

基于经典方法与新兴融合技术的 三维建模比较分析

方子威 文 静 黄于蒙 高 昕

江苏师范大学 物理与光电工程学院，徐州

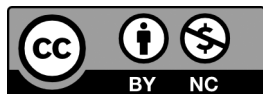
摘 要 | 从二维图像恢复三维结构的问题，长期以来备受计算机视觉与图形学领域关注。本文选取运动恢复结构（SfM）、多视图立体（MVS）、神经辐射场（NeRF）与3D高斯泼溅（3DGS）四条路线，以Mip-NeRF 360和Tanks & Temples两个基准上的公开数据集为基础，定量评估其重建精度、渲染效率、资源消耗与场景适用性。数据表明，3DGS以134~197 FPS的实时光栅化速率领先，约为Mip-NeRF 360的1000~3000倍，训练压缩至35~45分钟，PSNR（27.21 dB）与Mip-NeRF 360（27.69 dB）基本持平。经典工作流程在几何精度和可编辑性上依然领先。在稀疏视角下的新视角合成中，神经表征优势明显。

关键词 | 三维重建；神经辐射场；3D高斯泼溅；多视图立体；新视角合成

Copyright © 2026 by author (s) and SciScan Publishing Limited

This article is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

<https://creativecommons.org/licenses/by-nc/4.0/>



1 引言

1.1 研究背景与意义

三维重建的重要任务，是从图像、视频或深度数据中推算场景几何形状和表面材质，服务于自动驾驶、文化遗产存档、VR内容生产和工业逆向测量等场景。在过去二十年里，该领域先后经历了两次转向，即从纯几何驱动转为数据驱动，从基于点云、网格的显式表征，转为基于神经场的隐式表征。

早期研究以多视图几何理论^[1]为数学基础，实践上依次串联立体匹配^[2]、运动恢复结构（SfM）^[3]和多视

图立体（MVS）^[4]的环节，最终从图像集合中产出稠密点云或三角网格。这些工作流程将严格几何约束和SIFT等手工特征匹配作为算法主要内容，在光照稳定、纹理丰富的受控场景中精度很高，但一旦纹理缺乏，或光照剧烈波动，匹配就容易失败。

2020年，Mildenhall等人正式提出神经辐射场（NeRF）^[5]，采用多层感知器（MLP）将场景编码为连续辐射函数，经可微体积渲染合成新视角。NeRF的视觉真实感，远超此前方法，大量跟进技术随后应运而生。然而，训练耗时数十小时量级、渲染不足0.15 FPS的瓶颈，让它难以走出实验室。2023年，Kerbl等人倡导的3D高斯

基金项目：新工科2.0专项教育教学研究课题项目（XGKY202411）；江苏师范大学“十四五”第三批本科教改项目立项（课程建设编号：JYKC202423）；江苏师范大学博士学位教师科研启动项目（22XFRS034）。

通讯作者：高昕，江苏师范大学物理与光电工程学院讲师，研究方向：目标检测与分类、智能图像处理、统计信号处理、人工智能理论、数学建模等。

文章引用：方子威，文静，黄于蒙，等. 基于经典方法与新兴融合技术的三维建模比较分析 [J]. 社会科学进展, 2026, 8 (7): 638-647.

<https://doi.org/10.35534/pss.0807110>

泼溅 (3DGS) [6] 做了关键改动,即用显式高斯基元替代隐式MLP,使得渲染直接升到实时级别,且视觉质量无明显折损。

基于上述技术背景,本文系统化比较经典几何方

法与新兴神经表征方法的技术特点、性能边界和适用场景,对理解该领域的技术发展规律和实际工程选型,具有直接参考价值。

三维重建方法的分类与发展历程示意图,如图1所示。

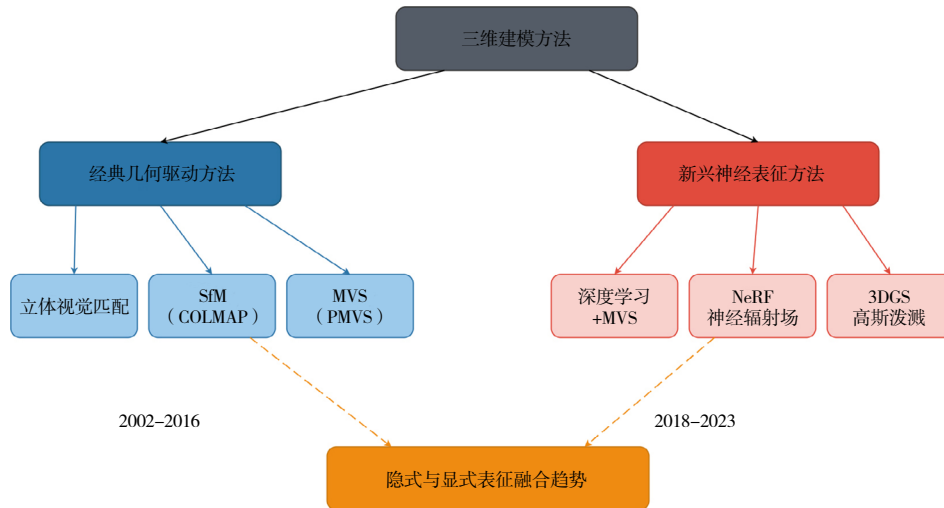


图1 三维重建方法分类与发展历程

Figure 1 Classification and development history of 3D reconstruction methods

1.2 国内外研究现状

2004年, Hartley与Zisserman合著的《多视图几何》[1], 为经典工作流程奠定了基础。编著者结合一系列数学方法,系统地整理了对极几何约束、基础/本质矩阵估计、三角测量,以及束调整(BA)的优化理论。在MVS这一环, Furukawa与Ponce的PMVS算法[4],以面片为基本单元,按“匹配→扩展→过滤”的流程,推进稠密重建,一度成为标准化模式。2016年, Schönberger等人研讨的COLMAP系统[3, 7],则将增量式SfM与逐像素深度估计,整合为统一内容。因其稳定性与通用性突出, COLMAP代表着学术界与工业界使用最广泛的开源 SfM-MVS 重建工具链之一。

深度学习技术的进步,为三维建模带来了新的工具。Yao等人的MVSNet[8],将MVS代价体构建变为可微操作,用3D卷积正则化后回归深度图,在DTU[9]上精度超过传统方法。Chen等人的Point-MVSNet[10],进一步将深度图转点云,通过特征增强与逐级精化提升细节。

NeRF[5]开启这一新方向后,训练慢和渲染慢,很快成为社区集中攻克的两个问题。Müller等人的Instant-NGP[11],率先用多分辨率哈希编码,将训练压至秒级。Mip-NeRF 360则在渲染质量上继续推进。Wang等人在综述[12]中对NeRF系方法做了系统整理。2023年, 3DGS[6]把可微高斯推到前台,依靠自适应密度控制动态调整高斯数量,配合瓦片光栅化器实现高效渲染。3DGS将训练时长压缩至数十分钟,同时可实现实时级渲染帧率,自

从该项研究成果问世后,在三维视觉领域获得大量跟进研究。

1.3 本文研究内容与组织结构

本文的对比研究范围集中在两条技术路线上,分别是基于几何理论的经典方法,以及近几年快速迭代的神经表征方法。作者选取三组方法对比分析。经典路线以COLMAP[3, 7]和PMVS[4]为代表,神经路线以NeRF[5]、Instant-NGP[11]和3DGS[6]为代表,学科交叉以MVSNet[8]和Point-MVSNet[10]为代表,体现深度学习与传统几何的融合。定性、定量评估相关方法时,均基于公开基准上的定量指标。

该论文共六章,后续章节安排如下:第2章介绍经典路线的理论基础与几种代表性算法。第3章考察三维建模的新兴技术,依次介绍深度学习增强的MVS、NeRF系列,以及3DGS。第4章从精度、效率、场景适用性等内容切入,对经典方法与新兴融合技术做横向比较。第5章汇总实验数据,并给出定量结论。第6章归纳全文研究工作,展望该研究后续可行研究方向。

2 三维建模经典方法

2.1 立体视觉与多视图几何基础

立体视觉的理论基础是对极几何。给定两幅存在重叠视野的图像,对应像素点 x 和 x' 之间满足对极约束关系 $x'^T F x = 0$,其中 F 为 3×3 的基础矩阵,编码实现两幅图像

间的相对位姿和内参信息^[1]。当相机内参已知时，基础矩阵可退化为本质矩阵 $E = K^T FK$ ，其中 K 为相机内参矩阵。通过对本质矩阵进行奇异值分解，即可恢复相机间的相对旋转和平移。

三角测量是从多幅图像的对应关系中恢复三维点坐标的基本操作。对于一对匹配的图像点，将其反投影射线在三维空间中求交点，即可得到空间点位置。在实际情况下，由于噪声的存在，两条射线往往无法精确相交，此时可采用直接线性变换（Direct Linear Transform, DLT）方法，求得最小二乘解^[1]。

在重建工作流程中，束调整（Bundle Adjustment, BA）承担着最后一道全局优化角色。该方法是公认的多视图重建金标准优化方法^[1]，但其计算复杂度随相机数量和三维点数量增长而急剧上升。它把相机参数和三维点坐标放在一起，视为待优化变量，构造非线性最小二乘问题，目标函数是所有可见图像点上重投影误差的平方和。求解时通常采用Levenberg-Marquardt迭代算法。BA的数学形式可表述为

$$\min \sum_{i,j} \rho \left(\|x_{ij} - \pi(P_j, X_i)\|^2 \right) \quad (2-1)$$

其中， P_j 为第 j 个相机的投影矩阵， X_i 为第 i 个三维点， $\pi(\cdot)$ 为投影函数， $\rho(\cdot)$ 为鲁棒核函数。

在双目或多目视觉中恢复深度信息，立体匹配是关键步骤。Scharstein与Szeliski将立体匹配归纳为四个阶段^[2]：匹配代价计算、代价聚合、视差计算与优化、视差精

化。这一分类框架，至今仍是评估和设计立体匹配算法的重要参照。

2.2 运动恢复结构（SfM）：以 COLMAP 为代表

运动恢复结构用于从一组无序的二维图像中同时恢复相机位姿和稀疏三维点云。高翔等人^[13]系统地综述了大规模场景运动恢复结构的方法体系，涵盖了增量式、全局式与混合式三类主流框架的发展历程与性能特征。

COLMAP^[3,7]是目前应用最广泛的增量式SfM系统之一，其工作流程如图2所示。

COLMAP的增量式SfM，由紧密衔接的几个步骤组成。第一步是对每幅图像提取SIFT 128维描述子，并做两两匹配。随后通过基础矩阵/本质矩阵的几何验证，筛选不可靠的匹配对，构建场景图。在初始化阶段，从图中选择一对重叠充分、基线适中的图像，利用双视图几何关系，恢复初始三维结构和相机位姿，作为增量扩展的种子。

增量扩展过程中，COLMAP用一套策略来抑制误差漂移，即优先挑选可见最多已重建点的图像，通过透视-n点（Perspective-n-Point, PnP）算法来估计其位姿。新图像注册后，立刻对可见匹配点执行三角测量产生新三维点，然后依次触发局部BA和全局BA。

总体而言，通过“注册—三角测量—局部BA—全局BA”的紧凑化迭代流程，可以使得COLMAP在精度和稳定性上均能明显超越早期的全局式SfM方法。

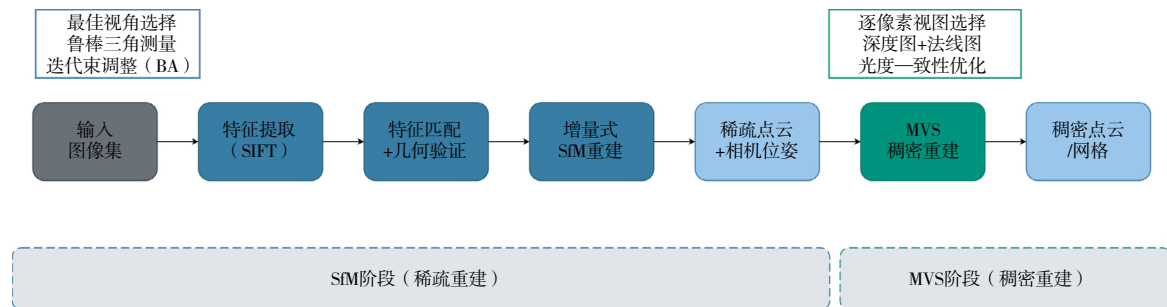


图2 COLMAP SfM-MVS重建流程

Figure 2 Reconstruction flowchart of COLMAP SfM-MVS

2.3 多视图立体（MVS）：PMVS 与 COLMAP-MVS

在SfM获取稀疏点云和相机位姿基础上，多视图立体进一步恢复场景的稠密几何结构，经典的方法，以PMVS和COLMAP-MVS为典型代表。

PMVS（Patch-based Multi-View Stereo）^[4]将场景表示为带法线方向的3D矩形面片，重建流程分为匹配、扩展、过滤三步。匹配阶段在参考图像网格上选取种子点，并做多视图法线优化。扩展阶段在新面片的邻域逐步填充间隙。过滤阶段利用光度一致性和可见性约束，

剔除错误面片。PMVS能较好处理弱纹理区域，但扩展过程对参数较敏感，而且大规模场景中的计算开销较大。

与PMVS不同，COLMAP-MVS^[7]选择了逐像素估计深度的实现方案。该系统对每幅参考图像独立估计深度图和法线图，以NCC作为匹配代价，联合优化光度一致性与几何先验。实现深度生成后，通过多视图几何一致性过滤离群值，再将各视图深度图融合为稠密点云或网格。逐像素的策略，从机制上提升了对遮挡和纹理变化的鲁棒性，完成融合任务后，可以直接做Marching Cubes网格重建。

2.4 经典方法的特点与局限

上述两种经典方法的长处,在于几何理论的严格性和可解释性。对极几何和束调整提供了数学保证,重投影误差可定量验证重建质量。输出的显式点云或网格,可直接测量、编辑和加工。值得注意的是,经典方法也

并非没有短板。遇上弱纹理、周期性纹理或强光照变化时,SIFT等手工特征匹配易失败。立体匹配中的遮挡处理依赖启发式规则,难以覆盖复杂场景的全部遮挡。束调整(BA)的计算量随重建规模急剧膨胀,大场景可能消耗数小时乃至数天的CPU时间。对比经典三维建模方法的主要特点,归纳总结如表1所示。

表1 经典三维建模方法对比

Table 1 Comparison of classical 3D modeling methods

方法	主要思想	输出形式	几何精度	主要局限
SfM (COLMAP)	增量式位姿估计 + BA	稀疏点云 + 相机位姿	高	大规模场景耗时; 依赖特征匹配质量
PMVS	面片匹配 - 扩展 - 过滤	稠密面片	中高	参数敏感; 计算开销大
COLMAP-PMVS	逐像素视图选择 + 深度图融合	稠密点云 / 网格	高	弱纹理区域深度估计不稳定

3 新兴融合技术

3.1 深度学习与传统几何的融合 (MVSNet、Point-MVSNet)

深度学习对MVS算法的优化改造,基本思路是以可训练神经网络模块替换人工设计的算法组件。传统MVS的工作流程中,从图像特征提取到代价匹配,每一步都由人工设计,而学习方法则采用卷积神经网络(CNN)来端到端地完成这些步骤。

MVSNet由Yao等人^[8]于2018年提出,网络工作流程以“特征提取—代价体构建—正则化—深度回归”为典型代表。2D CNN先从各输入图像提取多尺度特征,接着通过可微单应性变换(Homography Warping),将邻近视图特征映射到假设深度平面,构建存放代价的三维立方体,简称3D代价体。该可微操作的重要创新点,在于将传统MVS中手工设计的匹配代价,替换为梯度可回传的神经网络。该3D代价体经3D U-Net正则化后,沿深度方向做软argmin操作,回归出参考视图深度图。最后利用光度一致性,完成深度残差精化。

Chen等人2019年提出的Point-MVSNet^[10],瞄准了MVSNet的一个弱点,即深度图转点云过程中,三维一致性会损失。它的做法是先让MVSNet产出粗糙深度图,并转为初始点云,然后用特征增强模块,将2D图像特征投影到3D点上。精化阶段采用由粗到细(coarse-to-fine)策略,在三维空间中直接修正点位置。作者注意到,这种直接在3D空间操作的方式,比在2D深度图上回归更加贴近MVS问题本身的几何结构。

简言之,MVSNet和Point-MVSNet表面上是两种方案,底层的设计方案有一定的相通性。多视图几何里的单应性变换、代价体等重要概念得以保留,同时利用卷积神经网络(CNN)来学习特征和正则化处理,而非手工设计。

3.2 神经辐射场(NeRF)

NeRF^[1]由Mildenhall等人于2020年提出,其重要思

想是采用MLP模型,将连续的5D坐标 (x, y, z, θ, ϕ) ,映射为体积密度 σ 和颜色 c ,即 $(\sigma, c) = \mathbf{F}\Theta(\mathbf{x}, \mathbf{d})$ 。其中, $\mathbf{x} = (x, y, z)$ 代表空间位置, $\mathbf{d} = (\theta, \phi)$ 表示观察方向。MLP的输入除了空间坐标本身,还包含位置编码(Positional Encoding),即 $\gamma(\mathbf{x}) = [\sin(2^0\pi x), \cos(2^0\pi x), \dots, \sin(2^{(L-1)}\pi x), \cos(2^{(L-1)}\pi x)]$,可用于克服MLP对高频信号拟合能力不足的问题。

基于经典的体积渲染方程,可以实现对NeRF的渲染。对于每条相机射线,沿射线方向采样若干点,通过数值积分来累加各采样点的颜色和密度,其像素颜色的表达式为

$$\hat{C}(r) = \sum_i T_i [1 - \exp(-\alpha_i \delta_i)] c_i, \quad T_i = \exp(-\sum_{j=1}^{i-1} \alpha_j \delta_j) \quad (3-1)$$

其中, δ_i 为相邻采样点间距, T_i 为累积透射率。训练时,NeRF通过最小化渲染像素颜色 $\hat{C}(r)$ 与真实像素 $C(r)$ 之间的均方误差,实现对网络参数 Θ 的优化。

NeRF^[5]能产出高逼真度新视图图像的原因,在于它的隐式场景函数具有一定的连续性。在任意坐标上,MLP都能输出颜色和密度,不受网格分辨率约束。但持续性的代价是训练通常需要数十万次迭代,耗时数小时到数十小时。渲染一帧则需沿每条光线做密集MLP推理,速度被压在0.15 FPS以下。Instant-NGP^[11]通过多分辨率哈希编码,把训练压到秒级,但本身并没有实现对MLP的渲染瓶颈合理的解决方法。

3.3 3D高斯泼溅(3D Gaussian Splatting)

Kerbl等人于2023年提出了3D高斯泼溅(3D Gaussian Splatting, 3DGS)^[6],该方法代表了一种从隐式神经表征回归显式表征的转变。与NeRF使用MLP隐式编码场景不同,3DGS直接用一组可优化的3D高斯基元显式表示场景。每个高斯基元由以下参数定义,即中心位置 $\mu \in \mathbb{R}^3$ 、协方差矩阵 $\Sigma \in \mathbb{R}^{3 \times 3}$ 、不透明度 $\alpha \in [0, 1]$,以及用于表示视角相关外观的球谐函数系。

训练3DGS的过程中,依赖SfM^[3]预先产出的稀疏点云作为初始化数据。每个SfM点被初始化为一个各向同

性的高斯基元。在训练期间，自适应密度控制动态调整高斯总数。梯度大的区域表示欠重建，可执行克隆或分裂。对不透明度接近零的高斯基元，则做剪枝处理。这种增删策略使模型在保持紧凑的同时，捕捉场景的精细结构。

在渲染端，3DGS完全抛弃了NeRF沿光线逐点采样

的做法，改用基于瓦片（tile）的光栅化器。对于给定相机视角，所有3D高斯基元，先被投影到2D屏幕空间，按深度排序后，由GPU分块进行alpha混合。因为整个过程是高度并行的，渲染一帧的时间缩短到毫秒级别，同时保持了可微性，优化过程仍然可以通过梯度反传来实现。NeRF与3DGS的原理对比，如图3所示。

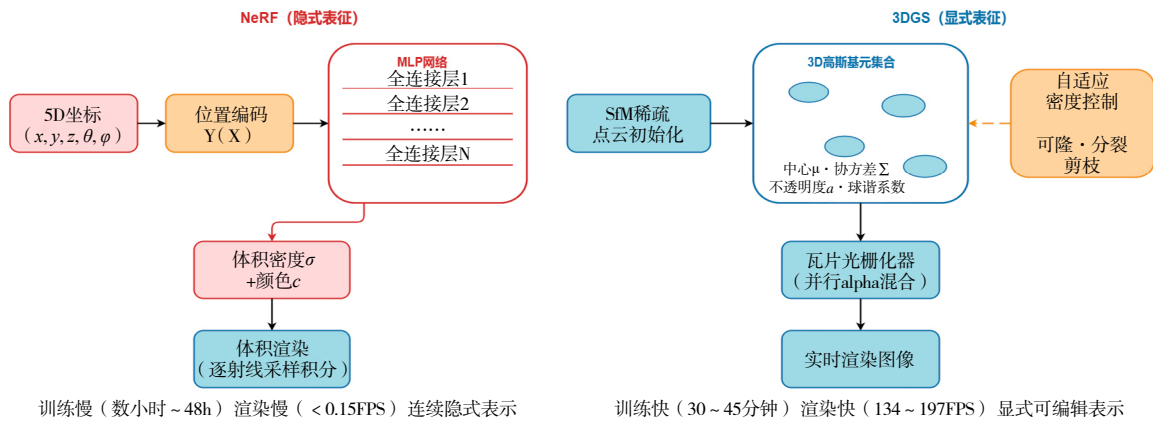


图3 NeRF与3DGS原理对比图
Figure 3 Principle comparison between NeRF and 3D Gaussian Splatting

3.4 隐式与显式表征的融合趋势

NeRF和3DGS恰好分布在场景表征光谱的两端。隐式端用诸如MLP的连续函数来编码场景，优点是内存消耗小、不受离散化限制，缺点是渲染过程需要密集采样的计算。显式端以高斯、点、面片为代表的离散基元，直接存储几何和外观，优点是渲染快、支持编辑，代价是内存随场景复杂度增长。

从技术发展历程来看，隐式与显式已开始相互渗透。可以将3DGS理解为对NeRF效率瓶颈的回应。它既保留了可微渲染的思想，也把场景从MLP搬进高斯基元。反之，社区也在尝试采用混合方案，即以显式几何引导神经网络训练，或用神经场来补充显式纹理细节。对几种新兴三维建模技术特征的对比分析，归纳总结如表2所示。

表2 新兴三维建模技术对比
Table 2 Comparison of emerging 3D modeling techniques

方法	表征类型	训练时间	渲染速度	视觉质量	可编辑性
NeRF ^[5]	隐式（MLP）	数小时 ~ 48 小时	<0.15 FPS	高	低
Instant-NGP ^[11]	隐式（哈希网格 + MLP）	秒级 ~ 5 分钟	11 ~ 17 FPS	中高	低
3DGS ^[6]	显式（高斯基元）	约 30 ~ 45 分钟	134 ~ 197 FPS	高	中
MVSNet ^[8]	显式（深度图）	预训练 + 推理	N/A	N/A	中

4 经典方法与新兴技术的比较分析

4.1 重建精度与视觉质量对比

评估三维重建方法时，精度和视觉质量是两项最重要的指标，但其关注点不完全重叠。经典方法以几何精度为导向，输出点云或网格，可用重投影误差和倒角距离（Chamfer Distance）来度量。COLMAP^[3, 7]和PMVS^[4]

在DTU^[9]上能给出高质量稠密重建。MVSNet^[8]和Point-MVSNet^[10]的几何精度，已在DTU上超越传统方法，印证了学习型特征在弱纹理区域的明显优势。

神经表征类方法优先考虑视觉质量，即考察合成图像看起来是否真实。量化评估指标上，则使用PSNR、SSIM和LPIPS。据3DGS论文展示^[6]，在Mip-NeRF 360上，3DGS-30K依次取得PSNR 27.21 dB、SSIM 0.815与

LPIPS 0.214的量化数值,同期Mip-NeRF 360则为27.69 dB、0.792、0.237。Tanks & Temples^[14]上,3DGS-30K为23.14 dB、0.841、0.183,而在室外大场景 Tanks & Temples 数据集的视觉渲染指标上,与Mip-NeRF 360相比,3DGS-30K具备明显优势。上述数据说明,显式高斯方法在真实复杂场景中的外观建模能力,不亚于隐式神经网络。这里需要增加一项界定。两类方法的“好”,是在不同维度上衡量的,不能简单地放在同一把尺子下做比较。经典方法产出的点云或网格,可直接用于测量和加工。神经方法的输出,本质上是一组渲染参数,主要服务于视觉呈现。由于逆向工程和尺寸检测这类任务对几何精度有硬性量化要求,在工业测量、精密逆向工程等场景下,经典几何重建路线的综合优势,短期内难以被现有神经表征方案替代。

4.2 计算效率与资源消耗对比

在计算效率维度上,新兴方法的优势明显。原始NeRF^[5]训练,需数小时到48小时。Instant-NGP^[11]以

多分辨率哈希编码为基础,压缩至5分37秒。而3DGS^[6]在6分25秒到41分33秒之间收敛,比Mip-NeRF 360快出60~80倍。

在渲染速度方面,两类方法的差距更为悬殊。对比各种数据集的渲染速度,Mip-NeRF 360仅为0.06 FPS,Instant-NGP为11.7 FPS,3DGS则达134~160 FPS。在Tanks & Temples上,3DGS的154~197 FPS,约是Mip-NeRF 360(0.14 FPS)的1000~3000倍。上述性能差距,源于3DGS的瓦片光栅化器对GPU并行计算的充分利用。

在资源消耗方面,3DGS模型文件在11~70 MB(T&T,7K步时达70 MB)和23~34 MB(Mip-NeRF 360)之间,比Plenoxels的2.1~2.3 GB低出两个数量级。但训练峰值GPU显存超过20 GB,对消费级显卡压力不小。Instant-NGP模型仅13 MB,且显存需求更温和。诸如COLMAP的经典方法,计算时主要集中在BA环节,大规模场景CPU耗时可达数天,但无需GPU。主要方法的渲染速度与训练时间对比,如图4所示。

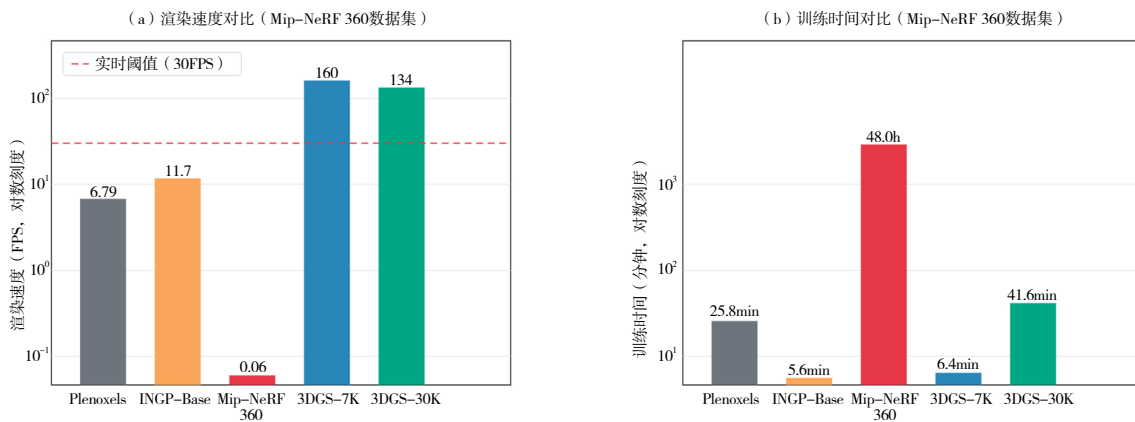


图4 主要方法渲染速度与训练时间对比

Figure 4 Comparison of rendering speed and training time among major approaches

4.3 适用场景与数据需求对比

不同方法对输入数据的要求和适用场景差异很大。经典SfM/MVS工作流程,通常需要几十到几百张图像,而且图像之间必须要保证充分重叠和合理的基线分布。它的好处是不需要GPU训练,诸如点云、网格的输出格式是标准的,可以直接丢进CAD软件做测量和编辑。

NeRF和3DGS对输入视角要求更为宽松,LLFF^[15]中每个场景,仅20~30张图像即可训练。但二者假设场景静态,动态场景需额外时序建模。在稀疏视角条件下,3DGS的表现优于NeRF,这是因为SfM点云初始化为高斯基元提供了几何先验。

与此同时,MVSNet^[8]和Point-MVSNet^[10]对多视角数据的密集度要求较高。DTU数据集^[9]中,每个场景包

含49个视角。由于预训练模型可直接泛化,这类方法的优势,在于推理阶段无需逐场景训练,但泛化能力受到训练数据覆盖的场景类型限制。

从输出过程看,经典方法输出网格/点云,适合需要几何精度的场景。NeRF输出隐式场,仅可实现渲染。3DGS输出显式高斯,既能实现渲染,也具备一定的编辑能力。可编辑性的排序可以表示为:经典>3DGS>NeRF。

4.4 综合对比

综合上述分析,在精度、效率、场景适用性和可编辑性等方面,各类方法均各有所长。三维建模方法的综合性能对比,如表3所示。

表 3 三维建模方法综合性能对比

Table 3 Comprehensive performance comparison of 3D modeling approaches

参数指标	NeRF ^[5]	3DGS ^[6]	MVSNet ^[8]	COLMAP ^[3,7]
几何精度	高	中	中	高
渲染速度	不适用	极慢 (< 0.15 FPS)	实时 (≥ 100 FPS)	不适用
训练时间	数小时 ~ 数天	数小时 ~ 48 小时	约 30 ~ 45 分钟	预训练 + 推理
数据需求	数十 ~ 数百张	稀疏 (~ 数十张)	稀疏 (~ 数十张)	多视角密集
静态 / 动态	静态	静态	静态	静态
输出类型	点云 / 网格	隐式 (渲染图像)	显式高斯 + 渲染	深度图 → 点云
可编辑性	高	低	中	中
PSNR (Mip-NeRF 360)	不适用	27.21 dB	27.69 dB	不适用
渲染 FPS (Mip-NeRF 360)	不适用	0.06	134 ~ 160	不适用

综合表3的对比信息，本章经过分析与推断，归纳出几点结论，简述如下。

在以工业测量、逆向工程为典型代表的场景，需要精确几何数据，经典方法凭借高精度点云/网格输出，以及与CAD/CAE工具链的天然兼容性，短期内仍是首选。在渲染质量和训练-推理效率之间，3DGS取得了最好的平衡，实时能力让它对VR/AR和数字孪生产生直接吸引力。NeRF虽已不是效率最优，但思想价值被后续工作持续吸收。MVSNet类方法在基准模板上，已经胜过传统工作流程，但跨场景泛化仍是硬约束。

5 实验与案例分析

5.1 实验设置与数据集

本文的定量分析基于3DGS论文^[6]中报告的公开实验数据，涉及四个评测数据集和五种代表性方法。实验数据集的选择，覆盖了合成场景和真实场景、室内物体和室外大场景，具有较好的代表性。

DTU^[9]含80场景×49视角，带结构光真值，用于桌面尺度的几何精度评测。Tanks & Temples^[14]面向真

实室外大场景，提供激光真值，可以分为Intermediate和Advanced两级。LLFF^[15]中包含8个前向面对场景，每个场景产生20~30张图像。Blender (NeRF Synthetic)^[5]含8个合成场景，包括100张训练视图与200张测试视图，图像分辨率为800×800。

参与本章实验对比的方法，包括以下几种：无神经网络的体素方法 (Plenoxels)、Instant-NGP^[11]基础配置 (INGP-Base)、NeRF系列高性能方法 (Mip-NeRF 360)、3DGS训练7K步 (3DGS-7K)，以及3DGS训练30K步 (3DGS-30K)，都在实验中有所体现。

所有实验在相同硬件条件下执行，图像量化评估指标包括三种：峰值信噪比 (Peak Signal-to-Noise-Ratio, PSNR, 越高越好)、结构相似性 (Structural Similarity Index Measure, SSIM, 越高越好)，可学习感知图像块相似度 (Learned Perceptual Image Patch Similarity, LPIPS, 越低越好)。此外，训练时间、渲染FPS以及模型内存占用，也是辅助的量化参数，用于评估实验的输出结果。

5.2 定量结果对比

Mip-NeRF 360数据集和Tanks & Temples数据集上的完整实验结果，分别列举如表4、表5所示。

表 4 真实场景Mip-NeRF 360数据集实验结果^[6]

Table 4 Experimental results on the real-scene Mip-NeRF 360 dataset^[6]

方法	PSNR ↑	SSIM ↑	LPIPS ↓	训练时间	渲染 FPS	内存
Plenoxels	23.08	0.626	0.463	25m49s	6.79	2.1GB
INGP-Base	25.30	0.671	0.371	5m37s	11.7	13MB
Mip-NeRF 360	27.69	0.792	0.237	48h	0.06	8.6MB
3DGS-7K	25.60	0.770	0.279	6m25s	160	23MB
3DGS-30K	27.21	0.815	0.214	41m33s	134	34MB

表 5 Tanks & Temples数据集实验结果^[6]

Table 5 Experimental results on the Tanks & Temples dataset^[6]

方法	PSNR ↑	SSIM ↑	LPIPS ↓	训练时间	渲染 FPS	内存
Plenoxels	21.08	0.719	0.379	25m5s	13.0	2.3GB
INGP-Base	21.72	0.723	0.330	5m26s	17.1	13MB
Mip-NeRF 360	22.22	0.759	0.257	48h	0.14	8.6MB
3DGS-7K	21.20	0.767	0.280	6m55s	197	70MB
3DGS-30K	23.14	0.841	0.183	26m54s	154	11MB

合成场景（NeRF Synthetic/Blender）上的各种方法平均PSNR结果，如表6所示。

表6的量化结果显示，合成场景中五种方法的差距

较小，3DGS以33.32 dB略优于其他方法，反映出在理想光照和几何条件下，显式与隐式表征的性能上限趋于一致。

表 6 合成场景中的不同方法平均PSNR结果

Table 6 Average PSNR of different approaches in synthetic scenes

方法	Plenoxels	INGP-Base	Mip-NeRF	Point-NeRF	3DGS-30K	平均
PSNR	31.76	33.18	33.09	33.30	33.32	32.93

5.3 结果讨论

上述表格中显示了五种方法在视觉质量与工作效率等关键基准点上的性能数据，可以从以下几个角度来解读这些结果。

视觉质量方面，3DGS-30K在两个真实数据集上的PSNR为27.21 dB和23.14 dB，与Mip-NeRF 360的27.69 dB和22.22 dB接近。但SSIM和LPIPS的数值分化更为明显。例如，Tanks & Temples上，3DGS-30K的SSIM为0.841（对手0.759，+10.8%），LPIPS为0.183（对手0.257，-28.8%）。推断可知，3DGS的显式高斯，在表示锐利边缘和高频纹理时更直接。而NeRF的体积分，则在边缘处天然引入平滑。

考察实现效率，可知上述方法具有数量级的差距。3DGS训练耗时在26分54秒至41分33秒之间，而Mip-NeRF 360则需48小时，前者相当于后者的1/70 ~ 1/107。渲染速度上，3DGS为134 ~ 197 FPS，Mip-NeRF 为0.06 ~ 0.14 FPS，前者是后者的1000 ~ 3000倍。Instant-NGP训练最快（约5分钟），但渲染速度仅为11 ~ 17 FPS，并且视觉上PSNR数值低约2 dB。这说明哈希编码优化了训练速度，但MLP推理延迟仍存在。而3DGS则用显式表示与光栅化的组合，同时解决了两端的问题。

从资源占用上看，3DGS的模型文件，在11 ~ 70 MB之间波动，与Instant-NGP（13 MB）和Mip-NeRF 360（8.6 MB）处于同一量级，比Plenoxels（2.1 ~ 2.3 GB）低出两个数量级。但在Tanks & Temples上7K步训练时，内存达到70 MB，这是其自适应密度控制在复杂场景中产生大量高斯导致的。训练期间峰值GPU显存超过20 GB，消费级显卡跑起来压力不小。

综合以上定量、定性分析，对各种方法的适用场景可简述如下。经典方法在几何精度和可编辑性上的优势，使其在工程测量和逆向设计领域不可替代。Mip-NeRF 360的渲染质量有竞争力，但48小时训练和0.06 ~ 0.14 FPS推理速度，限制了其工程应用的推广。Instant-NGP适合需要快速预览的场景。尽管3DGS的几何精度仍不及经典工作流程，大场景显存开销值得关注，但3DGS在质量、训练速度和渲染帧率上取得了最好的综合表现，其实时渲染能力，也为VR/AR和数字孪生提供了可用方案。

6 结论与展望

6.1 研究结论

本文系统比较了三维建模领域中基于几何驱动的传统方法与基于神经表征的新兴方法。基于文献调研和公开实验数据的评估，可得出以下结论。

经典方法与多视图几何理论^[1]一脉相承，其技术路径是先通过SfM^[3]来恢复相机位姿和稀疏点云，再由MVS^[4,7]来补全稠密几何结构。这种实现方案在几何精度和可编辑性上保持优势，适合工业测量和逆向工程。其技术短板在于依赖手工特征，并且在大规模场景下，BA的计算量膨胀得很快。

新兴融合技术中，各工作的贡献各有侧重。NeRF^[5]开创了MLP编码场景的方式，证明隐式辐射场在新视角合成上的可能性。3DGS^[6]以显式高斯基元为特征，在Mip-NeRF 360上以27.21 dB PSNR和134 FPS高效运行，其渲染速度比NeRF快三个数量级。Instant-NGP^[11]将训练压进秒级，适合快速预览。MVSNet^[8]和Point-MVSNet^[10]将深度学习引入MVS，在DTU^[9]上的几何精度，已超越传统方法，只是跨场景泛化仍是最大缺口。

从技术发展历程看，三维建模正由纯几何驱动逐渐向数据驱动发展，由隐式走向显式。3DGS的成功，说明显式表示的效率优势，足以弥补理论优雅性的不足。隐式与显式的融合，可能会代表下一阶段的重要发展方向。

6.2 未来发展方向

结合全文分析，三维建模领域各类新型融合技术，未来具备研究价值的发展方向如下。

当前各类方法中，动态场景建模是一项共同的短板。经典SfM/MVS、NeRF和3DGS均假设场景静态。将时序信息引入神经辐射场或高斯泼溅框架，实现非刚性变形和运动物体的三维建模，可以代表目前前沿的研究方向，具有广泛应用前景。

上述方法从实验室数据集搬到城市级，或更大尺度的场景中时，可扩展性问题立刻凸显。对3DGS而言，室外大场景会显著增加显存占用率。对经典方法而言，大规模重建中的BA计算开销同样不可忽视。常规的潜在发展方向，以层次化表示和流式处理为代表。

在同一类优化框架中,改善几何精度和视觉质量,值得进一步探索。现阶段的神经方法渲染质量出色,但几何精度不足。经典方法几何可靠,但视觉效果受限。将几何先验注入神经场训练,或者反用神经场来修饰显式几何的纹理,可以兼得两者之长。此外,在MVSNet和Point-MVSNet这类学习型方法中,跨场景泛化能力是其主要瓶颈之一。在大规模数据集上预训练其通用特征表示,结合小样本微调策略,可以提升方法在未见场景上的适应性。

参考文献

- [1] Hartley R, Zisserman A. Multiple View Geometry in Computer Vision [M]. 2nd ed. Cambridge: Cambridge University Press, 2004.
- [2] Scharstein D, Szeliski R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms [J]. Int J Comput Vis, 2002, 47 (1/3): 7–42.
- [3] Schönberger J L, Frahm J M. Structure-from-motion revisited [C] // Proc IEEE Conf Comput Vis Patt Recognit (CVPR), Las Vegas: 2016: 4104–4113.
- [4] Furukawa Y, Ponce J. Accurate, dense, and robust multi-view stereopsis [J]. IEEE Trans Patt Anal Mach Intell, 2010, 32 (8): 1362–1376.
- [5] Mildenhall B, Srinivasan P P, Tancik M, et al. NeRF: Representing scenes as neural radiance fields for view synthesis [C]. Proc Europ Conf Comput Vis (ECCV) Cham: Springer, 2020: 405–421.
- [6] Kerbl B, Kopanas G, Leimkühler T, et al. 3D Gaussian splatting for real-time radiance field rendering [J]. ACM Trans Graph, 2023, 42 (4): 139.
- [7] Schönberger J L, Zheng E, Pollefeys M, et al. Pixelwise view selection for unstructured multi-view stereo [C]. Proc Europ Conf Comput Vis (ECCV) Cham: Springer, 2016: 501–518.
- [8] Yao Y, Luo Z, Li S, et al. MVSNet: Depth inference for unstructured multi-view stereo [C]. Proc Europ Conf Comput. Vis. (ECCV). Cham: Springer, 2018: 767–783.
- [9] Jensen R, Dahl A, Vogiatzis G, et al. Large scale multi-view stereopsis evaluation [C]. Proc IEEE Conf Comput Vis Patt Recognit. (CVPR), Columbus: IEEE, 2014: 406–413.
- [10] Chen R, Han S, Xu J, et al. Point-based multi-view stereo network [C]. Adv Neural Inform Process Syst (NeurIPS). Vancouver: NeurIPS, 2019: 1538–1548.
- [11] Müller T, Evans A, Schied C, et al. Instant neural graphics primitives with a multiresolution hash encoding [J]. ACM Trans Graph, 2022, 41 (4): 102.
- [12] 李吉洋, 程乐超, 何靖璇, 等. 神经辐射场的研究现状与展望 [J]. 计算机辅助设计与图形学学报, 2024, 36 (7): 995–1013.
- [13] 高翔, 李梦晗, 申抒含. 大规模场景运动恢复结构研究综述 [J]. 计算机辅助设计与图形学学报, 2024, 36 (7): 969–994.
- [14] Knapitsch A, Park J, Zhou Q Y, et al. Tanks and temples: benchmarking large-scale scene reconstruction [C]. Proc IEEE Conf Comput Vis Patt Recognit (CVPR), Honolulu: IEEE, 2017: 3217–3226.
- [15] Mildenhall B, Srinivasan P P, Ortiz-Cayon R, et al. Local light field fusion: practical view synthesis with prescriptive sampling guidelines [J]. ACM Trans Graph, 2019, 38 (4): 1–14.

Comparative Analysis of Classical Methods and Emerging Fusion Technologies for 3D Modeling

Fang Ziwei Wen Jing Huang Yumeng Gao Xin

College of Physics and Optoelectronic Engineering, Jiangsu Normal University, Xuzhou

Abstract: Reconstructing 3D structures from 2D images has long been a focal research topic in computer vision and computer graphics. Four representative technical routes named as structure from Motion (SfM), Multi-View Stereo (MVS), Neural Radiance Fields (NeRF), and 3D Gaussian Splatting (3DGS) are selected by this paper. With respect to the public datasets of two benchmarks, Mip-NeRF 360 and Tanks & Temples, quantitative evaluations on their reconstruction accuracy, rendering efficiency, resource consumption, and scene applicability are conducted. Experimental results show that 3DGS achieves a leading real-time rasterization speed of 134~197 FPS, which is approximately 1000~3000 times faster than Mip-NeRF 360. The training time of 3DGS is reduced to 35~45 minutes, while its PSNR (27.21 dB) is nearly comparable to that of Mip-NeRF 360 (27.69 dB). The classic workflow still maintains advantages in geometric accuracy and editability, while neural representations demonstrate remarkable superiority for novel view synthesis under sparse-view settings.

Key words: 3D reconstruction; Neural radiance fields; 3D Gaussian splatting; Multi-view stereo; Novel view synthesis