生的稀疏点开始，文章用3D高斯（3D Gaussians）表示场景，3D高斯保留连续体积辐射场的理想属性以进行场景优化，同时避免了在空白空间中进行不必要的计算。

其次，文章对3D高斯进行了交错优化/密度控制，特别优化了各向异性协方差（anisotropic covariance）以实现场景的准确表示。

最后，文章开发了一种快速可见性感知渲染算法（fast visibility-aware rendering algorithm），该算法支持各向异性抛雪球（anisotropic splatting），既能加速训练，又能实现实时渲染。

**3 相关工作**

**3.1** **针对场景的传统重建和渲染方法：**

Structure-from-Motion（SfM）的出现使得一组照片即可用于合成新视图。SfM可估算相机校准期间的稀疏点云，最初用于3D空间的简单可视化。

multi-view stereo（MVS）的出现，促进了若干视图合成算法的出现，这些方法都将输入图像重投影（re-project）并混合（blend）到新视角相机中，并使用几何来引导重投影。但当 MVS 生成不存在的几何时，算法通常无法从“未重建”或“过度重建”的区域中完全恢复。

最近的神经渲染算法大大减少了此类伪影，同时避免了将所有输入图像存储在 GPU 上，在大多数方面都优于之前方法。

**3.2 针对神经渲染和辐射场**

深度学习与体积光线行进（volumetric ray-marching）相结合的技术，作用在连续可微密度场上，被用于表示几何。但由于查询体积需要大量样本，使用体积光线行进（volumetric ray-marching）进行渲染的成本很高。

NeRF引入了重要性采样和位置编码来提高质量，但使用了大量多层感知器拖慢了模型速度。目前用于新视图合成，具有SOTA图像质量的是Mip-NeRF360，虽然渲染质量非常出色，但训练和渲染时间仍然非常高（文章能够达到或在某些情况下超过这种质量，同时也提供快速培训和实时渲染）。

最新的方法主要通过利用三种设计来专注于更快的训练和/或渲染：使用空间数据结构来存储在体积光线行进期间插值的（神经）特征、不同的编码和 MLP 容量。这些方法包括空间离散化的不同变体，codebooks以及哈希表等编码，并允许使用较小的 MLP 或完全放弃神经网络。

其中InstantNGP使用哈希网格和占用网格来加速计算，并使用较小的 MLP 来表示密度和外观。Plenoxels使用稀疏体素网格来插值连续密度场，并且能够完全放弃神经网络。 两者都依赖于球谐波（Spherical Harmonics）：前者直接表示方向效果，后者将其输入编码到颜色网络。虽然两者都提供了出色的效果，但仍然难以有效地表示空白空间。另外，图像质量在很大程度上受到用于加速的结构化网格选择的限制，并且由于需要为给定的光线行进步骤查询许多样本，渲染速度会受到阻碍。

文章使用的非结构化、显式GPU友好的3D 高斯无需神经组件即可实现更快的渲染速度和更好的质量。

**3.3 针对Point-NeRF和Plenoxels**

**3.3.1 Point-NeRF:**

该方法采用点云进行NeRF渲染，即从点云出发构建辐射场，为每个点提取一个特征储存在点中，渲染时采用NeRF常见的体渲染。

本文也是从点云出发的辐射场渲染方法，但点中存储的是物理含义更加明确的3D高斯，避免了对抽象特征的学习过程，使得训练更易收敛。

**3.3.2** **Plenoxels**

该方法继承了许多NeRF的要素，如体素，三线性插值，体渲染，但抛弃了常用的MLP和隐式特征，采用球谐函数，实现了NeRF加速。

本文在渲染阶段采用Splatting，一种图形学中传统但高效的表达方式，提高了模型表达的下限。

**3.4 核心思想**

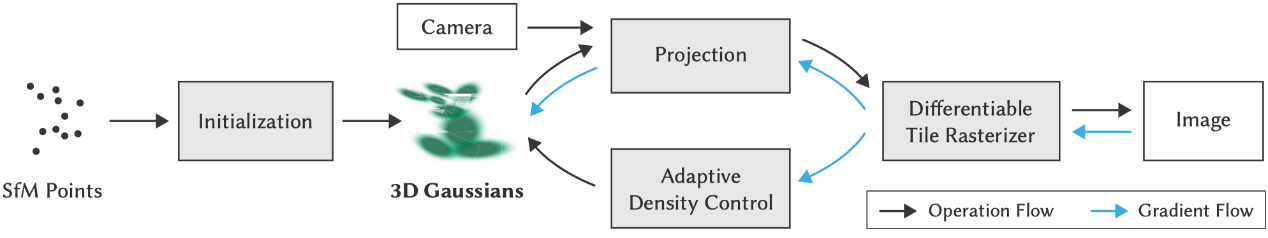


图2 3D高斯分布优化与基于图块的实时渲染流程

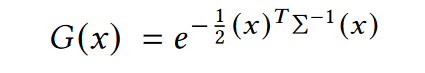
图3.1为整个系统的示意图，系统首先对SfM点云进行初始化，得到3D高斯，然后借助相机外参将点投影到图像平面上（即Splatting），接着用可微光栅化，渲染得到图像。得到渲染图像Image后，将其与Ground Truth图像比较求loss，并沿蓝色箭头反向传播。蓝色箭头向上，更新3D高斯中的参数，向下送入自适应密度控制中，更新点云。

**4 实现细节**

**4.1 可微3D高斯Splatting**

3D高斯点云中的每一个点里面储存了：位置坐标，协方差矩阵（决定高斯形状），不透明度（用于Splatting），球谐函数（拟合视角相关的外观）。

其中3D高斯点云作为一种图元，在拥有场景表达能力的同时可微，且显式地支持快速渲染，其中的参数可在迭代优化的过程中更新。系统采用Splatting的方法可轻易将3D高斯点云投影到2D图像上，实现快速的 α 混合（即渲染）。

3D高斯的定义为:

**4.2 优化与自适应密度控制：**

点云中点存储的参数通过随机梯度下降策略，快速光栅化实现优化。Sigmoid函数将参数中的不透明度限制在 [0, 1) 中。协方差矩阵中的尺度用Exponential激活函数。损失函数为L1*loss*。

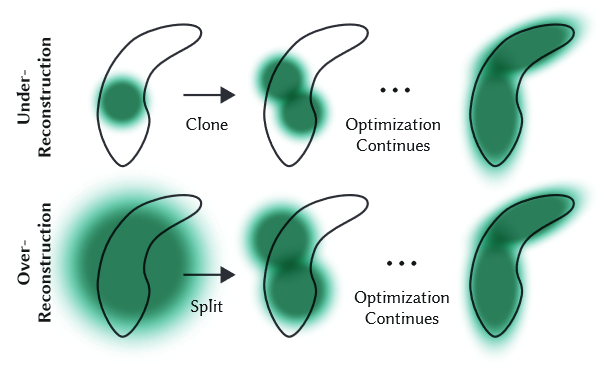


图3 自适应高斯致密化方案

该组件每100次迭代移除不透明度小于阈值的点。并对重建不充分的区域进行处理，针对“欠重建”（under-reconstruction）区域，对点云进行clone，对“过重建”（Over-reconstruction）的区域进行split。判断重建是否充分的依据为梯度，若该点梯度过大，则该位置误差较大，需进行操作，而判断是哪种操作则是通过方差，若方差大，则点过大，需分割，否则克隆。

该组件也周期性将不透明度重置为0，剔除不透明度太低的点，用于去除画面中靠近相机处学出的floaters，并周期性移除较大的高斯用于避免重叠。

**4.3 快速可微光栅化：**

该系统的渲染方法为Tiled-based Rasterizer，其流程为：

1. 将整个图像划分为16\*16个tiles，每个tile视锥内挑选可视的3D Gaussian。
2. 每个视锥内只取执行度大于99%的高斯，并将3D高斯实例化为高斯对象，并按对象中包含的深度信息排序。
3. 并行地在每个tile上splat。
4. 有像素的不透明度达到饱和就停止对应线程。
5. 反向传播误差时按tile对高斯进行索引。

**5 模型表现**

作者展示了他们的方法与之前的方法以及来自测试视图的相应地面实况图像的比较。这些场景从上到下分别是：来自 Mip-NeRF360 数据集的自行车、花园、树桩、柜台和房间； Playroom，来自深度混合数据集的 DrJohnson [Hedman 等人。 2018] 以及 Tanks&Temples 的卡车和火车。箭头/插图突出显示质量上不明显的差异。

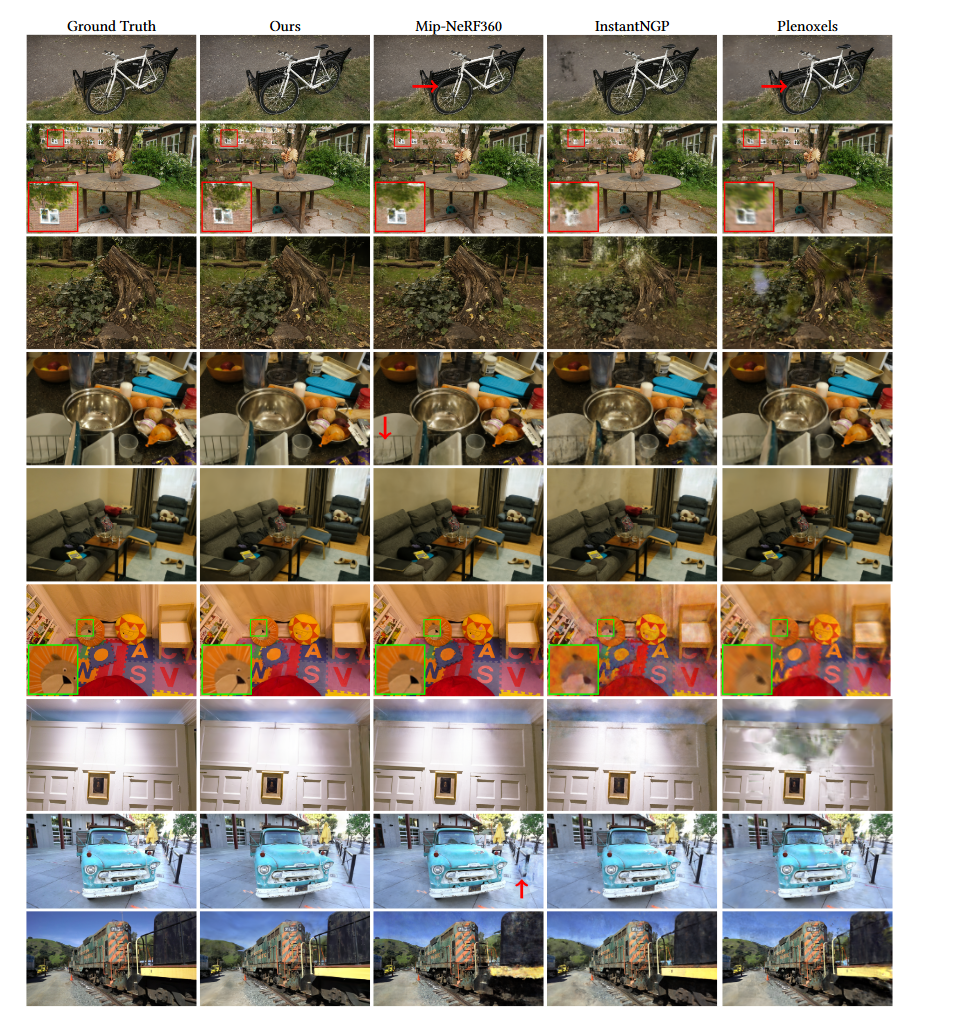


图4 与先前工作对比效果图。



图5 特定场景7K与30K迭代生成图。

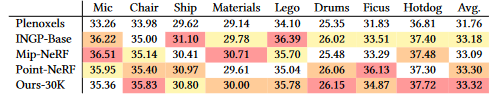


图6 合成 NeRF 的 PSNR 分数

**6 目前限制和展望**

在本文中，尽管提出的技术在多方面展现了优势，但在实际应用中也遇到了一些挑战。特别是在视角不足的区域，容易出现伪影，这主要是由于从有限的视角中重建场景时，信息不足以准确渲染那些细节较少的部分。同样，在反光区域，由于光线的复杂交互，伪影的产生也较为常见。这些问题突显了当前方法在处理特定类型场景时的局限性。

此外，高斯剔除机制和可见性算法的简单性也是导致问题的一个因素。这些算法在处理复杂场景时可能过于粗糙，无法精确处理场景中的细节，从而影响最终的渲染质量。在这方面，引入更复杂的正则化机制可能有助于改善这些问题，特别是在处理伪影和场景中的复杂光照效果时。

对于非常大型的场景，现有方法需要降低位置学习率才能实现收敛，这表明在处理大规模数据时，算法的效率和稳定性仍有待提高。此外，当前实现由于其基于点的方法特性，在显存占用上相对较高，达到了20GB以上的峰值，这在资源有限的环境中可能成为一个制约因素。

总体来看，尽管该方法在实时渲染方面展现了巨大潜力，但针对伪影问题、算法的精细度、以及资源效率方面的改进仍然是未来研究的重要方向。通过对这些挑战的深入研究和解决，有望进一步提升实时辐射场渲染技术的实用性和广泛应用。

参考文献

[1]. Kerbl B, Kopanas G, Leimkühler T, et al. 3D Gaussian Splatting for Real-Time Radiance Field Rendering[J]. ACM Transactions on Graphics, 2023, 42(4).