

**本科毕业论文（设计）**

**题目：** 基于高斯泼溅的三维重建系统设计与实现

学 院： 信息科学与工程学院

专 业： 人工智能

学 生 姓 名： 刘杨

学 号： 632102090923

指 导 教 师： 杨庆

评 阅 教 师：

完 成 时 间：

重庆交通大学

CHONGQING JIAOTONG UNIVERSITY

**本科毕业论文（设计）原创性声明**

本人严肃且认真地声明：所提交的这份毕业论文（设计），乃是本人在导师的悉心指导之下，独立自主开展研究工作从而获取到的相关成果。除去在文中已经明确注明引用来源的那些内容之外，本论文当中并未包含其他任何个人或者集体在此之前已经发表出来或者撰写完成过的相关作品成果。对于那些在本文的研究过程当中曾经做出过重要贡献的个人以及集体，也都已经在论文里面以十分明确的方式一一标明出来了。

本人完全意识到本声明的法律后果由本人承担。

作者签名（亲笔）： 年 月 日

-------------------------------------------------------------------------------------------------

**本科毕业论文（设计）版权使用授权书**

这篇毕业论文（设计）的作者对学校有关保留、使用学位论文的各项规定有着充分的了解。在本科生于校攻读期间，毕业论文（设计）工作所涉及的知识产权归属单位为重庆交通大学。作者同意学校把论文的复印件以及电子版予以保留，同时向国家相关的部门或者机构进行送交，也准许论文能被他人查阅以及借阅。本人还授权给重庆交通大学，使其能够将毕业论文（设计）的全部或者部分内容编入到有关的数据库当中来开展检索工作，并且可以运用影印、缩印或者扫描等复制的方式去保存、汇编毕业设计（论文）。

作者签名（亲笔）： 年 月 日

导师签名（亲笔）： 年 月 日

# 摘 要

随着计算机视觉以及三维重建技术持续取得发展，从多视图图像当中自动去还原出高精度的三维场景，这已然变成了研究领域的一个热点所在。传统的那些用于三维重建的方法，往往是要依靠对图像特征来做精确的匹配，并且要构建出稠密的点云，之后再据此去推断场景的几何结构。这类技术在碰到复杂光照环境或者是反射性材质等属于非朗伯面的情况时，其表现是不太好的，而且也很难在自由视角之下实现那种带有真实感的渲染效果。而基于神经网络的重建方法呢，它又存在着一些问题，比如训练周期比较长、渲染效率比较低，其隐式表示的方式还缺乏可解释性与编辑性，所以就很难在工业领域真正落地去使用。在2023年的时候，Kerbl等人提出了一种三维高斯泼溅方法，也就是3DGS，这个方法有效地对NeRF的效率短板进行了弥补，从而为三维场景建模以及实时渲染开辟出了新的可能性。

此篇论文着重围绕稀疏点云的构建以及优化方面展开探讨，同时也针对以高斯泼溅为基础的高质量三维重建的相关方法来深入开展研究工作，其意图在于搭建起一个能够实现高效化、真实感颇为突出、且具备可微优化特性的三维场景重建框架体系。

起初，本文凭借COLMAP算法以及开源软件来做完图像数据的预处理工作，这里面涵盖了特征提取环节、对特征匹配加以优化的操作，还顾及到增强点云在结构方面的一致性以及空间分布上的完整性情况，最终得以生成稀疏点云数据。在完成上述步骤的前提之下，着手去构建高斯点云表示形式，借助各个点的空间位置、各向异性尺度、透明度还有颜色等一系列参数来开展建模相关的事宜。

随后，本文针对基于3DGS的可微渲染方法展开了细致探讨，将球谐函数引入其中，用以对相关点的反射属性加以编码，进而达成方向感知层面的真实光照建模目的。此方法有着不错的可视化成效，并且在训练环节便于借助反向传播来对高斯点的参数做出优化。最后，论文还就高斯点的初始化策略、训练管线设计以及优化与压缩策略等方面展开了研讨，以此来达成三维重建在轻量化、高精度以及高效率等方面的要求。

本文所构建起来的那个依托于3DGS的三维重建系统，成功达成了快速渲染这一目标，并且还能够实现可交互式展示。如此一来，便给游戏开发领域、数字孪生方面以及虚拟现实等诸多应用开拓出了全新的技术路径。

关键词为三维重建、多视图几何、稀疏点云、高斯泼溅、可微渲染以及球谐函数。

Abstract

As computer vision and 3D reconstruction technologies develop quickly, the automatic reconstruction of high-fidelity 3D scenes from multi-view images has turned into a significant research focus. Traditional reconstruction methods usually depend on precise feature matching and the generation of dense point clouds to deduce scene geometry. Nevertheless, these approaches frequently encounter difficulties under complex lighting conditions or on non-Lambertian surfaces like reflective materials, and they are unable to achieve photorealistic rendering from any arbitrary viewpoints. Reconstruction methods based on neural networks, though showing promise, are confronted with challenges such as lengthy training times and low rendering efficiency. In 2023, Kerbl and others put forward the 3D Gaussian Splatting (3DGS) method, which successfully tackles the performance limitations of NeRF and creates new opportunities for 3D scene modeling and real-time rendering.

This paper zeroes in on the construction and optimization of sparse point clouds. It also deals with the implementation of a high-quality 3D reconstruction framework that is based on 3D Gaussian Splatting (3DGS). The goal is to bring about efficient, realistic, and differentiable scene reconstruction.

First off, COLMAP along with relevant open-source tools gets utilized to handle the preprocessing of image data. This involves tasks like extracting features, optimizing feature matching, as well as enhancing both the structural consistency and the spatial completeness. Through these steps, the generation of sparse point cloud comes about. Building on this foundation, a Gaussian point representation is put together. In this representation, each point is modeled by taking into account its spatial position, anisotropic scale, opacity, and color attributes.

After that, this paper delves into the differentiable rendering method which is based on 3DGS quite thoroughly. It takes spherical harmonics to encode the reflective features of each individual point and makes it possible to carry out direction-aware photorealistic lighting modeling. This particular approach offers high-quality visualization and at the same time helps with the gradient-based optimization of Gaussian parameters when it comes to the training process. The paper also looks into point initialization strategies, the design of the training pipeline, optimization methods, as well as compression techniques, with the aim of attaining lightweight, accurate and efficient 3D reconstruction.

Keywords: 3D Reconstruction, Multi-view Geometry, Sparse Point Cloud, Gaussian Splatting, Differentiable Rendering, along with Spherical Harmonics.

目 录

[摘 要 II](#_Toc1904)

[Abstract III](#_Toc27253)

[1 绪论 1](#_Toc31957)

[1.1 研究背景与意义 1](#_Toc32670)

[1.2 国内外研究现状 2](#_Toc17659)

[1.2.1 三维重建技术研究现状 2](#_Toc22892)

[1.2.2 3DGS研究现状 2](#_Toc17773)

[1.3 研究目标与内容 3](#_Toc16539)

[1.4 章节安排 4](#_Toc8175)

[2 三维重建相关技术综述 5](#_Toc9227)

[2.1 三维重建基础理论 5](#_Toc13151)

[2.1.1 多视图几何 5](#_Toc27801)

[2.1.2 相机模型与标定 7](#_Toc7353)

[2.1.3 常见的三维重建表达方式 9](#_Toc1180)

[2.2基于神经辐射场的体素渲染 10](#_Toc284)

[2.2.1.核心思想与数学原理 10](#_Toc22611)

[2.2.2.NeRF的训练过程： 12](#_Toc28922)

[2.2.3.NeRF 的优缺点 12](#_Toc32463)

[3 图像数据预处理与稀疏点云重建 13](#_Toc30179)

[3.1 基于COLMAP的相机位姿估计 13](#_Toc30492)

[3.1.1 相机模型与投影几何 13](#_Toc10120)

[3.1.2 特征提取与匹配 13](#_Toc944)

[3.1.3 增量式 SfM 重建流程 14](#_Toc18916)

[3.1.4 输出格式与可视化 14](#_Toc15986)

[3.2 稀疏点云生成 14](#_Toc17846)

[3.2.1 三角测量与深度恢复 14](#_Toc4565)

[3.2.2 稀疏点云的特性与局限性 14](#_Toc5542)

[3.3 基于点云神经网络的系数点云优化 15](#_Toc26581)

[3.3.1 点云处理挑战 15](#_Toc21973)

[3.3.2 PointNet架构原理 15](#_Toc9692)

[3.3.3 在稀疏重建中的应用 15](#_Toc20646)

[4 基于高斯泼溅的三维重建方法 16](#_Toc1242)

[4.1 高斯泼溅表示 16](#_Toc31482)

[4.1.1 高斯分布的数学模型 17](#_Toc3861)

[4.1.2 向量化与效率优化 17](#_Toc9025)

[4.2 可微渲染过程 17](#_Toc6523)

[4.2.1 渲染原理 18](#_Toc29186)

[4.2.2 可微渲染计算与梯度传播 18](#_Toc31867)

[4.2.3 引入球谐函数进行光照建模 18](#_Toc4051)

[4.3 参数初始化与训练策略 21](#_Toc12465)

[4.3.1 初始化策略 21](#_Toc9172)

[4.3.2 训练损失与优化目标 22](#_Toc21993)

[4.3.3 多阶段训练策略 22](#_Toc14113)

[4.4 优化与高斯压缩策略 23](#_Toc8005)

[4.4.1 点密度增长相关策略，也就是所谓的密度引导分割（Density-Guided Splitting）这一情况。23](#_Toc11494)

[4.4.2 点的合并以及压缩策略（也就是Gaussian Pruning）23](#_Toc13706)

[4.4.3 性能优化与加速策略 23](#_Toc15348)

[4.4.4 压缩率与质量的权衡 24](#_Toc26850)

[5 三维重建系统设计与实现 24](#_Toc32259)

[5.1 系统需求与分析 24](#_Toc5522)

[5.1.1 功能需求 24](#_Toc2567)

[5.1.2 技术需求 25](#_Toc31600)

[5.2 后端模块设计 25](#_Toc31307)

[5.3 前端渲染模块设计 25](#_Toc1579)

[5.3.1 场景构建与数据加载 25](#_Toc1059)

[5.3.2 高斯泼溅可视化实现 26](#_Toc13469)

[5.3.3 用户交互与界面设计 26](#_Toc13947)

[6 总结与展望 26](#_Toc32690)

[6.1 全文总结 26](#_Toc23945)

[6.2 展望 27](#_Toc26044)

[致 谢 29](#_Toc15726)

[参 考 文 献 30](#_Toc9508)

# 1 绪论

## 1.1 研究背景与意义

三维重建技术在计算机视觉以及图形学领域属于核心的研究走向，其要做的就是从二维图像或者传感器数据当中把三维场景的几何结构以及外观信息给恢复出来。传统的一些方法呢，主要是运用显式几何表示的方式，像是点云、网格、体素这些，或者也会采用隐式神经表示，就比如神经辐射场NeRF，不过这两种方式其实都有着比较明显的局限之处呀。

显式几何表示固然可以对场景结构予以直接表达，不过呢，在针对复杂材质以及光照变化展开建模工作的时候，它是存在着诸多不足之处的。而且呀，它还对多视图间的精确配准有着高度的依赖性，如此一来呢，就使得整体的流程变得颇为复杂，计算效率也呈现出低下的状况。与之相较而言，隐式神经表示是借助多层感知机来对辐射场实施隐式编码的，通过这样的方式能够生成那种高保真的渲染效果。然而呢，它的训练过程是极为耗时的，渲染速度也相当缓慢，并且还难以对场景编辑以及实时交互提供有力的支持，这一系列的情况也就限制了它在工业场景当中的实际应用程度呀。

在这样的背景之下，三维高斯splatting（也就是3DGS）技术被引入进来，这已然成为了可微渲染领域当中极为重要的一个突破点。3DGS它是利用显式高斯点云的方式来针对场景展开建模工作的，并且还和可微分光栅化管线相互结合起来。如此一来，它不但达成了能与NeRF相媲美的渲染质量，而且还能把训练时间大幅缩减，缩减到了分钟这样的级别，与此同时，其渲染帧率也得到了极大提升，能够提升至数百FPS，这般操作下来，计算效率以及场景可控性都有了明显的提高。靠着快速渲染以及轻量化这两大优势，该技术已经快速地在多个领域当中得到了应用，有力地推动了像移动增强现实（AR）、数字文化遗产等这些实际应用不断向前发展。

随着元宇宙、数字孪生以及智能机器人这些前沿领域不断快速发展，实时高保真三维重建的需求正变得日益迫切起来。就比如说，工厂场景的三维重建工作得在秒级时间内完成，并且还要能够支持交互式编辑操作；在医疗领域当中，为了能够还原器官的解剖结构，就要求达到亚毫米级的重建精度；而对于移动应用而言，则需要那种具备低功耗特点的轻量化算法，像3D Moment App就已经实现了用户级别的三维内容创作这件事。

总的来讲，3DGS所构建的三维重建系统，一方面有着重要的理论层面的研究价值，另一方面也实实在在地给实际应用给出了高效的解决办法。它的不断发展，是很有希望促使下一代三维感知、交互以及生成技术取得突破的。所以说，去展开针对3DGS技术的研究以及完成系统的实现工作，这在学术方面有着重要意义，在产业方面也有着广阔的发展前景。

## 1.2 国内外研究现状

### 1.2.1 三维重建技术研究现状

三维重建属于计算机视觉跟图形学之间极为重要的交叉领域方向，一直以来都是运用结构光、双目视觉以及激光雷达（LiDAR）等多种多样的方法去达成对场景三维信息的重建工作。在这之中，运动恢复结构（Structure from Motion, SfM）还有多视图立体（Multi-View Stereo, MVS）这两项技术实际上构成了传统三维重建流程里面的核心部分。这类方法大体上是要依靠图像特征的精准匹配操作，进而构建出稠密的点云，再依据这些来推断出场景具体的几何结构。不过呢，要是处于那种复杂的光照环境之下，或者是面对具有反射性、非朗伯特（non-Lambertian）材质的情况时，这些技术所展现出来的效果往往就不是那么理想了，并且想要在自由视角下实现那种具有真实感的渲染也是比较困难的事情。

当可微渲染（Differentiable Rendering）这一概念被提出来之后，研究者们便着手借助神经网络，采用端到端的方式去对三维结构和图像的一致性加以优化。在这当中，神经辐射场（Neural Radiance Fields, NeRF）的诞生可以说是一个极为关键的转折点。NeRF利用一个多层感知机（MLP）网络，能够把空间里的任意一个点映射成为相应的颜色以及密度值，并且与体积渲染过程相结合，进而从静态图像当中完成对连续场景表示的重建工作。这种方法在新视角图像合成这个方面已然取得了颇为显著的进展，在建筑建模、虚拟现实、文化遗产数字化等等诸多领域都得到了广泛的应用。

不过，NeRF模型明显存在着两个局限之处：其一，该模型所采用的隐式表示方式在可解释性以及可编辑性方面是有所欠缺的；其二，它需要依靠大量的光线采样以及前向网络计算，如此一来便使得训练周期被拉长了，而且渲染效率也比较低，很难去满足实时渲染以及交互式应用所提出的要求。针对这些存在的问题，学术界开始一步步地尝试引入一些更为高效的场景表示形式，像是体素网格（Voxel Grids）、有符号距离函数（TSDF）这类显式结构。虽然在效率与质量方面已经取得了一定的突破，可是在细节表达和性能二者之间依然需要去做一番权衡才行。

### 1.2.2 3DGS研究现状

在2023年的时候，Kerbl等人提出了三维高斯泼溅方法，也就是大家所熟知的3D Gaussian Splatting，简称为3DGS。这一方法成功弥补了NeRF在效率方面存在的不足，给三维场景建模以及实时渲染开拓出了新的发展方向。该方法是把场景通过一种方式表示出来，具体而言，是将其表示为一组带有高斯分布特性的点，并且运用了可微光栅化渲染管线，如此一来，便极大地提高了训练的速度，在推理效率方面也有明显提升，与此同时，还很好地保留了高质量的渲染效果。和NeRF不一样的是，NeRF需要针对整条光线展开密集的采样操作，而且还要多次调用MLP来进行推理，而3DGS则是直接把高斯点投影到图像平面之上，然后再对其进行加权融合，通过这样的操作，有效地削减了计算方面的开销。

3DGS的显式表示一方面增强了模型的可控性，另一方面还便于和传统图形学管线集成起来，进而拓展了它在实际应用当中的可行性，像城市级别场景、动态重建以及SLAM系统等方面都是如此。就拿VastGaussian来说，它引入了依据可见性判断的空间划分策略，从而实现了对大规模城市场景的高效重建。而4D-GS则是把HexPlane编码与轻量级MLP结合到一起，达成了对动态场景的时空建模。GS-SLAM通过高斯点的自适应更新机制，在实时定位与地图构建方面，跟NeRF-SLAM相比，展现出了更为出色的精度与速度表现。

虽说当下3DGS技术已经有了颇为明显的发展进程，可依旧存在着一些有待攻克的挑战。就好比说，在那种大规模的场景状况之下，针对高斯点展开的存储工作以及相关管理事宜，是很有可能会引发较为明显的内存方面的耗费情况的；而处于像复杂材质、镜面反射这类并非十分理想的场景当中的时候，它在针对光照物理展开建模的这一能力层面上，还是有进一步提升的必要的；除此之外，当下其在语义理解方面所具备的能力还是比较有限的，如此一来，便极有可能致使在开展虚拟世界建模这项任务的过程当中，在几何与语义两者之间出现不相符、不一致的情况。为了能妥善处理这些存在的问题，相关的研究者们正在努力尝试着从像是模型压缩、物理建模以及多模态语义融合等这样的一些不同角度出发，进而提出一些能够起到改进作用的方案。就拿LightGaussian来说吧，它借助八叉树结构来对高斯点集进行压缩处理，以此达到降低内存消耗的目的，而DrivingGaussian则是通过独立开展建模工作并且确保时间上的一致性，从而实现了对动态对象进行精确渲染的效果，进一步拓宽了3DGS技术在实际应用当中的范围边界。

从整体情况来讲，3DGS技术已然变成了在三维重建以及新视角渲染领域里的研究热点所在。它在重建精度这块，还有效率方面以及可控性等方面所具备的优势，实实在在地为构建下一代真实虚拟环境打下了牢固的根基。

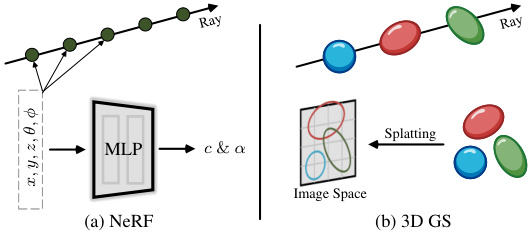


图1.2 Nerf和3DGS原理示意图

## 1.3 研究目标与内容

本研究着手对依靠三维高斯泼溅（3DGS）技术的三维重建系统展开探索。把传统图形学管线和现代神经渲染方法融合起来，以此来促使大规模三维场景建模以及实时渲染的效率、质量都能有所提升。该研究的关键目标是要处理现有方法在复杂场景当中所存在的效率、可扩展性以及适应性方面的诸多问题，进而提出一种全新的混合式三维重建的办法，目的是达成更高精度的三维重建成果，同时让系统的计算性能得以优化，尤其是在动态场景重建以及实时渲染这些方面更是如此。

本研究着重对3DGS技术的核心算法展开研究，同时剖析它在不同场景当中的具体表现，进而提出与之对应的优化方案。会细致探讨高斯点生成、存储、渲染以及合成的整个过程，还会摸索怎样借助算法的优化来提高重建以及渲染的效率。设计出一套完整的三维重建流程，涵盖从数据采集、做预处理，再到高斯点生成以及可微渲染等环节，以此来达成一个高效的三维重建框架。研究在大规模场景里面计算成本和重建精度二者间的平衡状态，确保系统能够高效地运转起来。

本研究还打算探讨怎样把深度信息、纹理信息以及光照模型等多模态数据引入到3DGS当中，以此强化系统针对复杂场景予以理解以及表达的能力。研究把语义分割、光照模型还有材质估计技术相互结合起来，促使三维重建的结果朝着更高精度以及更高真实感的方向去发展。除此之外，还提出一种依据分层存储与空间划分数据结构所制定的优化方案，该方案是用来优化大规模场景重建过程里的内存管理以及渲染性能的，并且要保证实时渲染的性能不会受到任何影响。本研究还会去探索点云优化、并行计算以及GPU加速等技术在场景重建方面的具体应用。

## 1.4 章节安排

这篇文章着重针对运用三维高斯泼溅（3D Gaussian Splatting，简称为3DGS）技术的三维重建系统展开相关的研究工作。在研究过程中，把通过COLMAP工具所获取到的稀疏点云数据加以整合，与此同时，还对高斯渲染模型予以分析并做出优化处理，如此一来，便达成了对三维场景进行可视化重建的目标，而且还能够在Web端实现实时展示的效果。整篇文章总共划分成了六个章节，其具体的章节安排情况如下所示。

首章是绪论部分。在这一章当中，详细讲述了本课题所涉及的研究背景，同时也点明了其具备的重要意义。还对国内外在三维重建以及3DGS相关方面的研究实际现状展开了阐述，进而清晰明确了本论文具体的研究目标，也确定了主要的研究内容。此外，对各个章节的结构安排情况也做了一番大致的概述。

第二章讲的是三维重建相关技术综述内容。在这一章里，系统地对和本课题联系很紧密的基础理论展开了回顾工作，像多视图几何、相机模型以及标定方法等都涵盖在内。并且还简单地介绍了那种基于体素的神经渲染方法，就比如NeRF之类的，同时也提及了它的发展趋向，从而给后续章节在技术选型方面给予相应的理论方面的支撑。

第三章所聚焦阐述的内容为图像数据预处理以及稀疏点云重建方面。在这一章节当中，着重对初始数据的预处理具体流程展开了介绍，详尽地说明了借助COLMAP工具来实施特征提取、进行相机位姿估计以及完成稀疏点云生成的整个过程，除此之外，还更进一步地针对点云优化方法以及该方法在稀疏重建领域当中的实际应用情况展开了探讨。

第四章内容聚焦于以3DGS为基础的三维重建方法，此部分属于论文核心所在。在这一章里，对3DGS的建模以及训练方法展开了较为系统的阐述。其中涵盖了对高斯分布进行数学建模的相关内容，也涉及可微渲染原理方面的情况，还包含球谐光照建模的具体事宜，并且对于训练损失设计以及多阶段优化策略也有所论述，另外还探讨了点云压缩与性能优化技术等方面的内容。

第五章讲的是三维重建系统的设计以及实现方面的内容。在这一章当中，把整个三维重建系统具体是怎样实现的过程给介绍了出来。主要呢，是依据系统功能方面的需求来着手进行相关操作的，对前后端的架构展开了设计工作。对于后端，细致地阐述了其文件管理以及任务调度所遵循的逻辑。而前端呢，围绕着基于React + Three.js的场景渲染的流程、高斯点可视化展示的情况，还有用户交互界面是怎样设计的等方面，都一一做了说明。

第六章讲的是总结以及展望相关内容。在这部分内容里，先是把本论文所做的各项研究工作都做了一番总结，接着对系统实际呈现出来的效果以及存在着的那些不足之处展开了细致分析，最后还从像是高质量重建呀、编辑功能进一步拓展啊以及系统如何去部署等等这些不同方面，针对未来的研究方向给出了相应的展望。

# 2 三维重建相关技术综述

三维重建技术其实就是依据多维信息来推导出三维几何形状这么一个过程。近些年来，计算机视觉、图形学以及深度学习都在快速地发展着，在这样的发展态势之下，三维重建的方法也随之不断地被优化，同时其应用范围也不断地得到拓展。三维重建在虚拟现实（VR）、增强现实（AR）、机器人导航以及自动驾驶等诸多领域当中，都有着极为重要的应用价值，并且在数字孪生、医学影像以及文化遗产保护等方面，同样也发挥着十分关键的作用。

若想达成高质量的三维重建目标，那就得熟知把图像信息转化成三维模型的办法，而这其中会涉及到诸如多视几何、相机模型及其标定、点云以及网格表示等好些个交叉学科的基础理论方面的知识。

## 2.1 三维重建基础理论

三维重建所涉及的基础理论，给从多视图图像里去恢复三维场景的几何结构这件事，搭建起了颇为坚实的理论框架。在这部分内容当中，会对三维重建里面的三个极为重要的核心理论展开介绍，这三个理论分别是多视图几何、相机模型与标定，还有那些常见的三维重建表达方式。

### 2.1.1 多视图几何

多视图几何（Multi-view Geometry）属于计算机视觉领域里极为重要的一个研究走向。它主要是凭借从各不相同的视角所拍摄下来的多张图像，进而去复原出真实世界当中场景所具有的三维结构，同时也能得出摄像机的运动轨迹。该项技术已然成为了现代三维重建、增强现实、无人驾驶导航以及文物数字化等诸多应用所依靠的最为核心的基础所在。

本质上而言，多视图几何是依托二维图像，凭借图像彼此间的几何对应关联以及投影模型，来反向推断出三维空间里物体的形状、所处的相对位置还有相机参数。它构建起了图像空间和三维空间相互间的映射关联，在三维计算机视觉领域充当着桥梁的角色。

多视图几何主要包括以下几个核心技术模块：

**1.特征提取与匹配：**

多视图几何开展的首要步骤便是从图像当中提取出既稳定又颇具描述性的局部特征点。在实际操作中，常用的特征提取算法有SIFT（也就是尺度不变特征变换）、ORB以及SURF等等。将特征提取出来之后，这些特征就会被应用到图像之间的匹配工作当中，借助这样的匹配操作，便能够在图像之间建立起点的对应关系。而这些所建立起来的对应点，实际上是后续开展几何推理以及进行三维重建工作的重要基础所在。

**2.立体匹配与视差估计：**

立体视觉（Stereo Vision）实则为多视图几何范畴里的一个基础性模型。一般来讲，它会以两张或者多张从各不相同角度所拍摄的图像作为依托，凭借特征点或者像素相互间的匹配操作，进而对图像之间存在的视差加以计算。要知道，图像里的同一个点在不一样的视角之下，其投影位置也是存在差异的，而这样的视差是能够被用来反向推算出该点所具有的深度方面的信息的。通常情况下，视差要是越大的话，那么相应的点距离相机也就会越近一些。并且，视差图自身的质量状况，是会对后续开展的三维重建工作的精度起到直接性的决定作用的。

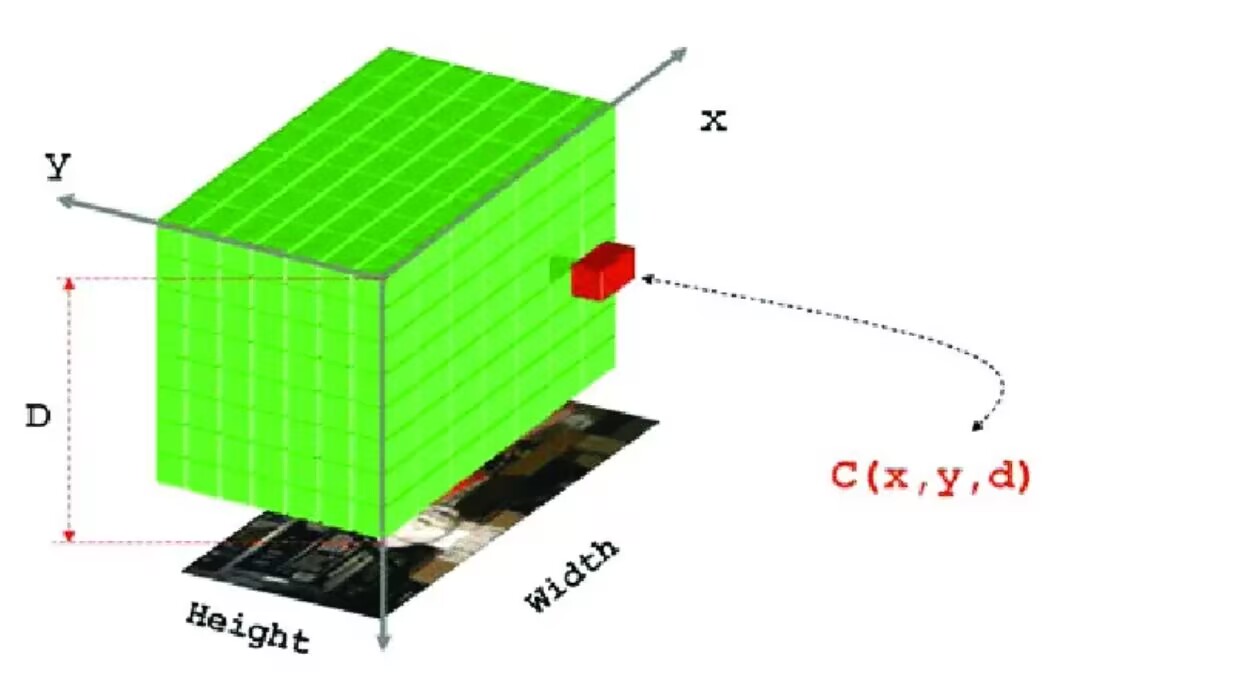


图2.2 视差空间示意图

**3.视图几何约束关系：**

图像间存在的几何关系能够凭借一系列数学工具予以刻画。其中颇为典型的有基础矩，它主要描述的是两个视图里对应点所具有的约束，并且满足极线几何关系；还有本质矩阵，在已知相机内参的情形下，其可用来描述相机彼此间的相对运动状况；再者是单应性矩阵，它所呈现的是同一平面上各点之间的映射关系，在平面场景的配准方面经常会被用到；另外极几何则是描述一对图像之间存在的几何对应关系，是展开匹配以及约束搜索极为重要的依据。这些几何约束关系能够借助RANSAC这类稳健估计方法，从匹配点对当中计算出来，进而可以被进一步用于三角测量以及相机姿态的估计工作当中。

**4.三角测量与三维点恢复：**

当图像间的匹配关系得以建立起来，并且相机的相对姿态也被成功获取到之后，便能够凭借三角测量这种方法，把二维图像里面那些相互匹配的点，反向推算出它们在三维空间当中所处的具体位置。三角测量所依据的基本原理呢，就是要去运用在两个视图里同一个三维点的投影情况，通过几何交会的方式来求解出这个点的空间坐标。从数学层面来讲的话，这一过程是可以经由求解一组线性方程组，又或者是通过最小化重投影误差的途径来达成的。最终所得到的三维点的集合，就构成了相应场景的稀疏点云，而这也为接下来要开展的表面重建工作或者是对三维模型进行优化等操作，预备好了基础方面的数据。

**5. The Structure from Motion (SfM) related to motion recovery:**

SfM is a classic method in multi-view geometry. Its aim is to jointly recover the three-dimensional scene structure (Structure) and the motion trajectory of the camera (Motion) from multiple images without the need to know the camera pose in advance. The core processes of it involve extracting image features and conducting global matching; estimating the camera pose either incrementally or globally; generating sparse point clouds through triangulation; performing global nonlinear optimization by means of bundle adjustment to enhance the reconstruction accuracy. SfM generally outputs sparse three-dimensional point clouds, the extrinsic matrix (pose) corresponding to each image, and the intrinsic parameters, which lay the foundation for subsequent dense reconstruction (such as MVS) and modeling. The system in question utilizes COLMAP to obtain sparse point cloud data. COLMAP is an open-source three-dimensional reconstruction toolbox based on Structure-from-Motion (SfM) and Multi-View Stereo (MVS).

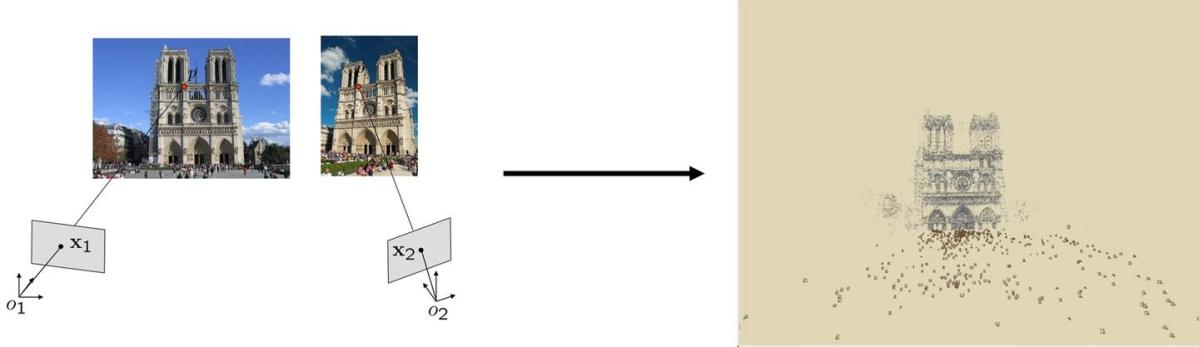


图2.3 运动恢复结构示意

**6.多视图几何的实际应用：**

多视图几何乃是三维视觉的根基所在，其在诸多实际领域均有着十分广泛的运用情况。就拿摄影测量以及地图制作来讲，能够借助航拍图像来对大范围的地形展开重建工作；在机器人视觉与同步定位与地图构建（SLAM）方面，可达成视觉里程计以及三维定位的目标；对于增强现实（AR）和虚拟现实（VR）而言，能够给予精确的场景建模以及相机追踪方面的支持；在文化遗产数字化领域，会针对文物、古迹等这类物理对象实施高精度的三维采集操作；而在自动驾驶与环境感知这块，能够从多个摄像头那里获取到车辆周边环境的三维结构相关信息。

### 2.1.2 相机模型与标定

相机模型以及标定算得上是三维重建系统里面的基础性环节啦，它最主要的目的呢，就是要精准地去描述在三维世界当中的那些点到底是怎样经由摄像机而投影到二维图像之上的。唯有相机模型是准确无误的，并且标定的结果也是足够精确的，如此才能够确保后续在进行几何恢复以及三维重建的时候可以达到相应的精度要求。所以呀，相机模型还有标定方法可不单单是在图像投影计算方面所依靠的数学依据，它们同时也是在对三维重建系统进行误差控制的时候极为关键的所在呢。

（一）相机模型的基本概念

相机模型的作用在于对现实相机的成像过程予以模拟，具体来讲就是去阐述三维空间里的点是怎样借助透视投影而映射至二维图像平面之上的。在计算机视觉领域当中，针孔相机模型（Pinhole Camera Model）属于最为常用的那一类模型。

（1）针孔相机模型

针孔模型属于一种理想化的简化模型，它把镜头那些复杂的光学结构以及畸变效应都给忽略掉了，仅仅留存下了投影变换当中最为核心的那部分内容。此模型假定成像的这个过程是线性透视投影的情况，在这种情况下，三维的点是会被投影到二维的图像点之上的，而它们二者之间的关系是能够凭借以下的投影矩阵来予以表示的。

其中：

为相机内参矩阵，包含焦距、主点、像素比例等；

为相机的外参矩阵，描述从世界坐标系到相机坐标系的变换；

表示齐次坐标的等价变换（即忽略比例因子）。

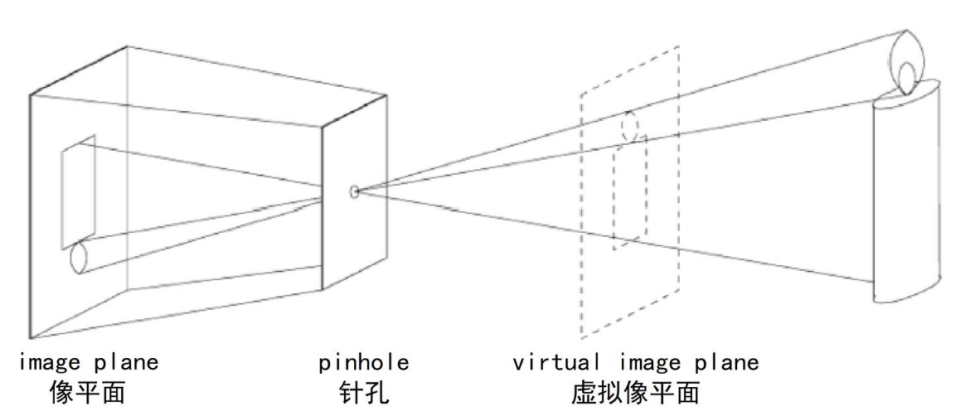


图2.4 针孔照相机模型示意

（2）畸变模型与广角镜头建模

现实世界里的相机镜头有着多种不同类型的成像畸变情况，在运用广角镜头以及鱼眼镜头这类并非理想成像系统的时候，这种成像畸变就显得格外突出。通常而言，常见的畸变类型主要有：

径向畸变：图像边缘位置会被拉伸或压缩，形成桶形或枕形失真；

切向畸变：由于镜头与图像平面不完全对齐导致的非对称畸变。

若想解决由这类畸变引发的误差情况，往往会在相机模型里引入诸如等这样的畸变系数，且在对图像进行处理的阶段开展图像去畸变的相关操作，以此来还原出真实图像原本的几何结构。

（二）相机模型中的坐标系

相机模型里是存在着4个坐标系的，分别为世界坐标系、摄像机坐标系、图像物理坐标系以及图像像素坐标系。所谓光心，指的就是图像的中心位置。而光轴呢，它是这样一条轴，是穿过图像的光心且和图像平面保持垂直的轴。

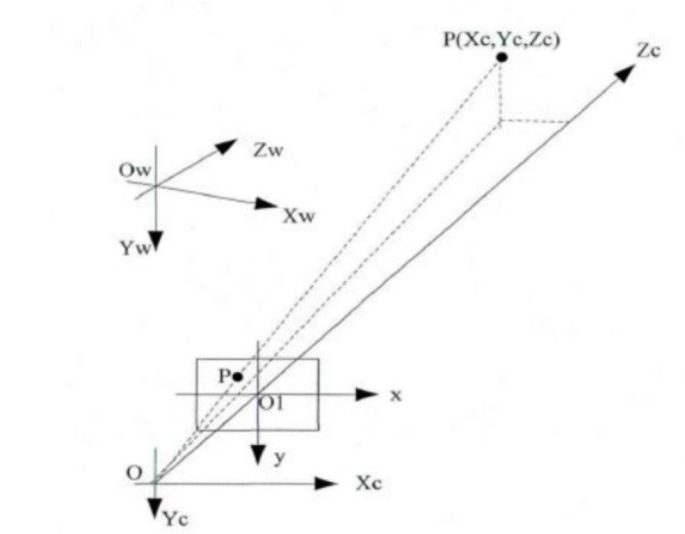


图2.5 相机模型中的坐标系

世界坐标系属于客观三维世界的绝对坐标系，又被称作客观坐标系，其表示的就是物体在真实世界里的坐标。世界坐标系会依据物体的大小以及位置的改变而发生变化，其单位为长度单位。

相机所采用的坐标系，其坐标原点设定在相机的光心位置。就其坐标轴而言，x轴与平行于图像的x方向保持一致，y轴则和平行于图像的y方向相同，而z轴是与光轴相互平行的。并且这x、y、z三轴彼此间呈垂直状态，其所用的单位为长度方面的单位。

图像物理坐标系的坐标原点设置在主光轴Zc和图像平面的交点处，其x方向以及y方向就如同所展示的图示那样，而它所采用的单位为长度方面的单位。

图像像素坐标系的情况是这样的，其会以图像的上顶点、下顶点、左顶点或者右顶点来当作坐标原点。而其中的u方向是和x’方向相平行的，v方向则与y’方向相平行，并且其单位为像素。

相机模型以及标定算得上是三维重建系统在理论方面的根基所在。准确的相机模型，一方面能够给出高质量的图像几何校正，另一方面还能给后续诸如SfM、MVS、3DGS等重建算法送去精确的几何方面的输入内容。在实际的系统当中，相机若能实现良好的标定，那么往往就会呈现出更高的重建精准程度，误差积累的情况也会更少，同时还能给三维高斯泼溅这类神经渲染技术给予可靠的初始化相关条件，并且能确保视角方面的一致性。

### 2.1.3 常见的三维重建表达方式

依据各异的重建办法以及数据所具备的特性，三维重建得出的结果能够运用多种多样不同的数据结构以及表达形式。常见的用于三维重建的表达方式包含有：点云（Point Cloud），还有网格模型（Mesh），再者体素（Voxel Grid），另外深度图（Depth Map），以及神经场（Neural Field），甚至三维高斯泼溅（3D Gaussian Splatting）等等。

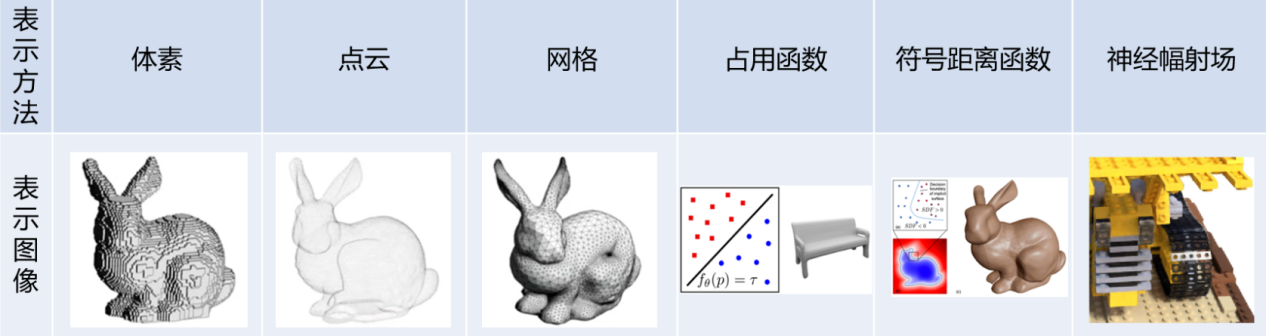


图2.6 常见的三维表达方式

**1、点云（Point Cloud）**

点云属于一种极为基础且通用性很强的三维表示形式，它是由数量众多的、处在三维空间里的离散点所构成的。其中，每一个点都会对其自身所在的空间位置，也就是(x,y,z)坐标进行记录，而且还能够附带像颜色、法线或者语义之类的一些属性呢。

其结构相对简单，能够较为容易地从诸如SfM、MVS这类多视图几何算法当中直接获取相关内容；并且其对于大规模场景可做出稀疏表示。不过它存在不足，缺少拓扑结构，所以无法表示出连续的表面信息；而且在进行渲染以及碰撞检测操作时，效率方面表现得比较低。它在诸多方面有所应用，像激光雷达扫描、SfM的初始输出环节、3D重建的初步展示阶段等等。

**2、三角网格（Mesh）**

三角网格是通过把点云中的那些点依照特定的拓扑关系连接起来，进而形成面片（一般来说这些面片通常呈现为三角形的形态），以此来构建出表面的。通常情况下，一个网格是由顶点的集合、边的集合以及面的集合共同组成的，其能够较为出色地对物体的几何形状予以表达。它有着较强的表达能力，能够被应用在精细建模方面，也能在纹理映射方面发挥作用，还可用于物理仿真等领域；而且主流的图形引擎对于Mesh也给予了很不错的支持。不过呢，网格的构建过程是比较复杂的，其重建的过程对于噪声极为敏感；并且它不太适宜用来表示体积方面的数据。它大多是在游戏建模、三维打印、仿真建模以及工业逆向工程等诸多方面得到应用的。

**3、体素网格（Voxel Grid）**

体素（Voxel）其实就是在三维空间当中被均匀划分出来的那种小立方体单元，这和二维图像里的像素是能够类比起来的。体素网格常常会被用于对实体物体或者体积密度场展开重建工作。它有着规则的结构，在处理以及分析方面都是比较容易着手的；也很适合用来表示体积对象，像是医学图像呀、构造块模型等等之类的情况都是如此。不过呢，它存在内存开销比较大的问题，而且分辨率方面也会受到一定限制；在去表示那些细节特别丰富的场景的时候，其效率往往是比较低的。大部分情况下它主要应用在医学CT/MRI建模、机器人路径规划以及体绘制（Volume Rendering）等方面的相关事宜当中。

**4、深度图（Depth Map）**

深度图属于二维图像范畴，算是它的一种拓展形式。在深度图里，每一个像素所具有的数值，实际上是在表明与之相对应的点到相机之间的深度方面的数值。要是把多张深度图拿来进行融合操作的话，那么就能够重新构建出三维模型来。它和图像在结构方面是保持一致的状态，采集起来比较方便，处理起来也不会太困难。而且它还很契合多视图融合以及立体视觉计算方面的相关需求。不过呢，它存在不具备空间一致性这样的情况，所以在进行融合的时候，难度是比较大的。并且它对噪声还特别敏感，稍微有点噪声干扰，可能就会影响到它的相关表现。它在很多时候主要是应用在像RGB-D相机采集、AR场景定位、SLAM系统、MVS重建输入等等这些方面。

**5、神经场（Neural Fields）**

神经场借助神经网络来对空间展开隐式建模操作，就像神经辐射场（NeRF）这类技术，便是凭借坐标-特征映射函数去构建整个三维场景的模型。如此一来，能够达成连续且无网格的高质量重建效果，其渲染出来的效果很逼真，还能支持新视角的合成工作。不过呢，它存在一些不足之处，比如训练所需要花费的开销比较大，渲染的效率又偏低，而且很难实现实时交互，也不太容易导出显式几何。

**6. The 3D Gaussian Splatting.**

三维高斯泼溅属于一种新出现的神经渲染表达方式。它是在三维空间里去分布高斯核的，这里面每个高斯都带有位置、方向、颜色以及透明度等方面的属性，然后再对其开展光栅化渲染操作，如此一来便能够达成近似于真实情况的视图合成成效。其重建的速度比较快，能够达成高质量的实时渲染效果；它还对显式空间结构予以支持，这对于交互而言是很便利的。不过它的精度会受到输入图像质量的限制；在结构较为稀疏的区域当中，有可能会出现冗余或者模糊的情况。它大多在快速三维扫描、人脸建模、轻量级数字孪生以及移动端AR渲染等方面得到应用。

## 2.2基于神经辐射场的体素渲染

### 2.2.1.核心思想与数学原理

神经辐射场（Neural Radiance Fields，简称为NeRF）属于一种借助神经网络来对三维场景加以建模的技术，它最初是在2020年的时候由Mildenhall等人所提出的。NeRF最为核心的理念在于：凭借一个多层感知器（也就是MLP）去学习一个函数，把三维空间当中的位置以及视角对应地映射到颜色以及体密度之上，以此达成对真实场景展开建模并且对新视角图像予以合成的目的。NeRF并非是直接输出那种显式的几何形状（比如点云、网格之类的），而是以一种‘体积渲染’的形式来间接地呈现出三维结构，这在神经渲染这个领域可以说是一项极为重要的突破。NeRF其关键所在正是一个神经网络函数：

（2.2）

其中：

：表示空间中的三维位置；

：表示光线的方向；

：RGB 颜色；

：体密度，用于体积渲染。

神经网络所承担的任务在于，针对空间里的任意一处点以及任意一个视角，要能够输出该点相应的颜色以及其密度情况。

【NeRF】并非直接对像素展开渲染操作，而是对光线穿过体积时所产生的累积效果予以模拟。它所涉及的颜色计算，其公式是这样的：

N：采样点数量；

：第 i 个采样点的采样间隔；

：第 i 个采样点的密度；

：第 i个采样点的颜色；

：第 i 个点未被遮挡的概率。

所谓体渲染，直观来讲，我们清楚相机是存在焦点的，将焦点和像素连接起来便能得出一条射线，针对这条射线上全部的点，我们能够通过做特定的求和操作，如此一来便可以获取到这个像素的颜色值。从理论层面看，对于这条射线所经过空间上的各个点，其密度仅和空间坐标存在关联，而颜色则同时取决于空间坐标以及入射角，针对这些情况做某种积分处理，就能得出每个像素的颜色。一旦每个像素的颜色都被计算好，那么在这个视角下的图像也就被成功渲染出来了。

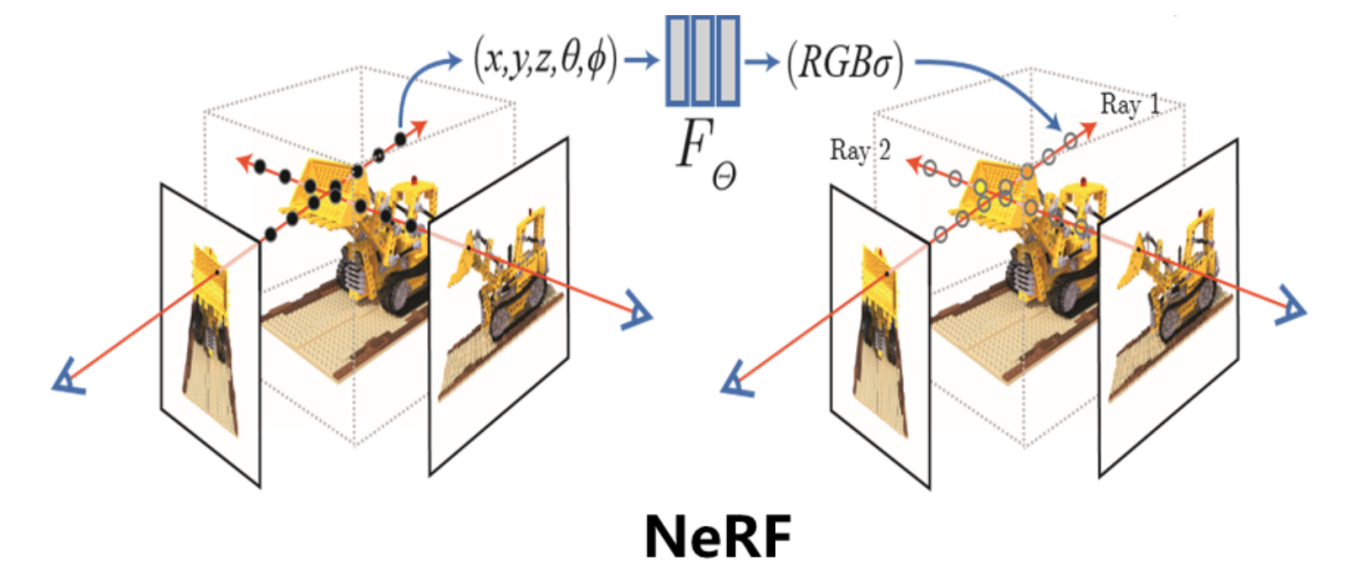


图2.7 NerF渲染原理示意图

### 2.2.2.NeRF的训练过程：

NeRF 的训练数据通常是多张具有已知相机位姿的图像。

训练步骤包括：

1. 相机参数估计：使用 COLMAP 等工具获取相机内参与位姿；
2. 射线生成：每个像素反投影为一条光线；
3. 关于采样以及渲染方面，在每一条射线当中去采若干个点，接着运用神经网络来对每一个点的颜色还有密度加以计算。
4. 损失函数：计算合成图像与真实图像之间的均方误差 (MSE)；
5. 反向传播训练：使用梯度下降算法优化网络参数。

### 2.2.3.NeRF 的优缺点

|  |  |
| --- | --- |
| **优点** | **说明** |
| 高质量合成 | 能够实现逼真的新视角图像生成，细节丰富 |
| 隐式建模 | 不需要显式构建网格，表示灵活 |
| 可扩展性 | 可以结合其他模块如语义、深度、时序等进行拓展 |

|  |  |
| --- | --- |
| **缺点** | **说明** |
| 训练慢 | 原始 NeRF 训练时间长（数小时以上） |
| 渲染效率低 | 推理时需大量采样，难以实时渲染 |
| 无显式几何 | 不方便与传统几何系统对接，如物理引擎或网格操作 |

# 3 图像数据预处理与稀疏点云重建

在以三维高斯泼溅（3DGS）为基础构建的三维重建系统里面，对输入的图像数据展开预处理工作以及初步构建点云，这两个环节可是整个系统流程当中极为重要的前置步骤呀。在这一章当中，会详尽地阐述怎样借助COLMAP来达成相机位姿的估计任务，并且完成稀疏点云的生成工作。除此之外，还会进一步去探寻结合其他一些优化手段，以此来对初始点云的质量实施结构性优化的具体办法呢。

## 3.1 基于COLMAP的相机位姿估计

### 3.1.1 相机模型与投影几何

【COLMAP】借助透视相机模型来给真实世界构建模型。存在一个三维点，它凭借相机内参矩阵以及外参矩阵，进而被投射至图像平面之上：

在这里，所标注的 代表着焦距的具体数值， 则明确指代主点所在的位置情况，而 所对应的是旋转矩阵相关内容， 为平移向量方面的相关量，至于 呢，其表示的是点在相机坐标系之下所具备的深度情况。

### 3.1.2 特征提取与匹配

系统一开始会针对所有输入的图像开展特征点提取方面的工作，在通常情况下，会默认运用SIFT（也就是尺度不变特征变换）算法来提取局部纹理方面的关键点。这些特征点有着尺度方面、旋转方面以及光照方面的不变特性，对于多视角图像的匹配来讲是比较适宜的。

特征匹配分为两个阶段：

先通过计算特征向量间的欧氏距离来做最近邻匹配，然后再结合像Lowe's ratio test这样的比值测试，把存在歧义的匹配给剔除掉。

精匹配方面，借助几何验证的方式，就像利用RANSAC来估计基础矩阵那样，进而从中筛选出那些能够满足极线约束条件的匹配点对。

### 3.1.3 增量式 SfM 重建流程

【COLMAP】运用增量式结构恢复的方式，从一对图像着手，一点点地增添新的视角，同步对已有的三维点坐标以及新相机的外参予以更新。在这一重建的过程当中，涵盖了如下几个颇为关键的步骤：

三角测量（Triangulation）这一概念具体而言是这样的：借助于两个视角下的匹配点来反向推算出三维点所处的位置。

位姿估计（PnP）指的是依照已知的三维点以及二维图像点来求出新相机所处的位姿情况。

全局优化也就是Bundle Adjustment，其主要目的在于把整个系统的重投影误差尽可能降低到最小程度。

这一优化凭借稀疏Levenberg–Marquardt算法得以高效达成。

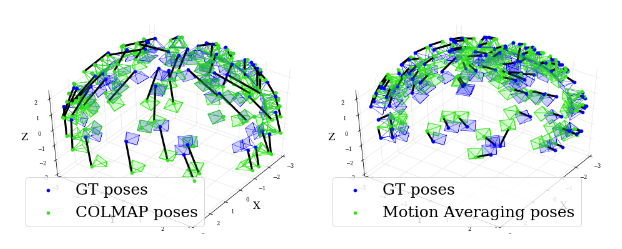


图3.1 COLMAP的特征匹配增强

### 3.1.4 输出格式与可视化

【COLMAP】其最终所给出的输出涵盖了多个方面的内容，具体来讲，其中包含了每张图像的相机内外参数，还有三维稀疏点集也在输出范畴之内。另外，每个点的颜色信息也会输出，这里的颜色是通过对图像进行平均采样而得的，并且与之相关的可视视角索引同样会给出。除此之外，匹配图像对的相关信息也属于最终输出的一部分。

另外，还能够输出可视化的相关结果，像是图像相互间的匹配连线图，还有稀疏点云视图等等，以此来辅助开展后续的模型质量评估方面的工作。

## 3.2 稀疏点云生成

### 3.2.1 三角测量与深度恢复

对于给定图像之间的相对位姿以及像素点匹配的情况，能够运用线性或者非线性三角测量的相关方法来算出匹配点的三维具体位置。要是以将重投影误差尽量减小作为目标的话，那么其中最为常用的当属线性三角测量方法了。

在这之中，它是由两个或者更多相机投影矩阵以及像素坐标共同构造而成的。所得到的点云会被存储于COLMAP数据库里面，通过这样的方式，初步完成了对场景三维几何结构的构建。

### 3.2.2 稀疏点云的特性与局限性

COLMAP 输出的点云具有以下特点：

点的数量级通常为 ，取决于图像数量和视角重叠程度；

关注点大多聚集在图像的重合区域，而在非纹理区域，像是天空、纯色墙面这些地方，则相对较少。

精度较高，但分布不均、存在遮挡与缺失区域；

部分点由于匹配误差或误差传播可能存在偏离或漂浮现象。

这类呈现出稀疏状态的点云，其特性使其比较适合当作结构方面的引导元素。不过呢，它却并不适用于直接开展渲染方面的工作。所以呀，针对这类稀疏点云，还需要对其实施进一步的优化操作，同时也要完成相应的补全工作才行。

## 3.3 基于点云神经网络的系数点云优化

### 3.3.1 点云处理挑战

在传统的三维重建系统里面，点云所涉及到的噪声方面的情况、高维特征提取相关事宜以及局部语义理解这类处理方面的问题，常常是很难依靠几何方法来妥善解决掉的。尤其是处在稀疏点云的情形之下，因为点的密度比较低，而且其分布还不均匀，所以传统的滤波方法或者插值方法想要取得很不错的效果，那是相当困难的。

点云神经网络（Point-based Deep Network）具备着颇为强大的结构感知能力。PointNet算得上是最早被广泛应用的点云处理架构当中的一种，它有这样的能力，就是能够直接从原始的点集去学习相关特征，在这个过程中，并不需要去构建网格或者体素。

### 3.3.2 PointNet架构原理

【PointNet】的核心思路重点在于针对无序点集展开“对称处理”。具体而言，其中的每个点都会被单独传送到一个共享的多层感知器（也就是MLP）当中，以此来提取出局部层面的相关特征。而所有这些被提取出来的特征呢，又会借助一个对称函数（比如说最大池化这种方式）来汇拢在一起，进而形成全局特征。

全局特征能够被应用在分类、回归以及分割等诸多任务当中，并且还可以对每个点的表示予以反馈优化。这种结构具备不少优点，其一是能兼容按照任意顺序所输入的点集；其二是有能力去处理那些呈现出不规则、稀疏状态以及非均匀采样特点的点云；其三则是其学习能力颇为出色，有着广泛的适应性。

### 3.3.3 在稀疏重建中的应用

在3DGS系统当中，能够把PointNet引入到稀疏点云的后处理环节，以此来完成下述这些任务：

（1）点云去噪与剔除

训练一个分类器网络来对异常点加以识别，这些异常点包含配准未能成功的漂浮点、由噪声图像特征所产生的伪点以及因少视角支持而出现的低可信度点。在这个过程中，引入已知场景的先验数据之后，便能够把这些点视作异常点并予以剔除，如此一来，点云的一致性以及可控性也能够得以提高。

（2）点云补全与插值

通过运用Point Completion网络，像Point Completion Network、PCN这类的，去从已有的点云当中来学习其完整的结构，并且在这个过程中把点插入到缺失区域之处，以此方式能够在一定程度上有效提升其稠密度。

（3）语义聚类与分类

【PointNet】能够对所涉及的点云展开相应的空间聚类操作，并且可以为高斯模型添加上具备局部结构特性的标签，像“边缘点”呀、“曲面点”之类的，还有“角点”等等这样的情况。而这方面的相关信息呢，是能够在对高斯参数展开初始化处理的过程中，被当作一种先验来加以运用的。

（4）多尺度结构理解

借助像PointNet++这样的扩展结构，能够在多尺度邻域当中来聚合相关特征，对处于不同尺度之下的结构有着敏感性，可以很好地适配建筑、自然场景之类多种不同类型的数据。

4 基于高斯泼溅的三维重建方法

## 4.1 高斯泼溅表示

在传统的三维重建相关方法里面，像三角网格（Mesh）、体素（Voxel）以及点云（Point Cloud）这些表示方法，往往会存在建模灵活性受到一定限制、表示精度有所欠缺或者数据存在冗余这类问题。高斯泼溅（Gaussian Splatting，简称为3DGS）则给出了一种处于稀疏点云和连续体之间的中间表示形式。它借助各向异性的3D高斯函数，以此来对三维空间里表面的密度以及颜色分布进行近似建模，进而能够以一种连续的、可微的并且较为紧凑的方式，把复杂场景所具有的细节以及结构表达出来。

在3DGS里呀，有这么一种情况，一个三维物体呢，它是由好些个所谓的“高斯点”来构成的。每一个这样的点呀，是被表示成一个呈椭球形的三维高斯核的模样。这些个高斯核呢，可不光是记录下了位置以及颜色方面的信息哦，它们还一并携带着像方向呀、尺度呀、透明度之类的诸多参数呢。而这些参数的存在，其作用就是要在渲染的整个过程当中，借助于图像空间投影的方式来达成融合的这么一个效果呀。

### 4.1.1 高斯分布的数学模型

我们用 表示第i个高斯点，其核心参数包括：

位置：

尺度（协方差矩阵）：，用于控制高斯在各方向上的扩展；

颜色：（RGB 格式）；

不透明度（density/opacity）：；

朝向（由视锥约束的缩放系数）：用于描述点的可见性与方向性。

在空间中，高斯点 对任意采样位置 的贡献可由如下公式表示：

​

此函数于空间当中生成了一个带有色彩的‘模糊球’，其所能产生影响的范围是由协方差矩阵来把控的。对协方差加以调节的话，高斯核便能够呈现出各向异性的形态，如此一来，也就可以更为灵活地去与场景里的各类不同几何细节相契合了。

### 4.1.2 向量化与效率优化

重建场景通常会涵盖数万个高斯点，要是直接去计算它的三维卷积，那所要付出的代价是非常高的。所以呢，3DGS方法一般会采用图像空间的二维渲染策略。具体而言，其做法是这样的：

首先把所有的3D高斯都投影至相机视平面。接着呢，在屏幕空间当中依照像素来聚合它们所产生的影响。最后利用透明度混合以及球谐光照之类的方法去合成颜色。如此一来，这种做法既能够保证三维信息的准确程度，又对二维渲染的高效性给予了兼顾，而且还天然地具备支持GPU并行化计算的特性。

## 4.2 可微渲染过程

高斯泼溅重建方法具备一个极为重要的特性，那便是其整个渲染流程是端到端可微的。这也就表明，渲染所得到的结果能够当作神经网络的输出，进而参与到损失函数的计算当中去，如此一来，便可以借助反向传播的方式对每个高斯点的各类参数，像是位置、颜色以及尺度等等，加以优化。而这样的可微特性，实实在在地为无监督学习以及图像到几何的端到端建模构筑起了相应的基础。

### 4.2.1 渲染原理

在进行渲染操作的时候，一开始会把每一个高斯点都投射到图像平面之上，接着依据它的空间所处位置、自身尺寸大小以及所处的视角等方面的情况，来明确它在屏幕上面所形成的那片覆盖区域，这里所说的覆盖区域其实就是2D椭圆啦。而每一个像素最终呈现出来的颜色呢，是由众多高斯点在该像素所处位置所做出的贡献，经过加权融合的方式而形成的。

具体渲染过程可表述如下：

其中：

：像素坐标；

N：所有参与该像素渲染的高斯点数目；

第i个高斯点在像素方面的透明度（此透明度是依据其投影以及协方差矩阵来加以计算的）

：该点的颜色；

该像素未被处于其前方的高斯点所遮挡的概率，也就是光线穿透率，是按照如下所阐述的方式来进行近似建模处理的：

该公式所呈现出来的是一种从前往后逐步叠加的透明度情况，它清晰地将前方点所产生的遮挡作用给体现了出来，这和在体积渲染过程当中针对光线累积透明度展开的计算颇为类似。

### 4.2.2 可微渲染计算与梯度传播

上述渲染公式里的各个部分，像高斯投影、遮挡概率以及透明度混合，它们每一个其实都是连续可导函数。正因为这样，就能针对损失函数去求偏导，进而得出高斯点参数的梯度。

​

这些梯度能够凭借反向传播机制被送回到神经网络或者优化器那里，进而被用于对模型参数做出更新。高斯点所处位置要是进行了调整的话，那么重建表面就会被推动着变得更加贴近真实几何形态；与此同时，针对颜色以及透明度所做的优化，也使得渲染图像在视觉方面的一致性得到了提升。

### 4.2.3 引入球谐函数进行光照建模

在高斯泼溅方法这块儿，为了让渲染过程里对真实光照变化的模拟能得到更进一步的强化，特别是在对方向性反射进行建模方面的能力得以提升，研究者把球谐函数（Spherical Harmonics, SH）给引入进来了，用它来呈现每个高斯点的颜色响应和观察方向二者之间存在的关系。这种方法在维持着相对较低的计算开销状况下，还可以很有效地去捕捉那些复杂多样的光照变化，进而变成了高斯泼溅渲染当中极为重要的一个组成部分。

**一、球谐函数概述**

球谐函数属于定义在单位球面上的一组正交函数系，其能够被视作是傅里叶级数在球面坐标系方面的一种推广形式。在诸多领域当中，它们有着较为广泛的应用，比如在计算机图形学领域、物理学领域以及量子力学领域等，都常被用于处理其中存在的球面对称问题。

球谐函数的标准形式为：

其中：

：阶数，表示频率层次；

：次数；

分别是球坐标系当中的极角（zenith）以及方位角（azimuth）。

​：关联勒让德多项式；

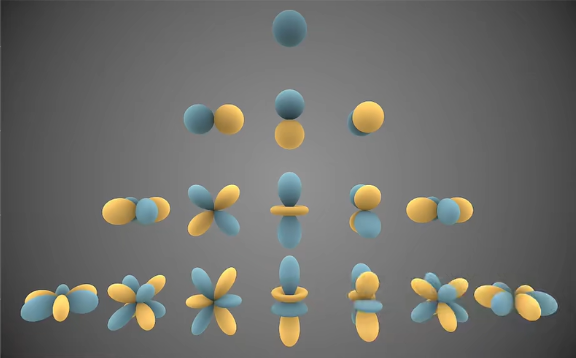
​：归一化系数，确保函数正交性。

球谐函数具有如下重要性质：

所谓完备性，就是说在球面上的任意连续函数，均能够展开成为球谐函数的加权和形式。

方向建模能力：能够表示随方向变化的函数（如反射、光照）；

低频有效性方面，在渲染环节常常会运用低阶球谐去对光照与反射函数加以逼近。如此一来，其在计算方面能够展现出较高的效率，同时在表达能力上也表现得颇为强劲。



**二、球谐函数在3DGS中的应用**

在高斯泼溅的可微渲染过程里面，每一个高斯点所携带的并非仅仅只是一个静态颜色哦，实际上呢，它还拥有一组球谐系数呢，而这组系数是专门用于对方向性反射情况进行描述的呀。在进行渲染操作的时候呀，就着当前像素的视角方向来看呢，我们会把这个视角方向在单位球面上所对应的方向向量当作自变量来使用哦，进而去计算该高斯点在这个特定方向之下所呈现出来的颜色响应情况呀。

具体的操作办法是把球谐函数（Spherical Harmonics, SH）引入进来，使其成为点的颜色表示当中的一部分内容：

其中：

是球谐展开的最大阶数（一般为2或3）；

：每个频率分量上的RGB系数；

：球谐基函数，输入为当前相机与点之间的方向角。

凭借这样的方法，各个点的颜色能够动态地依照光照情况或者视角的不同而发生相应变化，以此达成更为逼真的BRDF（双向反射分布函数）效果。该机制在3DGS当中被用来拟合更为繁杂的光照环境，从而让点云渲染的真实程度得以增强。

**三、SH建模在渲染中的实际应用**

在引入球谐光照建模这一操作之后，3DGS于渲染的过程里面呈现出了更为逼真的视觉特性，尤其适用于如下这些场景：

自然光照环境（IBL）方面，球谐函数比较适宜用于拟合自然图像里的低频全局光照情况。

细节方面是能够加以控制的：借助对球谐展开的阶数予以调节这一方式，便能够在渲染所呈现出的真实感以及计算效率二者之间达成一种平衡状态。

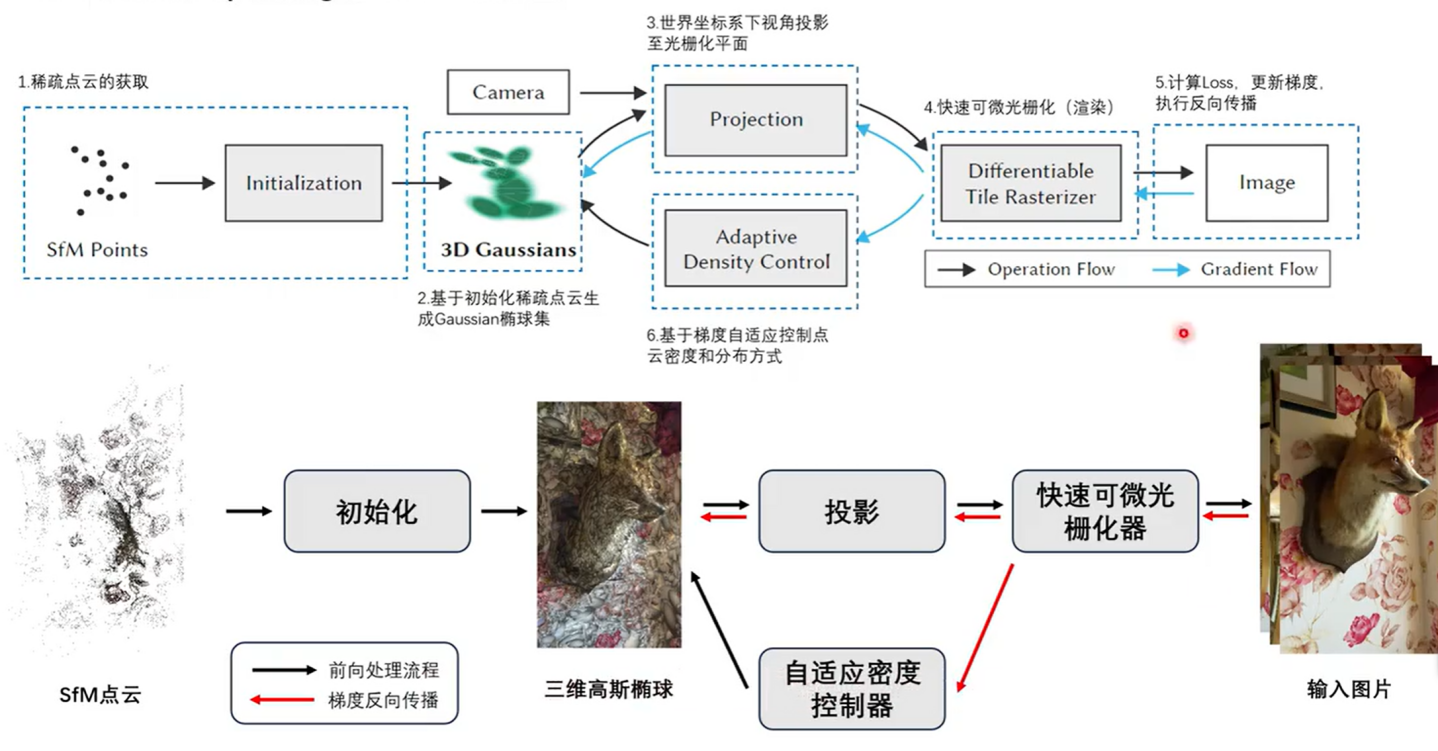
SH系数能够成为高斯点的构成部分，在开展网络训练时可直接对梯度进行回传，进而实现借助图像监督来完成三维重建任务，而且这整个过程是与神经网络相适配的。

在图形学领域当中，往往使用三阶球谐（此处无需对该名词做特别解释，按专业常规理解即可）便能够获取到相对不错的方向感知成效。仅仅需要9个SH基函数，就可以用来表示每个点的光照响应曲线。相较于传统的BRDF或者微面模型而言，其所需成本是极低的。

当引入球谐函数之后，高斯泼溅方法便拥有了能够表示更为丰富几何结构的能力，与此同时，其光照建模能力也得到了增强。如此一来，重建结果在视觉感知方面的真实度以及一致性都有了显著的提升，这无疑为实现高质量且能够实时可视化的三维重建给予了极为有力的支撑。

## 4.3 参数初始化与训练策略

质量颇高的三维重建并非仅仅依靠强大的模型结构，其同时也有赖于良好的参数初始化以及合理的训练策略。在高斯泼溅表示的情形下，每一个高斯点都带有多种参数，像位置、尺度、方向、透明度、颜色以及可能存在的球谐系数（SH）等，而这些参数的初始值不管是对于收敛速度来讲，还是就最终效果而言，均有着颇为重要的影响。



### 4.3.1 初始化策略

在正式开启训练之前呀，往往得借助稀疏点云来对高斯点的位置做初始化处理呢，就好比可以利用来自COLMAP的SfM结果那样的稀疏点云哦。这些被用来初始化的点呀，它们处于三维空间之中，不仅是稀疏的，而且其准确性也是比较可观的，同时还携带着颇为丰富的几何结构方面的信息呢。

每个点的初始参数如下：

位置 ：由 COLMAP 稀疏点云直接提供；

尺度（协方差）这一方面呢，一般情况下最开始是会被设定成各向同性的小球形态哦，也就是说呢，这里面会涉及到一个情况，那就是会存在一个相对来说比较小的常数 ；

颜色 ：可通过多个视角图像投影并平均采样得到；

就方向（用于椭球分布的情况）而言，一开始可将其设定成单位向量，或者通过主成分分析（PCA）来进行估计。

透明度或密度参数 ：初始化为常数或与深度成反比；

球谐系数往往会被设定成均值颜色与一个经验常数相乘的结果，或者直接设为零。这样的初始化策略能够确保模型从一开始便具备基本的几何以及颜色方面的一致性，进而能够切实减少在训练初始阶段所出现的不稳定性情况。

### 4.3.2 训练损失与优化目标

高斯点所设定的训练目标在于，要尽可能将渲染图像和真实图像之间所存在的差异缩小到最低限度。常见的损失函数涵盖了以下这些：

像素级重建损失（MSE）：

感知损失（Perceptual / LPIPS）这一概念是这样的，它会运用预训练网络来提取图像的相关特征，以此去对图像的相似性加以衡量。

深度一致性方面的损失情况：这是一种可选择采用的方式，具体而言，就是在存在深度图监督的情形下才会使用它。

正则项：限制点的尺度、密度、方向，避免训练时的参数发散：

### 4.3.3 多阶段训练策略

为了让训练的稳定性以及最终的质量都能有所提升，3DGS运用的是逐步细化的训练流程。

初始阶段：仅训练位置和颜色；

中期阶段：引入方向和尺度参数优化；

后期阶段：加入球谐系数、密度参数的联合优化；

分辨率提升：逐步引入高分辨率图像和更复杂视角；

在每一步的训练环节当中，均能够运用各不相同的学习率以及调度策略。而较为常用的优化器是Adam，其可与Cosine Annealing学习率下降曲线或者Exponential Decay学习率下降曲线相互搭配使用。

## 4.4 优化与高斯压缩策略

在训练环节里，高斯点的数量会逐步增多，其精度也会持续得到提升。要是不存在优化以及压缩方面的相关机制的话，那么模型就会碰到诸如渲染速度迟缓、显存被大量占用、训练缺乏稳定性等一系列的问题。所以说，对点数展开动态管理，这已然成为3DGS能够取得成功的关键要素之一。

### 4.4.1 点密度增长策略（Density-Guided Splitting）

随着训练不断推进，要在图像细节繁多或者重建误差较为明显的区域增添更多的点。常见的点增长形式包含：

图像误差驱动：在重建误差大的区域插入新点；

梯度驱动法：使用点的梯度大小判断其重要性；

光线采样密度控制：增加点来覆盖疏采样区域；

对于椭球拉伸复制操作而言，是把一个具有一定尺度的点，按照拉伸的方向去复制出多个点，这里所说的复制操作在相关语境下可称作split，通过这样的操作来使得边缘结构能够得以细化。

每一回有新增点出现之后，它的初始化是能够凭借邻域插值这种方式来达成的，或者可以通过权重传播的途径来完成，再者还能运用图像特征嵌入网络去生成呢。

### 4.4.2 点合并与压缩策略（Gaussian Pruning）

为了压缩模型体积，提高渲染效率，需要周期性压缩或合并冗余点：

空间聚类（也就是Spatial Clustering）这一操作呢，是把在空间位置上相互邻近的，并且颜色以及尺度都较为接近的那些点，统统合并成一个单独的点。

​

不可见点删除（Visibility Culling）这一操作具体是这样的：对每个点在训练图像里的可见帧数予以统计，要是某个点的可见帧数低于设定的阈值，那么就把这个点剔除掉。

关于权重贡献的剔除处理，需依据各个点在渲染图像当中所占据的权重情况，针对那些权重相对较低的点，采取相应的惩罚举措或者直接予以删除操作。

关于活跃性的检测是这样操作的：要是球谐系数在很长一段时间里都没有得到更新的话，那么就能够判定该点所做出的贡献是比较小的，基于此便可以将其移除掉。

### 4.4.3 性能优化与加速策略

为了平衡质量与效率，可进一步引入以下技术：

分块渲染也就是Tile-based Splats Rendering这种方式，它会把图像划分成一个个的tile，然后按照实际需求去加载高斯点。

视锥裁剪（Frustum Culling）这一操作，其主要做法便是仅仅加载在当下所处视角之下能够被看见的高斯点。

【KD】-tree或者Octree这类加速结构，能够实现对局部点邻域展开快速查询的操作，进而有助于提升渲染速度。

混合显存与磁盘缓存：将不活跃点转存磁盘，释放 GPU 显存；

### 4.4.4 压缩率与质量的权衡

在部署阶段，模型大小和渲染速度成为关键性能指标：

高压缩率意味着少量高质量的点；

高质量需求则要求保留更多局部细节点；

一般会运用PSNR、LPIPS以及FPS等多个指标来综合评定最终的效果。

# 5 三维重建系统设计与实现

此系统意在达成一个依托Web的三维高斯泼溅（即3D Gaussian Splatting，简称3DGS）可视化及编辑平台的创建。该平台能够以从COLMAP导出的稀疏点云以及位姿信息作为起始点，借助加载已经训练完毕的3DGS模型，进而在浏览器当中达成实时渲染以及交互的效果，同时还能给研究者与用户配备用于可视化调试、参数调整以及数据分析的相关工具。

## 5.1 系统需求与分析

### 5.1.1 功能需求

出于满足三维重建在教学、科研以及应用展示等诸多方面的需求考量，系统需要具备如下的一些功能模块：

相机姿态以及稀疏点云的获取过程如下：利用COLMAP工具来完成稀疏点云的重建工作，与此同时，通过解析的方式得到相机位姿，而获取到的相机位姿能够被用于对场景相机进行初始化操作。

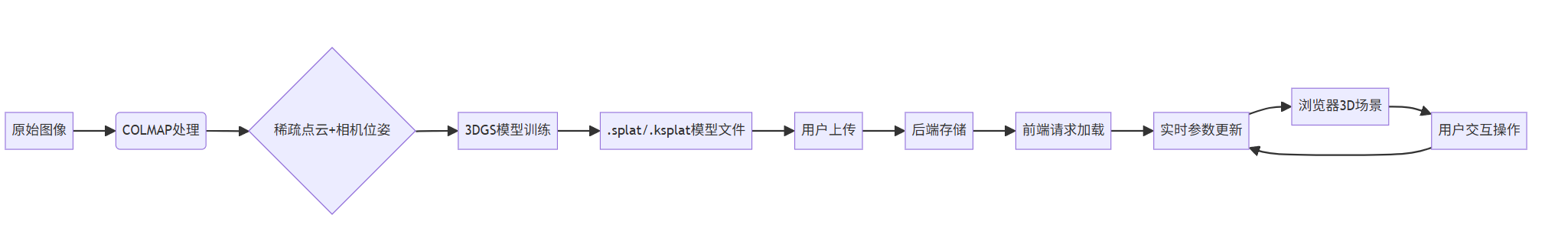
2.3DGS模型的加载与渲染这一环节，是支持用户从本地上传已经训练好的、呈现为.splat或者.ksplat格式的高斯点数据的。在此情况下，相应的系统便具备对这些数据进行解析的能力，进而能够将三维场景予以还原出来。

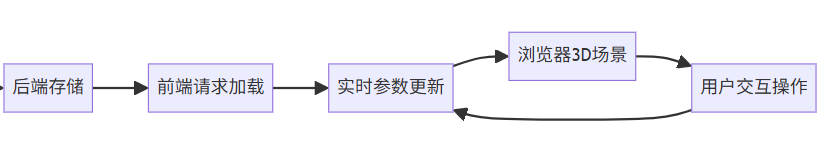
在三维可视化及漫游方面，Web端可提供交互式的三维场景浏览功能。用户能够借助鼠标，以旋转、缩放以及拖动等操作方式，在该场景中实现漫游体验。

4. 关于高斯泼溅参数的调节方面，能够实现对高斯点的诸多属性予以可视化的调整操作，这些属性涵盖了尺寸、颜色、方向以及透明度等等。

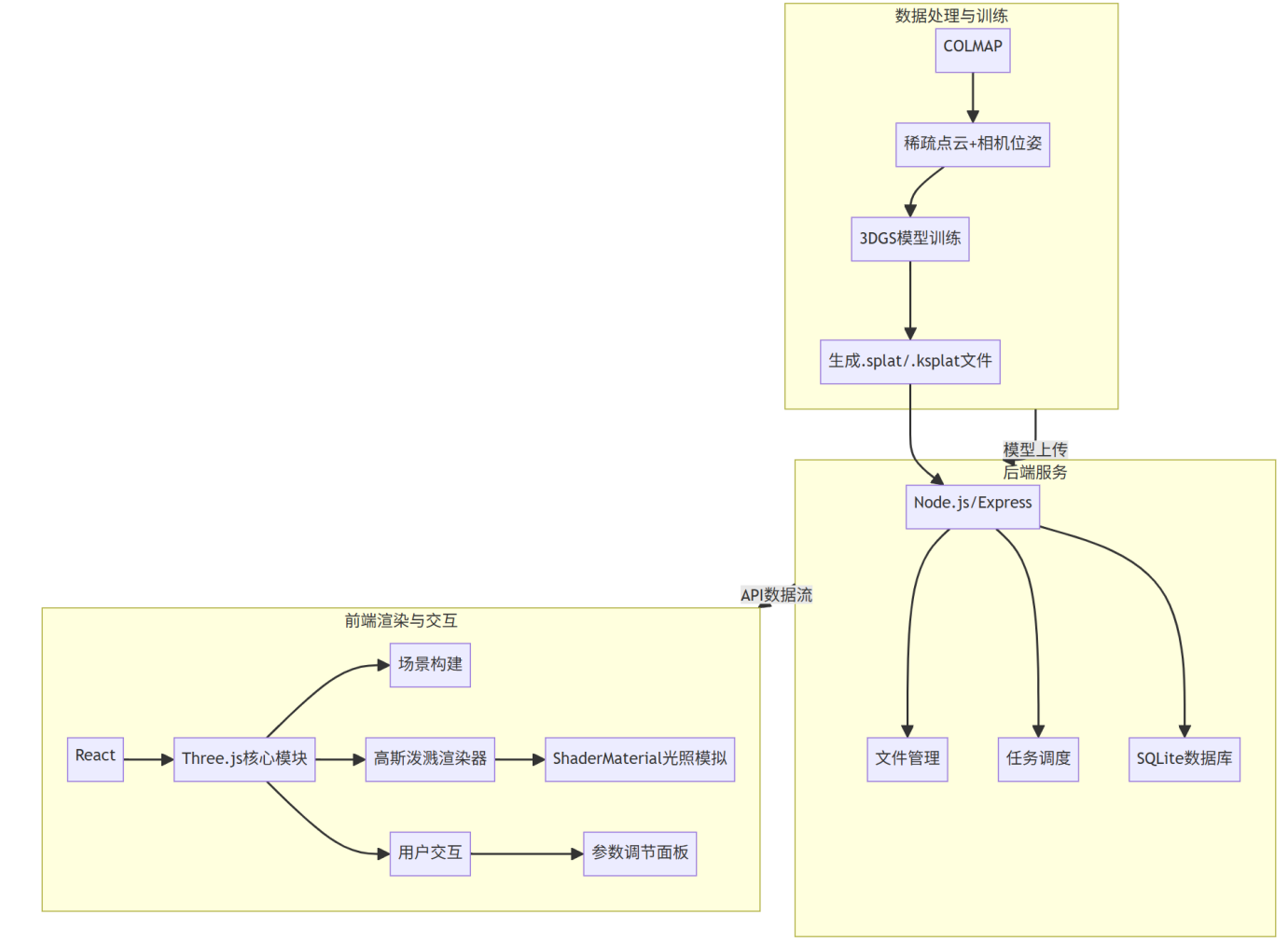
5. 关于光照控制以及SH拟合展示方面：准许用户对环境光的方向、颜色加以调节，进而展示出球谐光照模拟所呈现出的效果。

系统所具备的界面应当呈现出友好的特质，其体验也需足够流畅。具体而言，整体的系统界面要设计得简洁且直观，在接收到用户操作指令时能够迅速给出响应，从而为用户营造出良好的交互体验。





数据流图



### 5.1.2 技术需求

在前端部分，运用React以及Three.js来达成实时在Web端开展3D渲染的操作。

后端是借助Node.js以及Express框架来搭建起来的，其承担着文件管理方面的相关事宜，同时也负责对任务展开调度的工作。

模型的训练是依据原作者所提出的算法模型来开展的，针对点云数据展开相关的处理操作，进而生成格式为3dgs的模型。

## 5.2 后端模块设计

系统后端大体上是以Node.js平台为依托来达成实现的，与此同时，还借助SQLite对用户注册所涉及到的数据加以保存。

## 5.3 前端渲染模块设计

在前端部分，运用React来构建具备组件化特点的页面结构，同时将Three.js与之相结合，以此达成3D可视化的相关功能。其核心模块主要涵盖了：

### 5.3.1 场景构建与数据加载

利用THREE.Scene、THREE.Camera之类的各类来搭建起基础的渲染场景。

支持 GLTF 模型、ksplat格式点云的解析与转换。

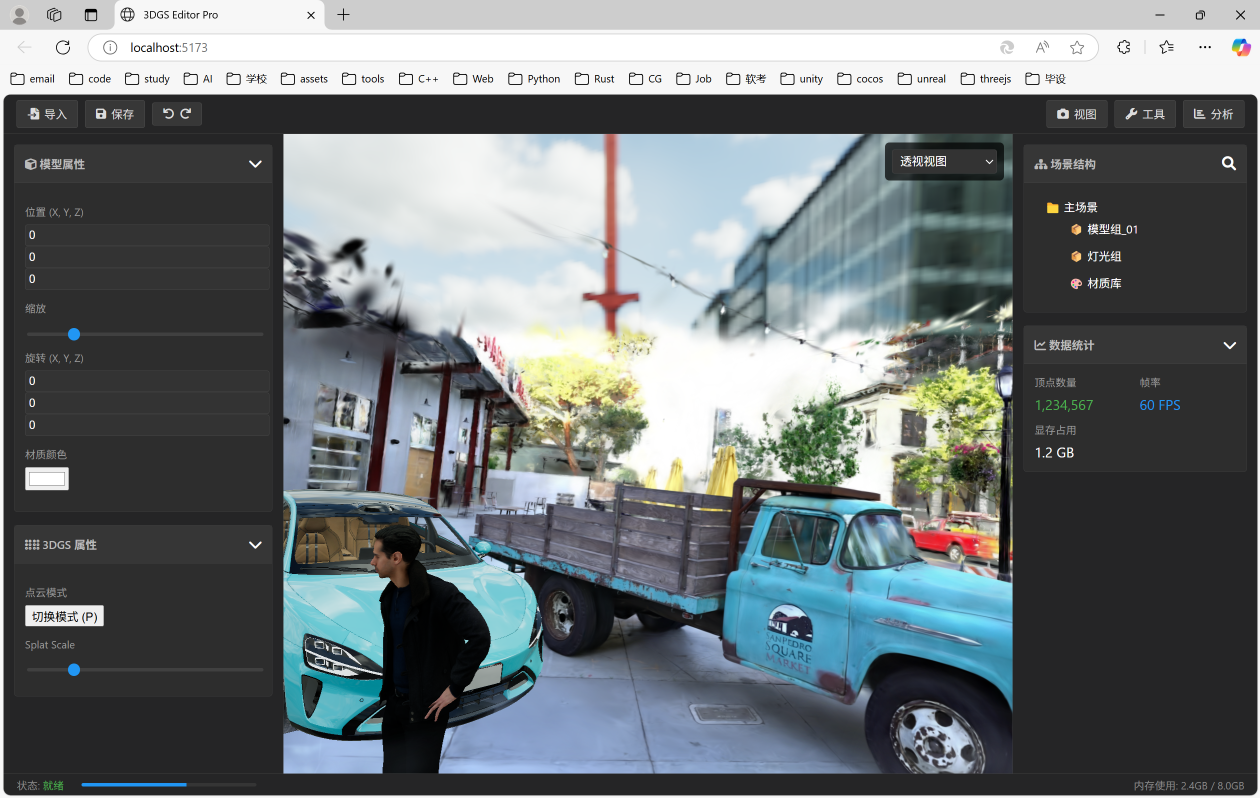
### 5.3.2 高斯泼溅可视化实现

模拟高斯分布呈现出的形状，像是球体或者椭球体之类的样子，并且对其半径、透明度以及颜色都要做出动态的调整。

运用依据ShaderMaterial所形成的GLSL来编写渲染材质，此做法能够对SH光照效果予以简单的模拟。

设置参数调节面板，让用户能够实时对高斯点属性加以调节，这里所说的高斯点属性包含但不限于尺度、方向以及颜色等方面。

能够编辑加载和处理的GLTF模型和动画。



### 5.3.3 用户交互与界面设计

提供鼠标操作支持，包括旋转、缩放、拖拽等；

引入调试面板提升交互体验。

# 6 总结与展望

## 6.1 全文总结

此篇论文着重针对以Gaussian Splatting为基础构建起来的三维重建系统展开相关的研究探讨，尤其把关注点放在了从图像作为输入一直到能够输出高质量点云渲染这一整套的技术流程之上。整个论文所涉及到的研究工作大体上能够划分成如下的几个不同方面。

起初，在图像预处理以及稀疏点云重建这块领域，本文借助COLMAP工具来针对多视图图像展开特征匹配以及束束优化方面的相关操作，进而获取到稀疏点云以及相机位姿方面的信息，以此为后续开展的三维建模工作给予可靠的初始结构方面的有力支撑。而为了能够让点云质量得到更进一步的提升，本文把PointNet网络引入进来，用其对稀疏点云实施感知优化方面的相关处理，在确保整体结构能够维持一致性的情形下，使得局部几何表达的精准性得以有效提升。

再者，就以Gaussian Splatting为基础的三维表示方法来讲，本文较为系统地阐述了三维高斯泼溅（3DGS）的表示原理、渲染机制以及参数建模的方式。和传统的体素、网格表示相比而言，3DGS凭借着大量带有空间分布以及方向特性的高斯点，达成了高效同时又可微的三维建模与渲染这一流程，从而给无网格重建技术开拓出了新的路径方向。

在渲染相关内容的阐述方面，本文着重对可微渲染的数学原理展开了分析。其中涵盖了基于密度以及透明度的融合机制，还有投影方面的机制，另外，遮挡建模以及梯度传播方法等内容也在分析范围之内。除此之外，还引入了球谐函数，用其来针对点的方向性反射特性进行建模操作。通过这样的方式，能够进一步去提升渲染光照所具有的真实感以及一致性，从而为Gaussian Splatting模型在那种较为复杂的光照环境之下的性能表现奠定相应的理论基础。

随后，本文对参数初始化以及训练策略展开了细致探讨，运用多阶段且有着多目标的优化方式，促使高斯点从一开始那种稀疏又较为粗略的初始状态，一步步地被优化成了密集且细致的三维表达形式。在此基础上，为了能够进一步对模型性能以及存储开销加以优化，本文还给出了一套关于高斯点压缩以及剪枝的机制，在推理阶段可以依据实际情况动态地去调整高斯点的数量，以此来确保模型具备高效性以及可用性。

最终，本文所取得的研究成果能够直接在无网格三维建模、全景重建以及AR/VR等前端应用场景当中加以运用，与此同时，还为点云渲染、可微图形等相关的研究领域给出了具有可推广性质的思路以及方法。

综合来看，这篇文章从稀疏点云怎样生成以及如何去优化这两方面着手，较为系统地搭建起了以3DGS作为核心的那种高质量的三维重建流程。在此过程中，对它在精度方面、速度层面以及可扩展性这些不同方面所具有的优势都进行了相应验证，进而将基于Gaussian Splatting的三维建模技术在未来发展进程里所蕴含的巨大潜力很好地展现了出来。

## 6.2 展望

虽说本文于三维高斯泼溅建模这一领域已然收获了特定的阶段性成果，不过就当下的系统来讲，在可扩展性以及优化的方向之上，依旧留存着为数不少的研究空间，后续的工作能够从如下的几个方面去展开进一步的探寻：

现今，稠密点云以及几何细节建模方面，相关系统大多是依靠稀疏点云还有单幅图像所包含的信息来开展重建工作的。虽说借助可微优化以及高斯点增强的手段，在相应程度上让重建的质量有所提升了，不过呢，在那些存在纹理重复情况或者纹理缺失的区域，要想把细节完整地还原出来，依旧是比较困难的事情。往后的话，可以试着去把多源的信息融合到一起（就像深度图、法线图、激光点云等等这些），如此一来，便能够得到在几何形态以及结构表达上质量更高的成果。

（2）高斯泼溅虽说有着不错的实时渲染能力，不过呢，在面对高点数场景的时候，还是有可能碰到带宽以及显存方面的瓶颈难题。往后呀，可以进一步去钻研依据硬件自身特性而开展的图形加速相关内容，就好比WebGPU、CUDA光线追踪等等，同时呢，也可以对压缩技术加以探究，像是点的层次LOD、稀疏索引结构之类的，通过这样的方式来促使系统的响应效率得以提升，并且让其对移动端以及Web应用的适配能力变得更强。

当前系统在设计上大多侧重于静态场景，而关于高斯泼溅在动态对象以及动态场景方面的拓展应用，这无疑是未来需要去探索的一个颇具潜力的方向。可以试着引入从时间维度出发对高斯点轨迹展开建模的方法，再结合基于目标跟踪来进行动态姿态预测的相关机制，此外，还可运用帧间可微约束策略，以此来达成对动态环境连续性的重建以及渲染的目的。

(4)就稀疏点云数据而言，尝试结合点云神经网络来对其加以优化。当下，以稀疏点云为基础展开的高斯泼溅建模，在针对纹理缺失区域进行几何恢复的时候，还是存在着一定的局限性的。在这种情形之下，可以把PointNet++这类点云神经网络引入进来，以此实现感知方面的优化。PointNet++ 它是凭借着层次采样以及局部特征聚合这样的机制，去提取出多尺度的几何特征的，并且还会运用编码 - 解码结构以及跳跃连接的方式，来把不同层级的信息加以融合。要是将高斯泼溅建模和点云神经网络两者结合到一起的话，那么就很有希望达成物理可解释性与深度学习能力的有机融合这样的一个效果。通过对高斯点的尺度以及方向做出动态的调整，能够让无纹理区域的细节重建工作得到增强。与此同时，还可以利用图神经网络去开展动态轨迹建模的相关工作，进而实现跨帧几何的一致性。

3DGS在交互式编辑与应用拓展方面有着独特优势，它不光能用于重建以及可视化场景，还拥有很不错的可编辑性以及物理可解释性。在未来，可以把它集成到三维编辑工具里面，如此一来，就能在高斯点层级实现诸如交互式修改、分割以及重定向等一系列操作，进而促使它在游戏制作、虚拟现实、数字文化遗产保护等诸多领域达成落地应用以及产业转化的目标。

# 致 谢

随着时光不断流逝，毕业论文的撰写也快要结束了，在这个时候，我满怀着诚挚的感激之情，向那些在我开展学习以及从事研究的整个过程当中，给予了我悉心指导以及诸多帮助的老师、同学还有亲友们，表达我深深的谢意。

首先得诚挚感谢我的导师杨庆老师呀，在本论文从选题开始，到后续展开研究，再到具体撰写等各个不同的阶段当中，都给予了我极为细致且周到的指导，还有那无私的帮助呢。导师那种严谨的治学态度哟，深厚的专业知识素养呀，以及踏实认真的工作行事作风，这些对我产生的影响可深远啦，实实在在让我在毕业设计这项工作上收获了诸多的益处呀。

再者，要感谢身处的课题组里的各位同学以及前辈们，在项目具体开展以及学术相互交流期间，他们给予了我颇多的启发以及鼓舞。尤其是在搭建实验、实现代码以及开展论文讨论这些方面，大家所给出的建议还有彼此间的协作，使得我避免了不少弯路，同时也让我深切感受到了团队合作所蕴含的力量。

特别要感谢信息科学与工程学院大数据与人工智能系的梁宗保教授、刘新龙老师、王笛老师以及杨翔立老师等，他们在学业的诸多方面给予了悉心的指导与切实的帮助；同时也十分感激辅导员张秋玉老师、金其奇老师，他们在本科阶段的学习生活里给予了细致入微的照顾与贴心的关怀。

我要衷心感谢我的同学李财华、甘孝平以及前专业的袁倩、吴智博。回想那四年的本科时光，是与你们一同度过的，其间留存了诸多美好的回忆。在那段日子里，大家彼此扶持、相互鼓励，同时也在不断地相互学习，就这样携手走过了这段无比美好的时光。你们无疑是我在求学探索之路上极为重要的力量源泉，在此，衷心祝愿你们未来事事皆顺遂。

在此期间，还得感谢我实习时所接触到的同校学长朱悦凯以及导师谢传金。在我初次踏入工作岗位之时，他们给予了诸多的帮助，生活方面也好，技术层面也罢，乃至个人发展等方面，都给我带来了极为深远的影响。

我要特别感谢我的家人，他们一直以来都给予我支持和鼓励，给我提供了坚实的后盾，这让我在碰到困难以及面临挑战的时候，依然能够充满勇气，保持动力。

最终，要衷心感谢母校，它给予了我十分优良的学习以及研究环境，让我得以有契机在图形学、人工智能这两个交叉的领域去展开较为深入的探索，从而为日后迈向更高的学术层级筑牢了颇为坚实的基础。

谨以此文，献给所有关心和支持过我的人。

参 考 文 献

1. 高建、陈林卓、沈秋等人针对基于三维高斯溅射技术的可微分渲染展开了相关研究，并梳理了其研究进展。相关成果发表于《激光与光电子学进展》，文章页码为1至23页，发布时间是2025年4月13日。
2. 朱东林、陈淼、毛宇岩等人所著的《三维高斯泼溅技术在场景重建中的现状与挑战》一文，发表于《集成技术》，其页码范围是1至17页，发表时间为2025年4月13日。
3. 韩开以及徐娟所撰写的《3D场景渲染技术—神经辐射场的研究综述》发表于《计算机应用研究》2024年第41卷第8期。
4. 侯耀斐、黄海松、范青松等人提出了一种基于改进多层感知机的神经辐射场三维重建方法，相关成果发表于《激光与光电子学进展》2024年第61卷第4期，其文章编号为0415004。另外，李吉洋、程乐超、何靖璇等人也针对神经辐射场展开了研究，他们对其研究现状进行了梳理并对未来发展做出了展望，相关内容发表在《计算机辅助设计与图形学学报》2024年的刊物上。
5. 王锋、银莹、王佳炎等人推出了一种基于高斯泼溅的轻量级重建场景分割方法，相关内容发表于《计算机学报》，其页码范围是从1到12页，发表时间为2025年4月22日。
6. 何天琪、宋佳洁、程景春等人开展了关于利用三维高斯泼溅技术来进行航拍场景建模方面的研究，相关成果发表于《火箭军工程大学学报》2025年第39卷第2期，其页码范围是从第1页至第12页，另外还涉及到第21页。
7. 宋明清、郭尧以及李晓峰等人的研究成果为《基于光照探针的三维高斯辐射场压缩算法》，其刊载于《计算机应用研究》，具体页码为1至8页，发布时间是2025年4月22日。
8. 张月婷、李文劼、郭嘉逸等人针对高斯基的SAR图像重建技术展开了相关的研究，其成果发表于《激光与红外》2024年第54卷第12期，所在页码为1936至1940页。
9. Kerbl B, Kopanas G, Leimkuhler T, and others have put forward the method of 3d gaussian splatting for the purpose of real-time radiance field rendering. Their work was published in ACM Transactions on Graphics in 2023, with the volume being 42 and the issue number being 4, covering pages from 1 to 14.
10. Xin YY, Zuo XX, Lu DY, along with others, came up with SimpleMapping. It's about real-time visual-inertial dense mapping that incorporates deep multi-view stereo. This was presented in the 2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), with the relevant content spanning from page 273 to 282 in 2023.
11. Guedon A along with Lepetit V. developed SuGaR, which pertains to Surface-Aligned Gaussian Splatting. This is for the purpose of accomplishing efficient 3D Mesh Reconstruction as well as achieving high-quality Mesh Rendering [C] // 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024: 5354-5363.
12. Liang Z, Zhang Q, Feng Y, and others. 'GS-IR: 3D Gaussian Splatting for Inverse Rendering'. It was presented at the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) [C] // 2024: 21644 - 21653.
13. Fei B, Xu J, Zhang R, along with others, have come up with '3D Gaussian Splatting as New Era: A Survey'. It's been published in IEEE Transactions on Visualization and Computer Graphics, specifically in the Early Access section, and the year of its release is 2024.
14. Zhu Z, Peng SY, Larsson V, along with others, presented NICE-SLAM: Neural Implicit Scalable Encoding for SLAM [C] // 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022: 12776-12786.
15. Zhu DL, Wang ZL, Lu TF and others came up with PMF-SLAM. This is a kind of Semantic SLAM which is designed for Autonomous Wheel Loader. It's based on Pose-Guided and Multiscale Feature Interaction. The related research was published in IEEE Sensors Journal in 2024. Specifically, it can be found in volume 24, issue 7, with the page range from 11625 to 11638.
16. Fridovich-Keil S, Meanti G, Warburg FR, Recht B, Kanazawa AK. K-planes: Explicit radiance fields regarding space, time as well as appearance. In the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, it can be found on pages 12479 to 12488, and this was in the year 2023.
17. Jiang Y., Yu C., Xie T., Li X., Feng Y., Wang H., Li M., Lau H., Gao F., Yang Y., and others have developed VRGS, which is an interactive Gaussian splatting system in virtual reality that is aware of physical dynamics. This work was presented in the arXiv preprint with the identifier arXiv:2401.16663 in the year 2024.
18. MILDENHALL B, SRINIVASAN P P, ORTIZ-CAYON R, and others came up with 'Local light field fusion: Practical view synthesis with prescriptive sampling guidelines'. It was published in ACM Transactions on Graphics in 2019, and it can be found in volume 38, issue 4, spanning from page 1 to page 14.
19. TANG J, REN J, ZHOU H, and others. Dreamgaussian: Generative gaussian splatting for the purpose of efficient 3d content creation [EB/OL]. (2023-09-28)[2024-05-27].
20. Rusinkiewicz, S. and Levoy, M. developed QSplat, which is a multiresolution point rendering system designed for handling large meshes. This work was presented in the Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, with the relevant content spanning from page 343 to page 352 in the year 2000.
21. Botsch, M. and Kobbelt, L. carried out high-quality point-based rendering on contemporary GPUs. This was presented in the Proceedings of the 11th Pacific Conference on Computer Graphics and Applications, with the relevant content spanning from page 335 to 343 in the year 2003.
22. Ge, W., Hu, T., Zhao, H., Liu, S., and Chen, Y. C. developed Ref-NeuS, which is a method for multi-view reconstruction with reflection. It focuses on reducing ambiguity in neural implicit surface learning. Their work was presented in the Proceedings of the IEEE/CVF International Conference on Computer Vision, spanning from page 4228 to 4237 in the year 2023.
23. Bonet, J., and Wood, R. D. came up with the work titled 'Nonlinear Continuum Mechanics for Finite Element Analysis'. It was published by Cambridge University Press in Cambridge, UK back in 2008.
24. Brooks T., Holynski A., and Efros A. A. focused on the study of 'InstructPix2Pix: Learning to follow image editing instructions'. Their work was presented in the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition in 2023, with the relevant content covering pages 18392 to 18402.
25. Seiskari O., Ylilammi J., Kaatrasalo V., Rantalankila P., Turkulainen M., Kannala J., Rahtu E., and Solin A. have explored Gaussian splatting in a scenario where the camera is in motion. They've focused on compensating for blur and the effects of rolling shutter, which occur naturally during camera movement. Their work is presented in the arXiv preprint arXiv:2403.13327, which was released in 2024.
26. Xiong B, Li Z, Li Z put forward GauU-scene. It is a scene reconstruction benchmark. It makes use of Gaussian splatting on the large scale 3D reconstruction dataset. This was presented in the arXiv preprint arXiv:2401.14032 in 2024.
27. HO J, Jain A, and Abbeel P. came up with the Denoising Diffusion Probabilistic Models [J]. It was presented in the Advances in Neural Information Processing Systems in 2020, with the relevant content spanning from page 336840 to page 336851.
28. Gao L., Yang J., Wu T., Yuan Y., Fu H., Lai Y., and Zhang H. developed the SDM-NET, which is a deep generative network designed for structured deformable mesh. Their work was published in ACM Transactions on Graphics in 2019, specifically in Volume 38, Number 6, with Article No. 243.
29. FANG J, WANG J, ZHANG X, and others. Gaussianeditor: Skillfully editing 3d gaussians by means of text instructions [EB/OL]. (2023-11-27)[2024-05-27].
30. ZHENXING M I and XU D came up with Switch-nerf. It's about learning scene decomposition by using a mixture of experts for those large-scale neural radiance fields. This was presented at The Eleventh International Conference on Learning Representations back in 2022.
31. SONG L, CHEN A, LI Z, and others have presented Nerfplayer. It's a streamable dynamic scene representation that comes with decomposed neural radiance fields, as detailed in the IEEE Transactions on Visualization and Computer Graphics in 2022, with the volume number being 29 and the issue number being 5, covering the page range from 2732 to 2742.
32. GAO C, SARAF A, KOPF J, and others. The synthesis of dynamic views from dynamic monocular video [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. In the year 2021, it was presented with the relevant content spanning from page 5712 to page 5721.
33. BARRON J T, MILDENHALL B, VERBIN D, and others. Mip-nerf 360: Neural radiance fields that are unbounded and anti-aliased [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 5470 - 5479. Here, the focus is on presenting the work related to Mip-nerf 360 which involves dealing with the concept of unbounded and anti-aliased neural radiance fields. The details of this work are laid out in the Proceedings of the mentioned conference in the year 2022 with the specific page range being from 5470 to 5479.