

内视流 SLAM: 带有流动约束的高斯点云实时内窥镜 SLAM

Taoyu Wu^{1,4} *^[0009-0008-7991-6869] *, Yiyi Miao^{2,3} *^[0009-0008-4488-1272],
Zhuoxiao Li¹^[0000-0002-4531-1959], Haocheng Zhao¹^[0000-00018932-8160], Kang
Dang²^[0000-0003-0613-2787], Jionglong Su²^[0000-0001-5360-6493], Limin
Yu¹ **^[0000-0002-6891-0604], and Haoang Li⁵ **

¹ School of Advanced Technology, Xi'an Jiaotong Liverpool University, Suzhou, China

Taoyu.Wu21@student.xjtlu.edu.cn, Limin.yu@xjtlu.edu.cn

² School of AI and Advanced Computing, Xi'an Jiaotong Liverpool University, Suzhou, China

³ School of Electrical Engineering, Electronics and Computer Science, University of Liverpool, Liverpool, United Kingdom

⁴ School of Physical Sciences, University of Liverpool, Liverpool, United Kingdom

⁵ The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China.

Haoangli@hkust-gz.edu.cn

摘要 高效的三维重建和实时可视化在如内窥镜手术等场景中至关重要。近年来, 3D 高斯点绘制 (3DGS) 在高效 3D 重建和渲染方面表现出色。大多数基于 3DGS 的同时定位与地图构建 (SLAM) 方法仅依赖外观约束来优化 3DGS 和相机姿态。然而, 在内窥镜场景中, 非朗伯表面引起的光度不一致以及呼吸造成的动态运动会影响 SLAM 系统的性能。为解决这些问题, 我们额外引入了光流损失作为几何约束, 这有效限制了场景的三维结构和相机运动。此外, 我们提出了一种深度正则化策略来缓解光度不一致的问题, 并确保 3DGS 在内窥镜场景中的深度渲染有效性。另外, 为了改进 SLAM 系统中的场景表示, 我们通过专注于与关键帧对应且具有次优渲染质量帧的视点, 改进了 3DGS 细化策略, 从而获得更好的渲染结果。在 C3VD 静态数据集和 StereoMIS 动态数据集上的广泛实验表明, 我们的方法在新视角合成和姿态估计方面优于现有的最先进的方法, 在静止和动态手术场景中表现出高性能。源代码将在论文被接受后公开。

* Co-first authors.

** Corresponding author.

Keywords: 内镜手术 · 新视角合成 · 3D 高斯投点法

1 介绍

精确的相机姿态估计和器官组织重建在医学应用中至关重要，尤其是在常用的内窥镜微创手术中 [17]。然而，有限的视野、动态组织变形和复杂的光照条件等挑战可能会削弱这些程序的精度 [18]。精确的相机姿态估计对于定义器官与手术器械之间的空间关系至关重要，这有助于外科医生导航和工具操作 [3,13,26]。密集场景重建也在手术期间的组织分析中发挥作用，并支持术后评估。视觉同步定位与地图构建 (SLAM) 技术为解决这些挑战提供了有希望的解决方案。

传统视觉 SLAM 方法通常生成稀疏场景重建，并假设环境是刚性的，这在内窥镜场景中往往不成立 [2,5]。这样的环境中纹理较弱，进一步复杂了基于特征的匹配。因此，将这些方法应用于内窥镜场景在准确性和鲁棒性方面存在显著挑战。密集 SLAM 系统已开发用于实时密集重建，它们通常依赖于 RGB-D 序列。然而，由于外科器械和切口的限制，在内窥镜环境中难以捕捉这些序列，导致场景表示不完整，受限内窥镜有限移动范围而被遮挡的部分。

为了克服这些限制，最近的研究将神经辐射场 (NeRF) [20] 和 3D 高斯点阵 (3DGS) [11] 集成到了 SLAM 系统中。NeRF 利用带有体积渲染的神经网络，实现了高精度的新视角合成，并填充地图中的未观测区域 [28]。基于 NeRF 的 SLAM 系统，如 ENeRF-SLAM [23] 和 Endo-Depth-and-Motion [22]，专门开发用于解决内窥镜环境中动态变化和空间约束问题。然而，NeRF 系统的训练和推理的高计算成本可能限制了它们在实时手术中的实际应用。相比之下，3DGS 在保持照片真实感图像的同时提高了渲染效率。通过结合显式场景表示与可微分高斯光栅化，3DGS 相比 NeRF [20] 实现了显著更高的渲染速度。基于 3DGS 的 SLAM，如 EndoGSLAM [25] 和 Free-SurGS [6]，已经展示了令人满意的相机跟踪和密集场景重建。

尽管取得了这些进展，上述挑战在将这些方法应用于实际内窥镜场景时仍然存在，特别是在处理由非朗伯表面引起的光度不一致、呼吸引起的动力变形以及缺乏深度信息的问题上。为了解决这些问题，我们提出了一种基于 3DGS 的内窥镜 SLAM 方法 **末端流-SLAM**，该方法能够在内窥镜环境中的高效相机跟踪、高质量密集重建和高保真新视角图像合成。我们的贡献可以总结如下三点：首先，我们将光流损失作为几何约束的一部分引入。这

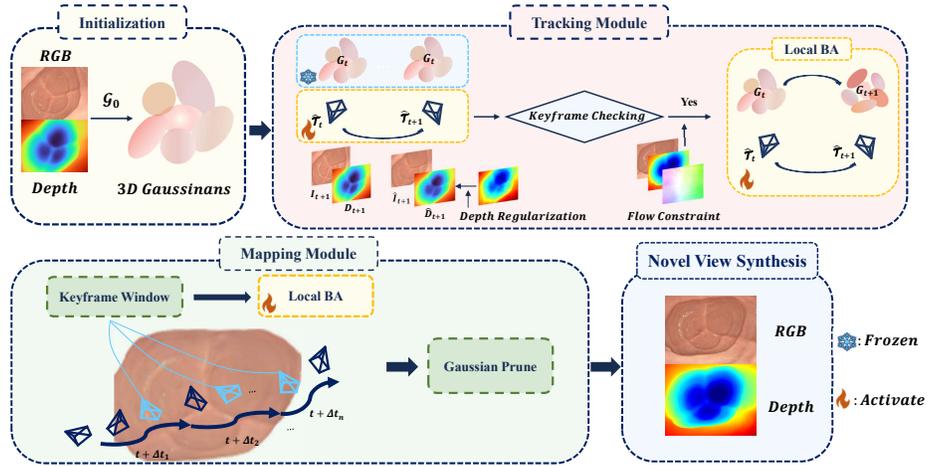


图 1. 方法概述。给定第一张 RGB-D 图像，我们初始化 3D 高斯分布，并随后迭代执行相机跟踪和建图。跟踪处理非关键帧并使用深度正则化以保持尺度一致性（仅优化相机姿态），而关键帧将光流作为几何约束纳入局部 BA 中同时优化姿态和 3DGS 原始模型。建图应用带有光流约束的 BA 来优化关键帧窗口内的姿态和 3DGS 原始模型。

一损失有效地限制了场景 3D 结构和相机运动的变化。使用我们基于 CUDA 的微分高斯栅格化技术，我们可以同时高效地优化 3DGS 和相机姿态。其次，为了解决由单目深度估计引入的比例模糊问题，我们提出了一种结合深度图归一化与比例不变损失公式的方法。第三，为了增强场景表示，我们在 SLAM 系统中的实时渲染后应用了改进的 3DGS 细化策略，在此过程中通过优先考虑关键帧视点实现了更准确的深度和彩色图像渲染结果。

2 方法论

2.1 初步：3D 高斯投点法

3DGS [11] 是一个可微渲染框架，使用一组高斯原语来建模三维场景。第 i 个高斯由中心 μ_i 、不透明度 o_i 和协方差矩阵 Σ_i 表征，并可以表示为：

$$G_i(\mathbf{X}) = o_i \cdot \exp \left\{ -\frac{1}{2} (\mathbf{X} - \mu_i)^\top \Sigma_i^{-1} (\mathbf{X} - \mu_i) \right\}, \quad (1)$$

其中 \mathbf{X} 表示三维空间中的任意一点。在渲染过程中，这些高斯分布沿着相机光线投影到二维图像平面上，并使用 alpha 混合方案合成它们的贡献。

如图 1 所示，基于 3DGS 的系统由三个主要模块组成：初始化模块、相机跟踪模块和映射模块。在初始化阶段，参照主流的基于 3DGS 的 SLAM 方法 [19,10,14]，我们从帧 I_{t_0} 初始化各向同性 3DGS \mathcal{G}_0 ，其中各向同性 \mathcal{G}_0 更适合内窥镜场景。在相机跟踪模块中，将相机姿态估计公式化为最小化与外观和几何约束相对应的光度和流残差。我们假设时间 t 处的相机姿态 \mathcal{T}_t 已经优化。我们的目标是优化时间 $t+1$ 处的估计相机姿态 $\hat{\mathcal{T}}_{t+1}$ 。在映射模块中，对于从关键帧窗口中的跟踪模块获得的粗略相机姿态，我们使用束调整 (BA) 联合优化关键帧窗口内的相机姿态和 3D 场景表示。

2.2 非关键帧优化

现有的基于 3DGS 的 SLAM 系统 [25,16,19] 通常采用光度 (\mathcal{L}_{rgb}) 和深度 ($\mathcal{L}_{\text{depth}}$) 误差最小化来进行相机姿态估计。然而，由于单目深度估计的限制，在内窥镜场景中这种范式面临着尺度模糊挑战 [27,15]。为了解决这一限制，我们提出了一种包含尺度不变损失 $\mathcal{L}_{\text{scale}}$ [21] 和深度梯度正则化 $\mathcal{L}_{\text{depth}}^{\text{reg}}$ 的修改优化框架。深度正则化项定义为估计深度图和真实深度图之间的加权梯度差： $\mathcal{L}_{\text{depth}}^{\text{reg}} = \frac{1}{N} \sum_{i=1}^N (w_h \cdot |\nabla_h d_i| + w_v \cdot |\nabla_v d_i|)$ ，其中 d_i 表示像素 i 的深度值， ∇_h 和 ∇_v 分别表示水平和垂直梯度， w_h 和 w_v 是相应的权重因子， N 表示图像像素的总数。估计的相机姿态 $\hat{\mathcal{T}}_{t+1}$ 在时间 $t+1$ 可以通过最小化光度、尺度不变和深度正则化损失来进行优化：

$$\hat{\mathcal{T}}_{t+1} = \underset{\mathcal{T}_{t+1}}{\operatorname{argmin}} \lambda_1 \mathcal{L}_{\text{rgb}} + \lambda_2 \mathcal{L}_{\text{depth}}^{\text{reg}} + \lambda_3 \mathcal{L}_{\text{scale}}, \quad (2)$$

2.3 关键帧优化与流约束

为了增强动态内窥镜场景中的几何一致性，我们在时间 $t+1$ 处将流损失作为关键帧优化的附加约束条件，从而实现同时优化的场景表示和姿态估计。

佩-高斯像素流。如图 2 所示，每个像素 \mathbf{x}_t 对应一组 3DGS，其中像素颜色是通过混合从多个 3D 高斯分布投影出的 2D 高斯分布得到的。基于 [4] 中的工作，在时间 t ，我们将第 i 个三维高斯分布使用相机姿态 \mathcal{T}_t 投影到二维图像平面上，结果得到像素 $\mathbf{x}_{i,t}$ 。该像素使用相应的第 i 个二维高斯分布的均值 $\boldsymbol{\mu}_{i,t}$ 和协方差矩阵 $\boldsymbol{\Sigma}_{i,t}$ 映射到规范空间。在时间 $t+1$ ，像素位置 $\mathbf{x}_{i,t+1}$ 是通过将 3DGS 投影到未知但需要求解的相机姿态 \mathcal{T}_{t+1} 来确定的，表示为：

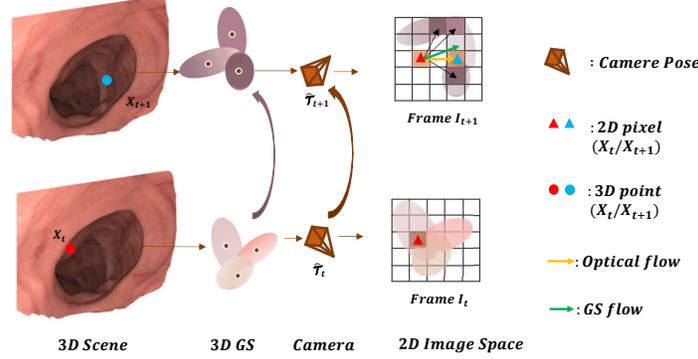


图2. 高斯流估计。在时间 t ，每个像素 x_t 都是由 K 个重叠的高斯分布产生的。在时间 $t+1$ ，每个 K 高斯分布将对应一个高斯流（黑色箭头）。通过累积这些高斯流，我们获得总体高斯流。我们的目标是通过优化相机姿态 $\hat{\mathcal{T}}_{t+1}$ 和三维 GS 原语 $\hat{\mathcal{G}}$ 来最小化 GS 流与光流之间的差异。

$\mathbf{x}_{i,t+1} = \pi(\mathcal{G}_t, \mathcal{T}_{t+1})$ ，其中 $\pi(\cdot)$ 表示透视投影。从这，我们可以得到对应的均值 $\boldsymbol{\mu}_{i,t+1}$ 和协方差矩阵 $\boldsymbol{\Sigma}_{i,t+1}$ 对于第 i 个高斯分布。第 i 个高斯分布的高斯流由位置偏移给出，表示像素位置之间的差异： $\text{flow}_i^G(\mathbf{x}_t) = \mathbf{x}_{i,t+1} - \mathbf{x}_{i,t}$ 。

通过流约束同时优化。不同于 [4]，在我们的 SLAM 系统中，高斯分布是各向同性的，其中两个协方差矩阵都是对称且正定的。因此，协方差矩阵 $\boldsymbol{\Sigma}_{i,t}$ 和 $\boldsymbol{\Sigma}_{i,t+1}$ 的 Cholesky 分解 [8] 简化为单位矩阵。这使我们能够表达第 i 个高斯等效的高斯流为： $\mathbf{f}_i^G(\mathbf{x}_t) = \boldsymbol{\mu}_{i,t+1} - \boldsymbol{\mu}_{i,t}$ 。对于每个具有 K 个重叠高斯像素，我们通过 alpha 加权混合计算复合流： $\mathbf{f}^G(\mathcal{T}_{t+1}, \mathcal{G}_t) = \sum_{i=1}^K w_i (\boldsymbol{\mu}_{i,t+1} - \boldsymbol{\mu}_{i,t})$ ，其中 w_i 表示沿摄像机光线第 i 个高斯的归一化混合权重。对于相邻帧 I_t 和 I_{t+1} ，我们使用现成的方法获得光流 $\mathbf{f}^{Gt}(\mathbf{x})$ 作为真实值。然后，我们将所有像素上的流损失定义为： $\mathcal{L}_{\text{flow}} = \|\mathbf{f}^G(\mathcal{T}_{t+1}, \mathcal{G}_t) - \mathbf{f}^{Gt}(\mathbf{x})\|_2$ 。

对于时间 $t+1$ 的关键帧优化，我们同时通过最小化以下目标函数来优化估计的相机姿态 $\hat{\mathcal{T}}_{t+1}$ 和 3DGS 原语 $\hat{\mathcal{G}}$ ：

$$\hat{\mathcal{T}}_{t+1}, \hat{\mathcal{G}} = \underset{\mathcal{T}_{t+1}, \mathcal{G}}{\operatorname{argmin}} (\lambda_1 \mathcal{L}_{\text{rgb}} + \lambda_2 \mathcal{L}_{\text{depth}}^{\text{reg}} + \lambda_3 \mathcal{L}_{\text{scale}} + \lambda_4 \mathcal{L}_{\text{flow}}). \quad (3)$$

相比于方程 (2)，方程 (3) 还额外涉及流动损失 $\mathcal{L}_{\text{flow}}$ 作为几何约束， $\mathcal{L}_{\text{flow}}$ 维持了运动的几何一致性。

2.4 关键帧优化和全局细化

关键帧导向的局部束调整。为了进一步提高场景的表达能力，我们还在映射中引入流损失作为几何约束。在相机跟踪模块之后，只有当帧 I_{t+1} 被指定为关键帧时，我们才执行 3D 高斯图表示优化。我们遵循来自 [19] 的关键帧管理策略。我们在关键帧窗口内的所有帧上执行 BA。通过最小化目标函数 (3) 中的损失函数 \mathcal{L}_{rgb} 、 $\mathcal{L}_{\text{depth}}$ 和 $\mathcal{L}_{\text{flow}}$ ，我们同时优化了估计的相机姿态 \hat{T}_{t+1} 和 3DGS 原始数据 $\hat{\mathcal{G}}$ 。

全局细化。

MonoGS [19] 在单阶段优化方法中存在局限性，其中随机视点选择无法有效细化低质量区域，并且仅依赖于光度损失会削弱几何约束。与 [19] 不同，在完成 SLAM 系统中所有帧的 3D 高斯图表示和相机跟踪后，我们实现了一个两阶段的修订全局优化策略。在第一阶段，我们优先使用关键帧选择视点，并额外涉及深度损失 $\mathcal{L}_{\text{depth}}^{\text{reg}}$ 和尺度不变性损失 $\mathcal{L}_{\text{scale}}$ 来优化全局场景，通过这个细化阶段提升场景质量。在第二阶段，我们采用全球视点选择策略，随机选取视点以确保在整个场景中进行全面的细化。

对于每个选定的视角点，使用以下损失函数进行高斯基元的优化：

$$\mathcal{L}_{\text{refine}} = (1 - \lambda_{\text{dssim}})\mathcal{L}_{\text{rgb}} + \lambda_{\text{dssim}}(1 - \text{SSIM}) + \|\nabla\hat{D} - \nabla D\|_2^2, \quad (4)$$

其中 λ_{dssim} 控制结构相似性指数 (SSIM) 的相对权重。SSIM 项确保了结构的保存，保持场景整体结构的完整性的同时最小化像素级差异。

3 实验

3.1 实现细节

实验设置。所有实验均在配备固定超参数的 NVIDIA RTX 4090 GPU 上进行。我们采用了 Adam 优化器 [12] 用于相机姿态参数，并使用标准 3DGS 学习率 [11] 用于高斯属性。参照 [19]，通过省略表示视点相关辐射的球谐函数 (SHs) 来简化高斯表示以求简便。每帧我们均使用了 15 次迭代进行相机跟踪和映射。

数据集。评估是在两个具有临床相关性的数据集上进行的，这些数据集捕捉了不同的内镜挑战。C3VD 数据集 [1] 提供带真实姿态的 RGB-D 数据，结合实际临床视频与合成结肠模型用于验证。参照 [25]，我们在 10 个序列（分辨

率 675 (E 540) 上进行了评估, 这些序列展示了典型的内镜挑战, 例如非朗伯表面。对于动态场景的评估, 我们使用了由达芬奇 Xi 手术机器人捕捉的 StereoMIS 数据集 [7], 该数据集包含因呼吸导致的自然组织变形特征。参照 [24], 我们选择了 P2_1 序列 (分辨率 640 (E 512)), 该序列包含了短段 (200 帧) 和长段 (1000 帧) 部分, 用于评估时间一致性。

评估指标。对于几何评估, 我们使用深度上的以毫米 (mm) 为单位的均方根误差 (RMSE)。相机跟踪使用绝对轨迹误差 (ATE), 单位为毫米 (mm) 进行评估。我们还使用峰值信噪比 (PSNR)、学习感知图像块相似性 (LPIPS) 和结构相似度指数测量 (SSIM) 指标来评估我们的新视图合成的渲染质量。

3.2 定量和定性结果

我们将我们的方法与四种代表性的 SLAM 方法进行了比较: NICE-SLAM [28]、Endo-Depth [22]、ESLAM [9] 和 EndoGSLAM [25]。我们的评估主要集中在新视角渲染和相机定位性能上, 使用了 C3VD 和 StereoMIS 数据集。我们在表 1 中总结了每个指标在十个场景中的平均值。我们的方法在 PSNR、SSIM、RMSE 和 ATE 方面始终优于其他方法。特别是与基于 NeRF 的 NICE-SLAM 相比, 我们基于 3DGS 的方法提供了更快的渲染和重建速度, 从而保证实时性能。此外, 如图 3 所示, 我们的方法在渲染质量上优于现有的基于 3DGS 的 SLAM 系统。另外, 我们在涉及呼吸内窥镜场景的 StereoMIS 数据集 [7] 上进行了实验。表 2 所示的结果表明, 我们的方法优于 NICE-SLAM、ESLAM 和 EndoGSLAM。如图 3 所示, EndoGSLAM 缺乏几何约束, 使其无法处理由呼吸引起的动态变化。因此, 整体图像显得模糊, 纹理细节明显缺失且形状不正确。

表 1. C3VD 数据集上的定量结果。

Methods	峰值信噪比 \uparrow	结构相似性指标 \uparrow	LPIPS \downarrow	均方根误差 (毫米) \downarrow	ATE (毫米) \downarrow
NICE-SLAM [28]	22.07	0.73	0.33	1.88	0.48
Endo-Depth [22]	18.13	0.64	0.33	5.10	1.25
EndoGSLAM-H [25]	22.16	0.77	0.22	2.17	0.34
我们的	25.18	0.82	0.27	1.54	0.23

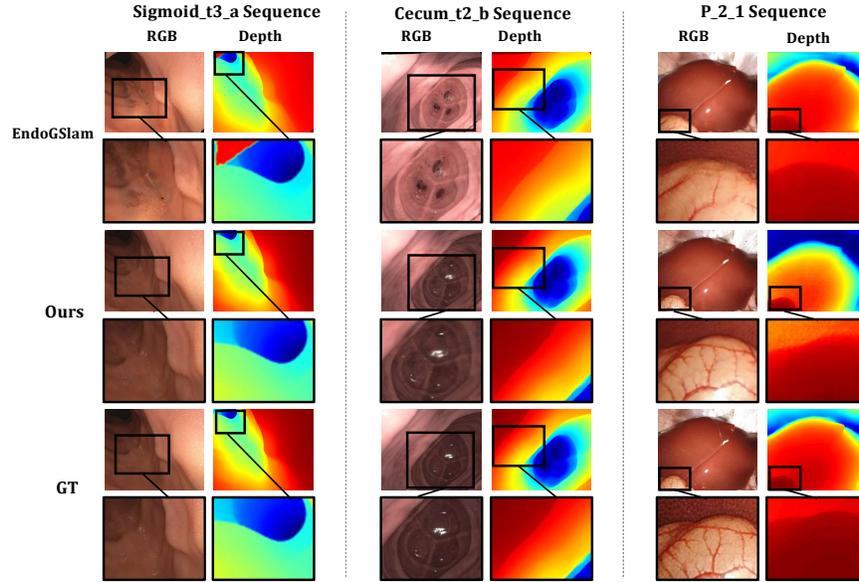


图 3. C3VD 和 StereoMIS 数据集的定性结果。

表 2. StereoMIS 数据集上的定量结果。

Methods	峰值信噪比 \uparrow	结构相似性指标 \uparrow	LPIPS \downarrow	ATE (毫米) \downarrow
NICESLAM [28]	13.07	0.49	0.61	38.24
ESLAM [9]	18.70	0.54	0.57	16.73
EndoGSLAM-H [25]	16.67	0.52	0.45	18.82
我们的	21.96	0.59	0.27	15.47

3.3 消融研究

表 3 展示了深度约束、优化策略和流约束模块的消融研究。结果显示，深度损失函数对于准确的深度重建至关重要，其缺失会导致深度图精度下降。集成外观和几何约束的优化模块提升了渲染的深度和颜色图像的质量，而不会影响相机跟踪精度。依赖于两种约束并用于相机跟踪的流模块同时提高了跟踪精度和渲染质量。

4 结论

在本文中，我们介绍了基于 3DGS 的 EndoFLow-SLAM 框架。该框架能够在内窥镜场景中实现精确的相机跟踪和高质量的新视角合成。通过将光

表 3. 我们方法在 C3VD 数据集上的消融研究

方法	峰值信噪比 \uparrow	SSIM \uparrow	LPIPS \downarrow	均方根误差 (毫米) \downarrow	自动三坐标测量机误差 (毫米) \downarrow
w.o. Depth	22.96	0.80	0.40	7.47	0.44
w.o. Refine	21.26	0.78	0.42	3.35	0.23
w.o. Flow	21.04	0.75	0.43	2.63	0.48
Ours	25.18	0.82	0.27	2.04	0.23

流作为几何约束纳入其中, EndoFlow-SLAM 更能处理由呼吸引起的现实世界内窥镜环境中动态变化。广泛的实验结果表明, 与传统的 SLAM 方法以及基于 3DGS 的方法相比, EndoFlow-SLAM 实现了更优的跟踪和渲染性能。未来工作将专注于对象级 3DGS 的构建和优化, 以更好地适应更复杂内窥镜场景中发现的强烈动态变化。

参考文献

1. Bobrow, T.L., Golhar, M., Vijayan, R., Akshintala, V.S., Garcia, J.R., Durr, N.J.: Colonoscopy 3d video dataset with paired depth from 2d-3d registration. *Medical Image Analysis* p. 102956 (2023)
2. Campos, C., Elvira, R., Rodríguez, J.J.G., Montiel, J.M., Tardós, J.D.: Orb-slam3: An accurate open-source library for visual, visual-inertial, and multimap slam. *IEEE Transactions on Robotics* **37**(6), 1874–1890 (2021)
3. Chen, L., Tang, W., John, N.W., Wan, T.R., Zhang, J.J.: Slam-based dense surface reconstruction in monocular minimally invasive surgery and its application to augmented reality. *Computer methods and programs in biomedicine* **158**, 135–146 (2018)
4. Gao, Q., Xu, Q., Cao, Z., Mildenhall, B., Ma, W., Chen, L., Tang, D., Neumann, U.: Gaussianflow: Splatting gaussian dynamics for 4d content creation. *arXiv preprint arXiv:2403.12365* (2024)
5. Grasa, O.G., Bernal, E., Casado, S., Gil, I., Montiel, J.: Visual slam for hand-held monocular endoscope. *IEEE transactions on medical imaging* **33**(1), 135–146 (2013)
6. Guo, J., Wang, J., Kang, D., Dong, W., Wang, W., Liu, Y.h.: Free-surgs: Sfm-free 3d gaussian splatting for surgical scene reconstruction. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 350–360. Springer (2024)

7. Hayoz, M., Hahne, C., Gallardo, M., Candinas, D., Kurmann, T., Allan, M., Sznitman, R.: Learning how to robustly estimate camera pose in endoscopic videos. *International journal of computer assisted radiology and surgery* **18**(7), 1185–1192 (2023)
8. Higham, N.J.: Cholesky factorization. *Wiley interdisciplinary reviews: computational statistics* **1**(2), 251–254 (2009)
9. Johari, M.M., Carta, C., Fleuret, F.: Eslam: Efficient dense slam system based on hybrid representation of signed distance fields. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 17408–17419 (2023)
10. Keetha, N., Karhade, J., Jatavallabhula, K.M., Yang, G., Scherer, S., Ramanan, D., Luiten, J.: Splatam: Splat track and map 3d gaussians for dense rgb-d slam. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 21357–21366 (2024)
11. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.* **42**(4), 139–1 (2023)
12. Kingma, D.P.: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014)
13. Lamarca, J., Rodríguez, J.J.G., Tardós, J.D., Montiel, J.M.: Direct and sparse deformable tracking. *IEEE Robotics and Automation Letters* **7**(4), 11450–11457 (2022)
14. Li, H., Meng, X., Zuo, X., Liu, Z., Wang, H., Cremers, D.: Pg-slam: Photo-realistic and geometry-aware rgb-d slam in dynamic environments. *arXiv preprint arXiv:2411.15800* (2024)
15. Li, W., Chen, W., Qian, S., Chen, J., Cremers, D., Li, H.: Dynsup: Dynamic gaussian splatting from an unposed image pair. *arXiv preprint arXiv:2412.00851* (2024)
16. Li, Z., Yao, S., Wu, T., Yue, Y., Zhao, W., Qin, R., Garcia-Fernandez, A.F., Levers, A., Zhu, X.: Ulsr-gs: Ultra large-scale surface reconstruction gaussian splatting with multi-view geometric consistency. *arXiv preprint arXiv:2412.01402* (2024)
17. Liu, X., Li, Z., Ishii, M., Hager, G.D., Taylor, R.H., Unberath, M.: Sage: slam with appearance and geometry prior for endoscopy. In: *2022 International conference on robotics and automation (ICRA)*. pp. 5587–5593. IEEE (2022)
18. Marcus, H.J., Cundy, T.P., Hughes-Hallett, A., Yang, G.Z., Darzi, A., Nandi, D.: Endoscopic and keyhole endoscope-assisted neurosurgical approaches: a qualitative survey on technical challenges and technological solutions. *British journal of neurosurgery* **28**(5), 606–610 (2014)

19. Matsuki, H., Murai, R., Kelly, P.H., Davison, A.J.: Gaussian splatting slam. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18039–18048 (2024)
20. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM **65**(1), 99–106 (2021)
21. Ranftl, R., Lasinger, K., Hafner, D., Schindler, K., Koltun, V.: Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. IEEE transactions on pattern analysis and machine intelligence **44**(3), 1623–1637 (2020)
22. Recasens, D., Lamarca, J., Fácil, J.M., Montiel, J., Civera, J.: Endo-depth-and-motion: Reconstruction and tracking in endoscopic videos using depth networks and photometric constraints. IEEE Robotics and Automation Letters **6**(4), 7225–7232 (2021)
23. Shan, J., Li, Y., Xie, T., Wang, H.: Enerf-slam: a dense endoscopic slam with neural implicit representation. IEEE Transactions on Medical Robotics and Bionics (2024)
24. Shan, J., Li, Y., Yang, L., Feng, Q., Han, L., Wang, H.: Dds-slam: Dense semantic neural slam for deformable endoscopic scenes. In: 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 10837–10842. IEEE (2024)
25. Wang, K., Yang, C., Wang, Y., Li, S., Wang, Y., Dou, Q., Yang, X., Shen, W.: Endogslam: Real-time dense reconstruction and tracking in endoscopic surgeries using gaussian splatting. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 219–229. Springer (2024)
26. Wang, Y., Long, Y., Fan, S.H., Dou, Q.: Neural rendering for stereo 3d reconstruction of deformable tissues in robotic surgery. In: MICCAI. pp. 431–441. Springer (2022)
27. Yang, L., Kang, B., Huang, Z., Zhao, Z., Xu, X., Feng, J., Zhao, H.: Depth anything v2. Advances in Neural Information Processing Systems **37**, 21875–21911 (2025)
28. Zhu, Z., Peng, S., Larsson, V., Xu, W., Bao, H., Cui, Z., Oswald, M.R., Pollefeys, M.: Nice-slam: Neural implicit scalable encoding for slam. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12786–12796 (2022)