arXiv:2601.11617v1 [cs.CV] 10 Jan 2026

PointSLAM++：基于鲁棒稠密神经高斯点云的同步定位与建图

Xu Wang[1*]、**Boyao Han**[1*]、**Xiaojun Chen**[1],**Ying Liu**[2†]、**Ruihui Li**[1†]

[1]中国湖南大学计算机科学与电子工程学院 [2]中国湖南师范大学信息科学与工程学院{xuwang0303, hby 22, chenxiaojun}@hnu.edu.cn, liu ying@hunnu.edu.cn, liruihui@hnu.edu.cn

摘要

实时三维重建对于机器人学和增强现实至关重要，但在存在深度噪声的情况下，当前的同时定位与建图（SLAM）方法往往难以保持结构一致性并实现稳健的位姿估计。本文提出 PointSLAM++，一种新颖的 RGB-D 同步定位与建图系统，它采用分层约束神经高斯表示，在生成用于场景建图的高斯基元的同时保留结构关系。该系统还采用渐进式位姿优化来减轻深度传感器噪声的影响，从而显著提高定位精度。此外，系统利用动态神经表示图，根据局部几何复杂度调整高斯节点的分布，使地图能够实时适应复杂的场景细节。这种组合实现了高精度三维建图和照片级真实感场景渲染。实验结果表明，PointSLAM++在重建精度和渲染质量方面优于现有基于 3DGS 的同步定位与建图方法，展现出其在大规模增强现实和机器人学领域的优势。

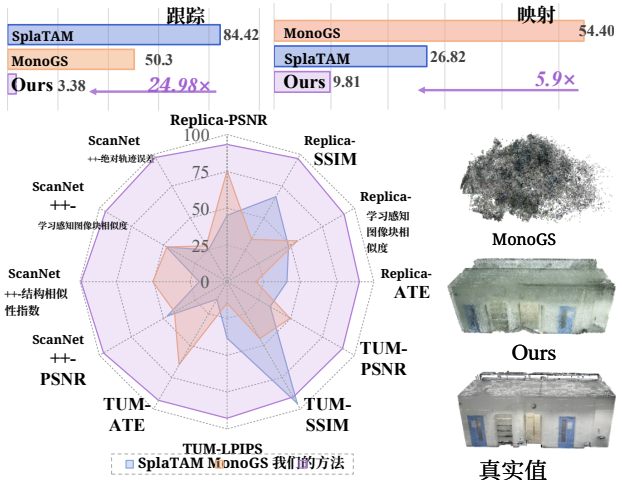

图 1：PointSLAM++在逼真的三维重建和相机跟踪方面优于 CVPR 2024 的最先进方法，并在关键指标上表现出色。在 MonoGS 失效的复杂环境中，它仍能保持准确定位和高质量三维建图。

引言

视觉同步定位与建图利用单目、双目或 RGB-D 图像重建三维场景并估计相机轨迹。它在机器人学、虚拟现实和增强现实中的广泛应用，进一步提高了对高保真渲染和稳健运动估计的需求（Zubizarreta、Aguinaga 和 Montiel，2020）。传统的 RGB-D 同步定位与建图（基于特征或体素）能够确保稳定跟踪，但重建细节较为粗糙（Du 等，2011；Keller 等，2013；Newcombe 等，2011）；而神经隐式方法（例如 iMAP（Sucar 等，2021）、NICE-SLAM（Zhu 等人，2022a））虽然能够实现稠密重建，却需要较高的计算成本，且适用规模有限。这种效率与质量之间的权衡推动了新型同步定位与建图表示的发展（Bylow 等，2013；Canelhas、Stoyanov 和 Lilienthal，2013；Dai 等，2017；Prisacariu 等，2017；Whelan 等，2013）。近期，三维高斯泼溅（3DGS）（Kerbl 等，2023）应运而生

作为一种混合范式，它采用各向异性高斯和可微分泼溅，将点云效率与神经渲染保真度相结合（Zhu 等人，2024b）。MonoGS（Matsuki 等人，2024）等系统展示了实时的照片级逼真重建以及位姿—几何联合优化，但传感器噪声、动态场景和大规模环境仍然带来了挑战（Huang 等人，2024；Keetha 等人，2024；Matsuki 等人，2024；Yan 等人，2024；Hu 等人，2024）。

当前的3DGS-SLAM方法面临三大挑战。第一，依赖精确的深度信息：过度依赖高质量深度图，导致在出现深度噪声、缺失数据或遮挡时，跟踪和重建的精度下降（Chung 等，2023；Deng 等，2024）。第二，高斯球过拟合：现有技术会针对每个训练视图拟合一个高斯球，却忽略局部场景结构，因此产生大量冗余，并限制了其在复杂大规模环境中的可扩展性（Zhang 等，2023）。第三，视角依赖性不足：视角相关效应被烘焙到单个高斯参数中，导致插值能力较差，且对大幅视角和光照变化的鲁棒性较低。解决这些问题对于实现可扩展、高保真度的3DGS-SLAM至关重要。

*这些作者贡献相同。
†通讯作者。

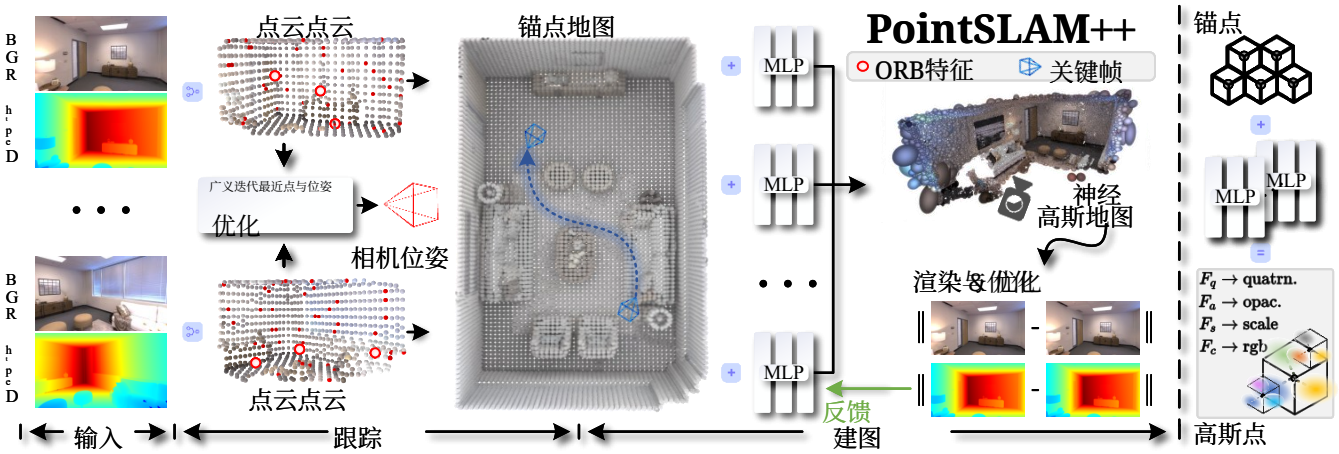

图 2：SLAM系统概述。系统的输入是一系列 RGB-D帧。我们对当前深度图像进行下采样和重投影，生成点云，并使用广义迭代最近点和 ORB特征估计当前位姿。在跟踪过程中，我们根据点云创建锚点，并利用神经网络预测场景中的高斯分布，再通过专用的高斯光栅化器渲染场景。在建图过程中，我们利用 RGB-D信息持续优化多层感知机（MLP）。右侧：我们将锚点和特征作为 MLP 的输入，以预测高斯点的各种属性。

本文介绍PointSLAM++，这是一种使用RGB-D相机的实时3D重建系统。该系统采用带有层次化约束和渐进式位姿优化的可学习神经高斯表示，能够实现精确建图和逼真渲染。

为摆脱对精确深度信息的依赖，我们提出一种渐进式优化方案，以实现准确的位姿估计：采用迭代最近点（ICP）算法对深度数据进行刚性配准，并结合基于ORB特征的视觉约束进行联合优化。在多尺度金字塔中，我们同时最小化深度点云残差和基于ORB的视觉重投影误差（Campos等人，2021年）。然而，在高速运动、低纹理区域或动态干扰等具有挑战性的条件下，跟踪仍可能丢失。为在传统跟踪失效时快速恢复相机位姿，我们引入了一种基于全局特征匹配的重定位机制。

为满足实时增量建图和高精度重建的需求，我们将跟踪得到的稳定 ORB 特征点云与深度传感器数据融合，用于地图构建。随后，将点云转换为具有位置、尺度及相关参数的神经高斯“锚点”，并根据场景的几何结构约束其分布。两级锚点机制包括用于全局结构的主锚点和用于精细细节的次锚点，能够在保持稳健位姿估计和几何精度的同时，实现增量更新并丰富细节。最后，将相机视角向量嵌入每个锚点的外观特征，以补偿视点相关光照和反射率效应，从而提升复杂光照和视点变化条件下的重建质量。

总而言之，我们的贡献如下：

- 我们提出了 PointSLAM++，这是一种实时 RGB-D 同步定位与建图系统，将几何与视觉位姿优化同分层神经高斯表示相结合，以实现精确建图和照片级真实感渲染。
- 我们开发了 a 渐进式多尺度 a位姿优化 a ，通过联合最小化 ICP 深度残差和 ORB 重投影误差来实现，并添加了一个全局特征重定位模块，以应对快速运动或低纹理跟踪失败。
- 我们提出了 a 双层神经 al-Gaussian a锚点框 a 架构，将 ORB特征云与深度数据融合，并嵌入相机视角向量，以自适应增强细节并补偿视角相关光照。

相关工作

经典 RGB-D同时定位与建图

传统 RGB-D 稠密同步定位与建图经历了三个阶段：KinectFusion（Newcombe 等，2011）通过迭代最近点（Gelfand 等，2003）实现了实时 TSDF 重建；ElasticFusion（Whelan 等，2016）通过基于图的非刚性变形，增强了大规模场景的拓扑一致性；BundleFusion（Dai 等，2017）则通过分层优化和全局闭环检测提升了鲁棒性。然而，基于体素和点的语义同步定位与建图仍然计算开销大且结果不一致。基于符号距离函数的方法虽然精度高，却因采用均匀网格而受限于内存占用和可扩展性（Wang、Skorokhodov 和 Wonka，2022）。此外，条件随机场分割等后处理方法会因优化相互独立而导致标签错位和边界模糊。



图 3：TUM 场景中的广义迭代最近点不匹配。正如 TUM 数据集中的示例所示，即使深度数据精度很高，真实场景噪声也可能导致基于广义迭代最近点的位姿估计出现对齐误差。

基于 NeRF 的 RGBD 同步定位与建图

神经辐射场（Mildenhall 等，2020）（NeRF）强大的场景表示能力推动了其在三维重建和定位中的应用。基于 NeRF 的同步定位与建图通过多层感知机构建连续隐式函数，将三维坐标映射为辐射度和密度（Wang、Wang 和 Agapito，2023；Johari、Carta 和 Fleuret，2023）。这克服了基于体素和点的模型在分辨率上的限制，使静态场景达到亚体素精度（例如 iMAP（Sucar 等，2021）、NICE-SLAM（Zhu 等人，2022a））。然而，优化效率、动态场景解耦和可扩展性方面仍面临挑战，阻碍了其在大型真实环境中的鲁棒性（Jiang 等，2023）。

基于高斯的 RGBD 同步定位与建图

三维高斯泼溅（Kerbl 等，2023）近年来重塑了 RGB-D 同步定位与建图。高斯泼溅同步定位与建图（Matsuki 等人，2024年；Liso 等，2024年）等方法利用解析雅可比矩阵和几何正则化实现照片级真实感重建，精度超越神经辐射场。GS-ICP SLAM（Ha、Yeon 和 Yu，2024年）进一步将高斯协方差整合进 G-ICP 框架，以两倍于传统迭代最近点的速度实现亚毫米级对齐。仍待解决的挑战包括深度噪声对协方差初始化的影响、对强正则化的需求，以及由于离线分割或简化的运动假设而导致的动态场景建模能力有限（Zhu 等，2024a；Mao 等，2024年）。增量优化也会在大规模建图中造成对齐误差。要将同步定位与建图实际部署，必须解决这些问题，实现抗噪正则化和可扩展动态建模。

方法

我们的重建流程概览如图 2 所示。在第 3.1 节中，我们首先介绍渐进式位姿优化（PPO）框架。然后，在第 3.2 节中，我们描述神经高斯表示及其密度自适应优化策略。最后，在第 3.3 节中，我们详细介绍基于神经高斯表示和渐进式位姿优化的完整在线重建过程。

渐进式位姿优化

在 c 问题中，首要目标是在全局坐标系内准确估计相机轨迹。为实现高度鲁棒的位姿定位，我们提出了一种渐进式优化框架，通过级联阶段系统地细化位姿估计，并将点云配准与基于特征的细化相结合，以减轻现实场景中的常见误差来源。

我们首先采用广义迭代最近点（GICP）作为前端里程计，用于粗略位姿估计。给定来自前一帧的局部点云 $\mathcal{P} = \{p_i \mid i = 1, 2, \ldots, N\}$ 和来自当前帧的局部点云 $\mathcal{Q} = \{q_i \mid i = 1, 2, \ldots, N\}$，最优变换 $\mathbf{T} \in SE(3)$ 通过最小化配准误差来求得：

$$\mathbf{T}^* = \arg\min_{\mathbf{T}} \sum_{j=1}^{N} \sum_{i \in C(j)} d(\mathbf{T} \cdot q_j, p_i)^2 \tag{1}$$

其中，$C(j)$ 表示点 q_j 的对应集，$d(\cdot)$ 表示距离度量。GICP 通过迭代优化对应关系和变换，得到初始位姿估计。

尽管准确的深度数据有助于提升广义迭代最近点的效果，但真实场景噪声常会导致错配（图 3）。为缓解这一问题，我们利用 ORB特征的鲁棒性，并采用更可靠的深度数据。通过多尺度金字塔提取并与深度数据融合后，这些特征形成局部点云，并通过点到平面约束配准至全局地图；配准以粗略 GICP位姿$\mathbf{T}$ 初始化，从而实现更快的收敛和更高的精度。

基于这一基础，我们的框架进一步通过光束法平差最小化重投影误差来细化位姿，同时优化相机位姿和三维路标点。这种渐进式策略独特地将里程计、特征驱动配准和优化无缝融合：使用 GICP+ORB 进行初始化，将第一帧固定以消除尺度歧义，在 $SE(3)$ 流形上对位姿进行参数化，并采用带有 Huber损失和深度先验的鲁棒 Levenberg–马夸尔特求解器，以实现稳定收敛。核心目标函数将这些要素统一起来：

$$\min_{\xi} \sum_{k=1}^{K} \rho\left(e_k^T \Sigma_k^{-1} e_k\right) \tag{2}$$

其中：- $\xi \in \mathfrak{se}(3)$ $T = \exp(\xi^\wedge) \in SE(3)$ 对位姿进行参数化；- K 表示观测特征的数量；- $e_k = x_k - \pi(KTX_k) \in \mathbb{R}^2$ x_k 表示重投影误差，其中 为观测像素，π 为投影函数，K 为相机内参，X_k 为三维点；- $\Sigma_k^{-1} \in \mathbb{R}^{2\times 2}$ 为信息矩阵；- ρ 为用于离群点剔除的鲁棒核。

位姿恢复与优化。 在高速运动、模糊、稀疏纹理或动态场景等具有挑战性的条件下，跟踪可能会失效。我们的恢复机制创新性地将特征匹配与成熟技术相结合，以可靠地恢复位姿。通过当前特征与

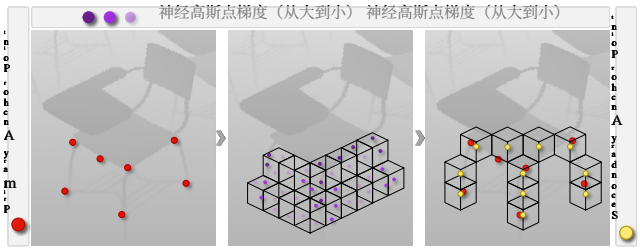

图4：**分层锚点优化**。 初级锚点（红色）源自ORB特征，并在整个同步定位与建图过程中保持稳定；次级锚点（黄色）则根据对神经高斯点进行的逐体素梯度检查（紫色）动态引入或剔除。

地图点之间的描述子匹配，形成初始对应集 。随后，嵌入RANSAC 循环中的透视-n-点（PnP）求解器根据公式 (2)，通过最小化重投影误差并结合鲁棒核 ρ 处理错误匹配，从而计算出稳健的初始位姿。

随后，非线性优化使用与公式（2）相同的李代数表达和目标函数对该估计进行细化。这种分层架构——粗略GICP、ORB增强配准、光束法平差以及PnP+RANSAC恢复——使我们的方法脱颖而出，通过提供自适应鲁棒性，超越了现有位姿跟踪系统中对这些组件的孤立应用。当内点率超过预定义阈值时，重新定位能够以较高置信度成功，从而增强系统在复杂环境中的整体韧性。

神经高斯表示

高质量的三维重建依赖于准确的地图初始化。COLMAP生成的稀疏点云虽可作为有效先验，但并不适用于同步定位与建图应用。为此，我们提出利用同步定位与建图过程中提取的稳定ORB特征点云，并辅以深度传感器数据，实现高效、轻量的初始化。

启动时，建图线程使用第一帧的ORB特征点云初始化全局地图。随后，线程逐步纳入每个后续关键帧中的特征点，持续更新并扩展地图。为实现高保真场景重建，我们将传统ORB点云转换为表征能力更强的神经点云：

$$N_v = \{(p_v, f_v, l_v, O_v)\} \tag{3}$$

每个神经高斯点由体素 v的中心位置$p_v \in \mathbb{R}^3$、局部上下文特征 $f_v \in \mathbb{R}^{32}$、缩放因子 $l_v \in \mathbb{R}^3$以及 k 可学习偏移量 $O_v \in \mathbb{R}^{k\times 3}$组成。与Scaffold-GS方法（Lu等人，2024年）类似，我们将这些点称为“锚点”。

每个锚点通过多层感知机（MLP）生成 K 神经高斯模型。每个神经高斯模型定义如下：

$$G = \{(\mu, \alpha, q, s, c)\} \tag{4}$$

其中， $\mu \in \mathbb{R}^3$ 是三维位置， $\alpha \in \mathbb{R}$ 是不透明度，$q \in \mathbb{R}^4$ 是控制协方差的四元数， $s \in \mathbb{R}^3$ 是缩放向量， $c \in \mathbb{R}^3$ 是颜色。

具体而言，神经高斯模型的三维位置与其可学习偏移量之间的关系如下：

$$\{\mu_0, \ldots, \mu_{k-1}\} = x_v + \{O_0, \ldots, O_{k-1}\} \cdot l_v \tag{5}$$

其中， $\{O_0, O_1, \ldots, O_{k-1}\} \in \mathbb{R}^{k\times 3}$ 表示一组可学习偏移量， l_v 是与锚点相关联的缩放因子。

每个锚点生成的 k 神经高斯的属性集合 A由多层感知机网络F_a计算得出，如下所示：

$$\{A_0, \ldots, A_{k-1}\} = F_a(\hat{f}_v, \delta_v c, \vec{d}_v c) \tag{6}$$

其中， A 表示神经高斯的任意属性（不透明度 α、颜色 c、四元数 q或缩放 s）。$\{A_0, \ldots, A_{k-1}\}$ 表示单个锚点生成的 k 神经高斯的该特定属性集合。随后，使用高斯渲染器进行渲染。 F_a 表示用于计算特定属性 A的多层感知机网络。 $\delta_v c$ 和 $\vec{d}_v c$ 分别表示相机与锚点之间的相对观测距离和方向。具体计算细节将在补充材料中提供。

分层锚点优化。 与通过运动恢复结构重建初始化的点云不同，同步定位与建图算法维护的地图是一种动态、增量式表示。在跟踪线程终止之前，建图与跟踪线程共享同一张地图。因此，对锚点进行大幅修改会直接影响相机位姿估计。

为了解决这一问题，我们引入了分层锚点。由 ORB特征生成的锚点被指定为主锚点，不允许分裂或删除，并以较小的学习率进行优化。由于 ORB特征在纹理稀疏区域的建模能力有限，我们引入了由深度传感器数据和更小体素生成的次级锚点。主锚点和次级锚点统称为 N_v。

为确定哪些位置需要次级锚点，我们构建尺寸为 ϵ_g 的体素，对神经高斯点进行空间量化。对于每个体素，我们计算经过 N 次训练迭代后的平均梯度 ∇g。如果 $\nabla g > \tau_g$ 且该体素中不存在锚点，则利用该体素中心的深度信息部署一个新的次级锚点。如果该体素中已存在次级锚点，则将其不透明度减半，直至低于阈值并被剔除。

视点相关环境补偿。 相机运动期间，场景外观会因视点变化而改变。这些变化源于环境光照变化对相机曝光的影响，以及镜面反射等材质属性所导致的视点相关辐射。这些因素要求位于同一位置的神经高斯能够从不同视点渲染出不同的像素值。然而，传统高斯模型受各向同性假设限制，即使采用球谐函数，也无法充分处理这些视点相关的外观变化。

为解决这一局限，我们将相机观察方向编码为单位向量$\mathbf{v} \in \mathbb{R}^3$，并将其与锚点外观特征$\hat{\mathbf{f}}_a$一同嵌入上下文特征中：

$$\mathbf{f}'_a = \mathrm{Embed}(\hat{\mathbf{f}}_a,\ \mathbf{v}) \tag{7}$$

其中，$\hat{\mathbf{f}}_a$是锚点外观特征（即 RGB 颜色），$\mathbf{v}$是归一化的视线方向向量，Embed(·)由多层感知机实现。输出$\mathbf{f}'_a$是视线方向增强的上下文特征，它作为上下文特征$\mathbf{f}_v$的初始形式，并在训练过程中持续更新。因此，我们将该初始上下文特征记为$\mathbf{f}_a$。这种方法显式地将视线方向与外观特征相结合，克服了传统方法在动态视点方面的局限性。

SLAM系统

We h我们开发了一种基于神经高斯表示和渐进式位姿优化框架的实时 SLAM 系统。假设我们有一张由一组三维高斯表示的地图，该地图由之前第 1 帧至 t 帧相机帧构建。对于新的 RGB-D 帧 $t+1$，我们的 SLAM 系统执行以下步骤：

- **相机跟踪：** 我们利用当前 RGB-D 帧 $t+1$ 的图像和深度信息，通过最小化当前帧与全局地图之间的重投影误差和深度误差来估计相机位姿。首先使用广义迭代最近点获得粗略估计，然后基于 ORB 特征的点云约束和重投影误差进行优化。
- **高斯点加密：** 被当前帧关键帧时，我们根据该帧的 ORB 特征点及其对应的深度信息，向地图中添加新的锚点。这增强了地图表示细节的能力。
- **地图更新：** 如果当前帧 a帧 $t+1$ 被选为 as a 关键帧，我们通过最小化当前关键帧以及与其存在重叠的历史关键帧上的 RGB 和深度误差，来更新地图中所有神经高斯点的参数。如果当前帧不是关键帧，则会随机选取一部分历史关键帧进行优化，以持续优化地图并保持一致性。

关键帧选择。 我们采用一种动态关键帧选择策略，类似于 GS-ICP SLAM（Ha、Yeon 和 Yu，2024），并以几何一致性为依据。当当前帧与地图之间的平均点云距离超过阈值，或低于该阈值的点过少时，就会选取新的关键帧。为减轻误差累积，仅在新关键帧的非重叠区域添加新的高斯点。

地图更新。 在锚点优化过程中，我们采用多种损失函数——颜色损失、SSIM（基于三维高斯泼溅）损失和几何损失——以确保渲染图像的颜色保真度、结构一致性和几何连贯性。此外，尺度正则化项进一步引导高斯椭球趋向各向同性球体，从而减少各向异性，并在跟踪精度与渲染真实感之间取得平衡。

$$\begin{aligned}\mathcal{L}_{mapping} = \lambda_{I_1}\mathcal{L}_1\left(I, I_{gt}\right) + \lambda_{I_2}\mathcal{L}_{D\text{-}SSIM}\left(I, I_{gt}\right) \\ + \lambda_D\mathcal{L}_1\left(D, D_{gt}\right)\end{aligned} \tag{8}$$

实验

实验设置

实现细节。 我们的同步定位与建图系统运行于一台配备 Intel Xeon Silver 4314 CPU 和 Nvidia RTX 3090 GPU 的台式计算机上。建图和跟踪组件使用 Python 实现，并基于 PyTorch 框架构建。

数据集。 我们在三个数据集上评估所提方法。Replica（Straub 等，2019年）提供高精度合成 RGB-D 图像，适用于基础验证。TUM-RGBD 数据集（Sturm 等人（2012））具有运动捕捉系统提供的精确相机位姿，是评估同步定位与建图跟踪精度的标准数据集，尽管其图像质量较低且深度数据存在噪声。ScanNet++（Yeshwanth 等，2023）是一个大规模数据集，包含高质量和普通质量的室内场景数据，具有较大的相机位姿间隔和深度图误差，可用于测试系统在相机快速运动或纹理稀疏条件下的鲁棒性。

基线方法。 我们将所提方法与最先进的 RGB-D 同步定位与建图方法进行比较，包括 NICE-SLAM（Zhu 等，2022b）、Point-SLAM（Sandstr¨om 等，2023），以及同期的高斯 SLAM 方法，例如 SplaTAM（Keetha 等，2024）、MonoGS（Matsuki 等人，2024）、Photo-SLAM（Huang 等，2024）和 GS-ICP SLAM（Ha、Yeon 和 Yu，2024）。为确保结果可信，我们使用各方法提供的官方代码复现了实验。对于 ScanNet++ 数据集（Yeshwanth 等，2023），由于其视点分布稀疏，我们将所有方法的场景（即 SLAM 中的地图）更新频率提高一倍，以确保比较公平。

重建地图的质量

Ta表 Point SLAM++在 TUM-RGBD（Sturm 等人（2012））和 Replica（Straub 等，2019年）数据集的 RGB-D 场景新视角渲染中，表现显著优于其他方法。

在 Replica（Straub 等，2019年）数据集上，由于其深度数据质量高且场景较简单，PointSLAM++ 在所有序列中均取得最佳结果。在对比方法中，最先进的 GS-ICP SLAM 渲染精度排名第二。PointSLAM++ 利用神经高斯技术，成功超越 GS-ICP SLAM。其独特的锚点优化技术巧妙地解决了外观变化问题，使 PointSLAM++即使在理想条件下也表现出色，渲染质量优于 GS-ICP SLAM。

TUM-RGBD数据集（Sturm 等人（2012））在小物体、运动模糊和不完整深度图方面带来了更大的挑战。尽管存在这些困难，PointSLAM++ 仍保持了最高的渲染准确率

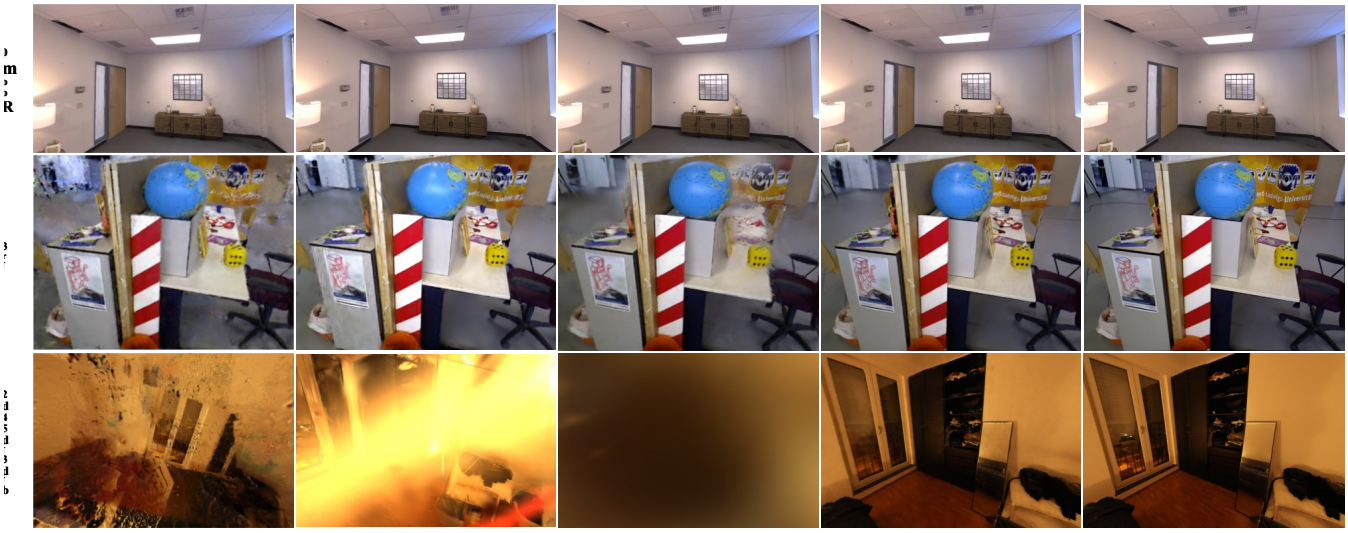

图5：**渲染结果对比**。我们展示了来自三个数据集的场景：Room0（Replica）、fr3（TUM-RGBD）和 bfd3fd54d2（ScanNet++）。我们的方法取得了与真实值高度吻合的渲染结果，展现出很高的准确性。更多结果请参阅补充材料。

方法	Replica				ScanNet++			
	峰值信噪比↑	结构相似性指数↑	学习感知图像块相似度↓	绝对轨迹误差均方根↓	峰值信噪比↑	结构相似性指数↑	学习感知图像块相似度↓	绝对轨迹误差均方根↓
NICE-SLAM（Zhu 等人，2022a）	24.42	0.809	0.233	1.95	-	-	-	-
Point-SLAM（Sandstr¨om 等人，2023）	5.17	0.975	0.124	0.52	-	-	-	-
SplaTAM（Keetha 等，2024年）	34.11	0.970	0.100	0.36	14.32	0.482	0.599	287.36
MonoGS（Matsuki 等人，2024年）	37.5	0.960	0.070	0.58	12.90	0.648	0.603	101.19
Photo-SLAM（Huang 等，2024年）	34.96	0.942	0.059	0.61	×	×	×	×
GS-ICP SLAM（Ha、Yeon 和 Yu，2024年）	38.83	0.975	0.041	**0.16**	14.94	0.776	0.446	111.37
Ours	**39.46**	**0.979**	**0.027**	0.19	**26.51**	**0.905**	**0.148**	**6.73**

表1：在 Replica（Straub 等，2019年）和 ScanNet++ （Yeshwanth 等，2023）数据集上，针对 RGB-D 相机对我们的方法与现有方法进行定量评估，其中“–”表示系统无法提供有效结果，“×”表示系统无法完成跟踪或重建。更多定量结果见补充材料。

在所有序列中均是如此。GS-ICP SLAM由于过度依赖深度图，在此表现不佳。PointSLAM++ 主要利用深度来增强定位和几何，而不是让深度噪声直接影响渲染，从而显著提升了性能。

在 ScanNet++ （Yeshwanth 等，2023）上，PointSLAM++ 取得了压倒性胜利，而 GS-ICP SLAM 等方法无法完成相机位姿定位。PointSLAM++ 能够实现高质量的地图重建和新视角渲染，同时保持稳定的定位能力，展现出其应对具有挑战性的真实世界扫描数据的卓越能力。

相机跟踪精度

Ta　表1展示了我们的方法在Replica数据集上的卓越跟踪精度，与其他方法相比，轨迹误差降低了50%以上。通过利用G-ICP提供的高精度深度图像和三维信息，我们优于依赖二维空间误差的方法。我们的性能与GS-ICP SLAM相当，因为高精度深度 en-

通过广义迭代最近点实现准确的相机位姿预测。

表3展示了我们在TUM-RGBD数据集上的结果；我们的方法通过渐进式位姿优化实现了出色的跟踪精度。通过引入ORB特征，我们减轻了深度噪声的干扰，增强了定位稳定性。相比之下，GS-ICP SLAM在深度噪声较大的数据上表现不佳，因为其GICP算法难以提供可靠的结果。

ScanNet++数据集的评估结果列于表1。在其他方法失效的场景中，我们的算法展现出良好的跟踪稳定性。它能够应对该数据集面临的挑战：较大的相机位姿间隔、稀疏纹理和深度干扰。在这些条件下，GICP方法会失效，而依赖二维空间误差的方法在位姿间隔较大时也难以应对；相比之下，我们的方法依然鲁棒，展现出对复杂环境的适应能力。更多结果见补充材料。

方法	指标	fr1/desk	fr2/xyz	fr3/office	Avg.
NICE-SLAM	峰值信噪比↑	13.83	17.87	12.89	14.86
	结构相似性指数↑	0.569	0.718	0.554	0.614
	学习感知图像块相似度↓	0.482	0.344	0.498	0.441
Point-SLAM	峰值信噪比↑	13.87	17.56	18.43	16.62
	结构相似性指数↑	0.627	0.708	0.754	0.696
	学习感知图像块相似度↓	0.546	0.585	0.448	0.526
SplaTAM	峰值信噪比↑	22.00	24.50	21.90	22.80
	结构相似性指数↑	0.860	**0.950**	0.880	**0.897**
	学习感知图像块相似度↓	0.230	0.100	0.200	0.177
MonoGS	峰值信噪比↑	23.69	24.68	22.83	23.7329
	结构相似性指数↑	0.786	0.793	0.770	0.783
	学习感知图像块相似度↓	0.245	0.225	0.277	0.249
Photo-SLAM	峰值信噪比↑	20.87	22.10	22.74	21.90
	结构相似性指数↑	0.743	0.765	0.780	0.763
	学习感知图像块相似度↓	0.239	0.169	0.145	0.184
GS-ICP SLAM	峰值信噪比↑	17.95	23.13	20.50	20.53
	结构相似性指数↑	0.710	0.829	0.758	0.766
	学习感知图像块相似度↓	0.296	0.141	0.232	0.223
Ours	峰值信噪比↑	**25.05**	**27.11**	**26.33**	**26.16**
	结构相似性指数↑	**0.876**	0.899	**0.881**	0.885
	学习感知图像块相似度↓	**0.123**	**0.078**	**0.118**	**0.106**

表 2：TUM-RGBD 上的渲染性能（Sturm 等人（2012））。

方法	fr1/desk	fr2/xyz	fr3/office	Avg.
NICE-SLAM	4.26	31.73	3.87	13.29
Point-SLAM	4.34	1.31	3.48	3.04
SplaTAM	3.35	1.24	5.16	3.25
MonoGS	**1.50**	1.44	1.49	1.48
Photo-SLAM	2.60	0.35	**1.00**	1.32
GS-ICP SLAM	3.50	1.76	2.74	2.67
Ours	1.56	**0.33**	1.34	**1.08**

表3：TUM-RGBD 上的跟踪性能（Sturm 等人（2012））（绝对轨迹误差均方根 [厘米] ↓）。

效率与内存比较

Ta表现比较了 PointSLAM++ 在 ScanNet++ （Yeshwanth 等，2023）上的效率和内存占用，并与现有方法进行了比较。与 SplaTAM 和 MonoGS 等神经方法相比，得益于我们的非神经跟踪优化，PointSLAM++ 能够以显著缩短的处理时间实现高质量重建。

消融研究

我们从渐进式位姿优化框架和视点相关环境补偿中分别提取了两个关键模块，并通过针对性实验评估其影响。定量结果见表 5。

渐进式位姿优化框架。我们比较了 PointSLAM++采用与不采用渐进式位姿优化（PPO）时的性能。结果显示，仅依赖广义迭代最近点算法会导致跟踪性能显著下降。在存在深度噪声的情况下，这一问题尤为突出，因为仅使用广义迭代最近点会在几何特征模糊的场景中造成累积误差

方法	建图时间↓	跟踪时间↓	FPS↑	GPU 内存（GB）↓
SplaTAM	41分54秒	117分49秒	0.053	20.32
MonoGS	29分3秒	17分29秒	0.191	23.76
GS-ICP SLAM	**1分10秒**	1分10秒	2.84	**9.62**
Ours	8分13秒	**1分00秒**	**3.33**	10.11

表4：在 ScanNet++ 上的时间效率与内存消耗对比（Yeshwanth 等，2023）。

方法	峰值信噪比↑	结构相似性指数↑	学习感知图像块相似度↓	绝对轨迹误差均方根↓
不含 PPO	21.47	0.784	0.201	17.33
不含 $NeuGS$&VDC	24.06	0.812	0.168	2.32
不含 VDC	25.36	0.856	0.128	1.10
Ours	**26.16**	**0.885**	**0.106**	**1.08**

表 5：在 TUM-RGBD（Sturm 等人（2012））上的渐进式位姿优化与神经高斯表征消融实验。

。相比之下，渐进式方法通过多阶段调整有效抑制了深度噪声。该方法提升了系统整体的鲁棒性和准确性，在相机快速运动或光照剧烈变化等具有挑战性的场景中尤为有效。

神经高斯与视角相关环境补偿。我们设置了两个消融变体：一个同时移除神经高斯（NeuGS）和视角相关补偿（VDC），另一个仅移除 VDC。结果表明，与基线相比，引入神经高斯能够提升重建质量，尤其能增强细节表征和几何一致性。尽管这一中间版本捕获了更丰富的几何与外观信息，完整的 PointSLAM++ 模型在不同视点下表现出更高的渲染一致性，并且更能适应光照条件。这证实了视角相关补偿机制对于解决视点大幅变化导致的渲染伪影至关重要；在光照变化剧烈的场景中，其改进尤为明显。此外，我们还发现，尽管各变体的跟踪组件相同，神经高斯特有的优化过程仍能生成更准确的地图。地图保真度的提升进而有助于实现更精确的位姿估计。

结论

我们提出了 PointSLAM++，这是一种基于神经高斯表示的实时 RGB-D 重建系统。通过整合渐进式位姿优化和视角相关补偿，该系统能够减轻深度噪声的影响，捕捉精细的几何结构与外观，并确保跨视角渲染的一致性。实验表明，该系统在机器人学、混合现实和智能交互领域具有出色的重建精度和鲁棒性。尽管该方法效果显著，但会增加计算成本，因此未来工作将致力于提高效率，以支持更广泛的设备部署。

致谢

本研究得到国家自然科学基金（编号 U25A20421、编号 62202151、编号 62202152）和中国国家重点研发计划（编号 2025YFB3003601）。

参考文献

Bylow, E.; Sturm, J.; Kerl, C.; Kahl, F.; and Cremers, D. 2013. 使用有符号距离函数进行实时相机跟踪和三维重建。在*Robotics: Science and sys-tems (RSS) conference 2013*，第 9 卷。Robotics: Science and Systems。

Campos, C.; Elvira, R.; Rodr´ıguez, J. J. G.; Montiel, J. M.; and Tard´os, J. D. 2021. Orb-slam3：一个用于视觉、视觉-惯性及多地图同步定位与建图的高精度开源库。*IEEE*机器人学汇刊，37(6)：1874–1890。Canelhas, D. R.; Stoyanov, T.; and Lilienthal, A. J. 2013. SDF tracker：一种用于基于深度图像在线位姿估计和场景重建的并行算法。载于*2013*年*IEEE/RSJ*智能机器人与系统国际会议，3671–3676。IEEE。Chung, C.-M.; Tseng, Y.-C.; Hsu, Y.-C.; Shi, X.-Q.; Hua, Y.-H.; Yeh, J.-F.; Chen, W.-C.; Chen, Y.-T.; and Hsu, W. H. 2023. Orbeez-slam：一种采用ORB特征和神经辐射场实现的建图的实时单目视觉同步定位与建图方法。载于*2023*年*IEEE*机器人与自动化国际会议（*ICRA*），9400–9406。IEEE。Dai, A.; Nießner, M.; Zollh¨ofer, M.; Izadi, S.; and Theobalt, C. 2017. Bundlefusion：通过即时表面重新整合实现实时、全局一致的三维重建。*ACM*图形学汇刊（*ToG*），36(4)：1。Deng, T.; Shen, G.; Qin, T.; Wang, J.; Zhao, W.; Wang, J.; Wang, D.; and Chen, W. 2024年。Plgslam：通过从局部到全局的光束法平差实现渐进式神经场景表示。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，19657–19666。Du, H.; Henry, P.; Ren, X.; Cheng, M.; Goldman, D. B.; Seitz, S. M.; and Fox, D. 2011. 使用消费级深度相机对室内环境进行交互式三维建模。载于第*13*届普适计算国际会议论文集，75–84。Gelfand, N.; Ikemoto, L.; Rusinkiewicz, S.; and Levoy, M. 2003. 用于迭代最近点算法的几何稳定采样。载于第四届三维数字成像与建模国际会议（*2003*年）。*3DIM 2003*。会议论文集。，260–267。IEEE。Ha, S.; Yeon, J.; and Yu, H. 2024年。RGB-D GS-ICP同步定位与建图。载于欧洲计算机视觉会议，180–197。施普林格。Hu, J.; Chen, X.; Feng, B.; Li, G.; Yang, L.; Bao, H.; Zhang, G.; and Cui, Z. 2024年。CG-SLAM：在一致的不确定性感知三维高斯场中实现高效稠密RGB-D同步定位与建图。载于欧洲计算机视觉会议，93–112。施普林格。Huang, H.; Li, L.; Hui, C.; and Yeung, S.-K. 2024年。Photo-SLAM：面向单目、立体和RGB-D相机的实时同步定位与照片级真实感建图。载于*IEEE/CVF*计算机视觉与模式识别会议论文集。

Jiang, Y.; Yu, C.; Xie, T.; Li, X.; Feng, Y.; Wang, H.; Li, M.; Lau, H.; Gao, F.; Yang, Y.; et al. 2024年。Vr-gs：一种用于头戴式显示器交互式逼真人-场景交互的物理动力学框架。载于*SIGGRAPH Conference Papers '24*，1–1。Jin, H.; Li, I.; Qin, T.; and Shen, S. 2024年。基于有符号距离场的高效稠密同步定位与建图系统。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，17408–17419。

Keetha, N.; Karhade, J.; Jatavallabhula, K. M.; Yang, G.; Scherer, S.; Ramanan, D.; 以及 Luiten, J. 2024年。Splatam：用于稠密RGB-D同步定位与建图的 Splat 跟踪与三维高斯建图。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，21357–21366。

Keller, M.; Lefloch, D.; Lambers, M.; Izadi, S.; Weyrich, T.; 和 Kolb, A. 2013. 在动态场景中使用基于点的融合进行实时三维重建。载于*2013*年第三维视觉国际会议（*3DV 2013*），1–8。IEEE。Kerbl, B.; Kopanas, G.; Leimk¨uhler, T.; 和 Drettakis, G. 2023. 用于实时辐射场渲染的3D高斯泼溅。《ACM计算机图形学汇刊》，42(4)：139–1。Liso, L.; Sandstr¨om, E.; Yugay, V.; Van Gool, L.; 和 Oswald, M. R. 2024. Loopy-slam：带闭环检测的稠密神经SLAM。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，20363–20373。Lu, T.; Yu, M.; Xu, L.; Xiangli, Y.; Wang, L.; Lin, D.; 和 Dai, B. 2024. Scaffold-gs：用于视角自适应渲染的结构化三维高斯。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，20654–20664。Mao, Y.; Yu, X.; Zhang, Z.; Wang, K.; Wang, Y.; Xiong, R.; 和 Liao, Y. 2024. Ngel-slam：基于神经隐式表示的全局一致低延迟SLAM系统。载于*2024*年*IEEE*机器人与自动化国际会议（*ICRA*），6952–6958。IEEE。Matsuki, H.; Murai, R.; Kelly, P. H.; 和 Davison, A. J. 2024. 高斯泼溅SLAM。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，18039–18048。Mildenhall, B.; Srinivasan, P. P.; Tancik, M.; Barron, J. T.; Ramamoorthi, R.; 和 Ng, R. 2020. 神经辐射场：将场景表示为神经辐射场以进行视图合成。载于欧洲计算机视觉会议（*ECCV*），405–421。口头报告。Newcombe, R. A.; Izadi, S.; Hilliges, O.; Molyneaux, D.; Kim, D.; Davison, A. J.; Kohi, P.; Shotton, J.; Hodges, S.; 和 Fitzgibbon, A. 2011. Kinectfusion：实时稠密表面建图与跟踪。载于*2011*年第十届*IEEE*混合与增强现实国际研讨会，127–136。IEEE。Prisacariu, V. A.; K¨ahler, O.; Golodetz, S.; Sapienza, M.; Cavallari, T.; Torr, P. H. S.; 和 Murray, D. W. 2017. InfiniTAM v3：用于大规模三维重建与闭环检测的框架。arXiv:1708.00783。Sandstr¨om, E.; Li, Y.; Van Gool, L.; 和 R. Oswald, M. 2023. Point-SLAM：基于稠密神经点云的

SLAM。载于*IEEE/CVF*国际计算机视觉大会（*ICCV*）会议论文集。Straub, J.; Whelan, T.; Ma, L.; Chen, Y.; Wijmans, E.; Green, S.; Engel, J. J.; Mur-Artal, R.; Ren, C.; Verma, S.; Clarkson, A.; Yan, M.; Budge, B.; Yan, Y.; Pan, X.; Yon, J.; Zou, Y.; Leon, K.; Carter, N.; Briales, J.; Gillingham, T.; Mueggler, E.; Pesqueira, L.; Savva, M.; Batra, D.; Strasdat, H. M.; Nardi, R. D.; Goesele, M.; Lovegrove, S.; 和 Newcombe, R. 2019. Replica数据集：室内空间的数字副本。《arXiv预印本 arXiv:1906.05797》。Sturm, J.; Engelhard, N.; Endres, F.; Burgard, W.; 和 Cremers, D. 2012. RGB-D同步定位与建图系统的评测基准。载于国际智能机器人系统大会（*IROS*）论文集。Sucar, E.; Liu, S.; Ortiz, J.; 和 Davison, A. J. 2021. iMAP：实时隐式建图与定位。载于*IEEE/CVF*国际计算机视觉大会论文集，6229–6238。Wang, H.; Wang, J.; 和 Agapito, L. 2023. Co-slam：用于神经实时SLAM的联合坐标与稀疏参数化编码。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，13293–13302。Wang, Y.; Skorokhodov, I.; 和 Wonka, P. 2022. Hf-neus：利用高频细节改进表面重建。《神经信息处理系统进展》，35：1966–1978。Whelan, T.; Johannsson, H.; Kaess, M.; Leonard, J. J.; 和 McDonald, J. 2013. 用于稠密RGB-D建图的鲁棒实时视觉里程计。载于*2013*年*IEEE*机器人与自动化国际会议，5724–5731。IEEE。Whelan, T.; Salas-Moreno, R. F.; Glocker, B.; Davison, A. J.; 和 Leutenegger, S. 2016. ElasticFusion：实时稠密SLAM与光源估计。《国际机器人研究期刊》，35(14)：1697–1716。Yan, C.; Qu, D.; Xu, D.; Zhao, B.; Wang, Z.; Wang, D.; 和 Li, X. 2024. Gs-slam：基于3D高斯泼溅的稠密视觉SLAM。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，19595–19604。Yeshwanth, C.; Liu, Y.-C.; Nießner, M.; 和 Dai, A. 2023. ScanNet：高保真3D室内场景数据集。载于*IEEE/CVF*国际计算机视觉大会论文集，12–22。Zhang, W.; Sun, T.; Wang, S.; Cheng, Q.; 和 Haala, N. 2023. Hi-slam：利用混合隐式场进行单目实时稠密建图。《IEEE机器人与自动化快报》，9(2)：1548–1555。Zhu, L.; Li, Y.; Sandstr¨om, E.; Huang, S.; Schindler, K.; 和 Armeni, I. 2024a. Loopsplat：通过配准3D高斯泼溅实现闭环检测。《arXiv预印本 arXiv:2408.10154》。Zhu, S.; Wang, G.; Blum, H.; Liu, J.; Song, L.; Pollefeys, M.; 和 Wang, H. 2024b. Sni-slam：语义神经隐式SLAM。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，21167–21177。Zhu, Z.; Peng, S.; Larsson, V.; Xu, W.; Bao, H.; Cui, Z.; Oswald, M. R.; 和 Pollefeys, M. 2022a. Nice-slam：用于SLAM的神经隐式可扩展编码。载于会议论文集

*IEEE/CVF*计算机视觉与模式识别会议，12786–12796。Zhu, Z.; Peng, S.; Larsson, V.; Xu, W.; Bao, H.; Cui, Z.; Os- wald, M. R.; and Pollefeys, M. 2022b. NICE-SLAM：用于同步定位与建图的神经隐式可扩展编码。载于*IEEE/CVF*计算机视觉与模式识别会议论文集 *(CVPR)*。Zubizarreta, J.; Aguinaga, I.; and Montiel, J. M. M. 2020. 直接稀疏建图。*IEEE*机器人学汇刊，36(4): 1363–1370。