NFL-BA: Near-Field Light Bundle Adjustment for SLAM in Dynamic Lighting

Andrea Dunn Beltran*[1], Daniel Rho*[1], Marc Niethammer[2], Roni Sengupta [1]
[1] University of North Carolina at Chapel Hill [2] University of California San Diego
* Equal contribution
{asdunnbe, dnl03c1, ronisen}@cs.unc.edu, mniethammer@ucsd.edu

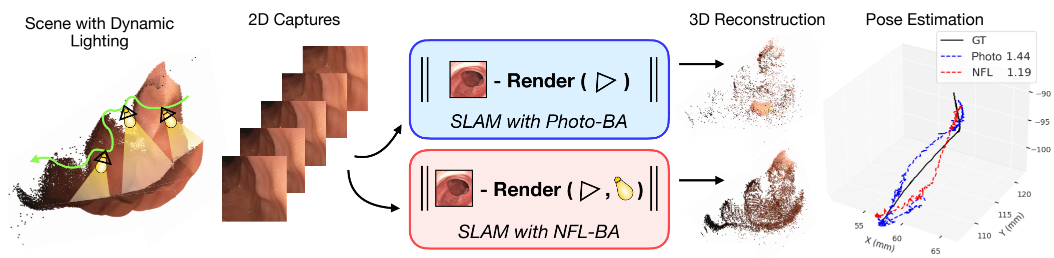

Figure 1: NFL-BA enhances tracking and mapping in neural rendering-based SLAM (e.g., MonoGS [23]) by explicitly modeling dynamic near-field lighting, with applications in endoscopy.

Abstract

Simultaneous Localization and Mapping (SLAM) systems typically assume static, distant illumination; however, many real-world scenarios, such as endoscopy, subterranean robotics, and search & rescue in collapsed environments, require agents to operate with a co-located light and camera in the absence of external lighting. In such cases, dynamic near-field lighting introduces strong, view-dependent shading that significantly degrades SLAM performance. We introduce Near-Field Lighting Bundle Adjustment Loss (NFL-BA) which explicitly models near-field lighting as a part of Bundle Adjustment loss and enables better performance for scenes captured with dynamic lighting. NFL-BA can be integrated into neural rendering-based SLAM systems with implicit or explicit scene representations. Our evaluations mainly focus on endoscopy procedure where SLAM can enable autonomous navigation, guidance to unsurveyed regions, blindspot detections, and 3D visualizations, which can significantly improve patient outcomes and endoscopy experience for both physicians and patients. Replacing Photometric Bundle Adjustment loss of SLAM systems with NFL-BA leads to significant improvement in camera tracking, 37% for MonoGS and 14% for EndoGS, and leads to state-of-the-art camera tracking and mapping performance on the C3VD colonoscopy dataset. Further evaluation on indoor scenes captured with phone camera with flashlight turned on, also demonstrate significant improvement in SLAM performance due to NFL-BA.

1 Introduction

Simultaneous Localization and Mapping (SLAM) enables autonomous agents to build a spatial map of an unknown environment while estimating their own poses within it, with wide-ranging applications in robotics, computer vision, autonomous vehicles, and scientific imaging. Most SLAM systems [37, 33, 6, 59, 62, 57, 48, 19, 27, 15] assume an autonomous agent navigating an environment with distant, static illumination, e.g., a self-driving car in the streets, and they optimize a Photometric Bundle Adjustment loss where they minimize an error between the captured image and the re-rendered image using estimated 3D scene and camera poses.

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

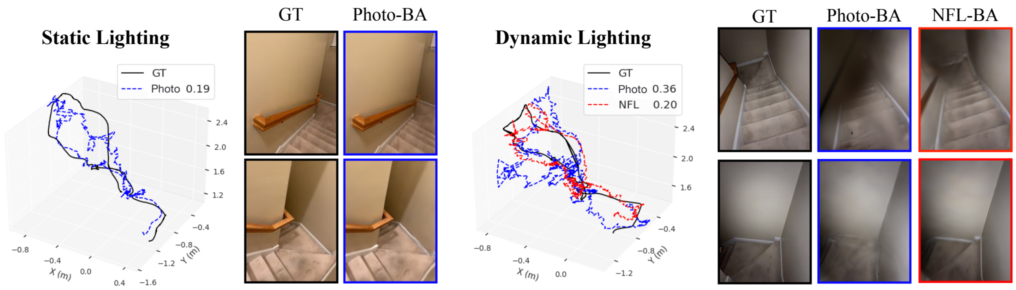

Figure 2: We compare MonoGS performance under (1) distant static lighting and (2) dynamic near-field lighting from a co-located flashlight. Standard photometric BA performs well under static lighting but fails under dynamic lighting, degrading both trajectory and map quality. NFL-BA restores performance under dynamic lighting, matching the quality of the static-light setup.

However, many scientific and safety-critical applications demand that autonomous agents operate in environments devoid of external illumination, relying instead on self-mounted light sources. For example, in endoscopy procedures, a slender flexible tube with a co-located light and camera is used to inspect internal organs such as the airway and the colon [8, 45, 32, 22, 50, 16]. Accurate trajectory estimation is crucial for reliably guiding instruments to areas of interest, mapping anomalies, and avoiding tissue damage during navigation. In subterranean search-and-rescue or collapsed-building inspection, robots rely on onboard lamps to explore unstable voids; slight errors in pose estimation can accumulate into large drift, leading to misaligned maps, missed victims, or costly back-tracking.

Despite the prevalence of these use cases, current SLAM systems perform poorly under such conditions (see Fig. 2). This performance drop is primarily due to the effects of *dynamic near-field lighting*, where the only illumination is co-located with the camera and moves with it. Dynamic near-field lighting causes different points of the surface to receive different intensities of light at each time step, depending on the distance and orientation of the point to the camera, introducing strong, view-dependent shading. These lighting artifacts significantly impair both feature-based and direct (photometric) tracking, resulting in substantial failures in mapping accuracy and pose estimation.

To alleviate these issues, we propose a new Bundle Adjustment loss that accounts for dynamic near-field lighting. Our key intuition is that the shading effect of the captured image can provide valuable information about the relative distance and orientation between the surface and the camera. With this, we formulate a Near-Field Lighting Bundle Adjustment loss, NFL-BA, where we optimize the surface geometry and the camera parameters such that the rendered image has shading variations that match the relative distance and orientation between the surface and the camera. Our NFL-BA loss can be applied to any neural rendering-based SLAM algorithm, i.e., with neural implicit and explicit 3D Gaussian scene representation.

In this paper, we specifically focus on demonstrating how NFL-BA can improve the performance of existing SLAM systems for 3D reconstruction and localization from endoscopy videos. SLAM can enable autonomous navigation through internal organs and guide physicians to unsurveyed regions to improve physicians' situational awareness by providing 3D visualizations, and can help measure organ shapes, e.g. the cross-sectional area of the upper airway to diagnose airway abnormalities [53]. We evaluated NFL-BA with two state-of-the-art 3DGS-based SLAM systems, general-purpose MonoGS [23] and endoscopy-specific EndoGSLAM [45], and one neural implicit SLAM, NICE-SLAM [59], by replacing their Photometric Bundle Adjustment loss with NFL-BA loss. We observe that the NFL-BA loss improves the performance of all SLAM algorithms on average when using ground-truth or estimated depth maps on the C3VD colonoscopy dataset. For example, NFL-BA significantly improves over MonoGS by reducing camera tracking error by 37% (3.48mm to 2.18mm) and camera mapping error by 38% (1.59mm to 0.99mm) when initialized by PPSNet depth[34].

Additionally, we also demonstrate the effectiveness of NFL-BA on indoor rooms captured with a moving co-located light and camera without any external light source, mimicking agent navigation during search & rescue and covert military operations. By replacing incorporating our NFL-BA loss, we see an average improvement of ∼35% in pose estimation across all scenes.

2 Related Works

Dense SLAM and Bundle Adjustment. Early SLAM pipelines focused on sparse feature matching for pose estimation and mapping [30, 5, 42, 10]. With advancements in neural scene representations several proposed SLAM frameworks [59, 61] generate dense, pixel-level that yield more detailed and robust reconstructions. More recently, 3D Gaussian surface methods have demonstrated real-time rendering with high-fidelity mapping[19, 27, 48, 15, 9, 52].

These dense SLAM approaches all rely on a core Bundle Adjustment step. Bundle Adjustment (BA) alternatively optimizes camera parameters and surface geometry by minimizing errors across multiple frames. Traditional geometric BA aligns detected 2D feature points to their 3D counterparts by minimizing reprojection error, assuming static lighting and Lambertian surfaces [13]. Although effective in controlled environments, it struggles in complex or low-texture scenes. Photometric BA (Photo-BA) [1] incorporates pixel intensities into the optimization process, minimizing photometric re-projection errors and proving advantageous in environments where feature matching fails [10]. However, Photo-BA does not exploit the correspondence cues provided by dynamic or near-field lighting where image intensities vary across frames.

Near-field Lighting models. Near-field lighting has been leveraged for 3D reconstruction tasks like monocular depth and surface normal estimation [34, 58] and Photometric Stereo [21]. Some of these approaches [34, 20] use a near-field lighting representation as input to a CNN along with captured images for predicting surface normal and geometry. In the context of Endoscopy, LightDepth [36] and PPSNet [34] demonstrated the effectiveness of near-field lighting to enhance depth estimation. LightNeus [2] exploited the inverse-square law for light decay to improve endoscopic surface reconstruction, however with known camera parameters and pre-operative 3D CT scan.

It has never, however, been used for Simultaneous Localization & Mapping (SLAM) problems, let alone in combination with neural rendering methods. To this end, we propose a Bundle Adjustment Loss with Near-Field Lighting (NFL-BA), considering the most commonly available single co-located camera & light in the endoscope or other autonomous agents.

Dynamic Lighting in SLAM. Visual SLAM performance often degrades under illumination changes such as exposure shifts, specularities, and varying color temperature. Early photometric calibration methods jointly optimize camera intrinsics, exposure, and scene depths to normalize brightness variations in real time [10] while probabilistic SLAMs with unscented filtering further stabilizes pose estimates under uncertain lighting conditions [26]. More recently, learning-based matchers [41, 55] adapt descriptors to cope with complex lighting variations. None of these methods, however, explicitly model near-field lighting geometry to handle this co-located light setting.

SLAM in endoscopy. Early works [39, 11] demonstrated the feasibility of applying SLAM in such environments by addressing dynamic lighting and tissue deformation. Researchers have often used a mixture of supervised learning on synthetic and self-supervised learning on real endoscopy datasets for tailoring SLAM frameworks to endoscopy with complex camera motion [25, 56, 46] and developed novel endoscopy SLAM frameworks [35, 29, 17]. However these techniques often struggle with challenging sequences from both synthetic and clinical data. Recently, neural rendering-based methods [38, 22, 45, 50, 12, 14] have proved especially effective in generating high-quality details and modeling textureless regions with a large number of Gaussians. In this work, we adopt neural rendering approaches and explicitly model the near-field lighting effects, alleviating dynamic lighting challenges and improving performance.

3 Background

In this section, we review the general framework of neural rendering-based SLAM. We represent the camera at time t by its extrinsics $P_t = [R_t, T_t] \in \mathbf{SE}(3)$ and known intrinsics K, yielding the projection $\pi_t = KP_t$. We assume the camera intrinsic K to be the same for all frames and known or calibrated ahead of time. Pixels are denoted p and 3D camera-space points by x.

In neural rendering, scene parameters Θ, whether in the form of neural networks or primitives, encode visual and geometric information, such as colors c_i and occupancy α_i. Given Θ and P_t, we can get the color $\hat{C}(\cdot)$ and the depth $\hat{D}(\cdot)$ of a pixel p from a frame at time t as follows [28, 19]:

$$\hat{C}(p) = \sum_{i \in \mathcal{N}} c_i \alpha_i \Pi_{j=1}^{i-1}(1-\alpha_j)\,, \quad \hat{D}(p) = \sum_{i \in \mathcal{N}} z_i \alpha_i \Pi_{j=1}^{i-1}(1-\alpha_j) \tag{1}$$

where $\mathcal{N}$ denotes the group of samples for a pixel p, with α_i representing the occupancy of the i-th sample, and z_i denotes its distance from the camera center.

To optimize P_t and Θ, dense SLAM methods typically use rendering loss $\mathcal{L}_{ren}$, reducing the rendering errors between the rendered and captured images [59, 23, 48] and, if estimated or ground truth depth maps are available, an additional depth loss $\mathcal{L}_{geo}$ can be added [43]. Typically, these losses take the form of L^p norm as follows with variations with M_t as a pixel-wise mask:

$$\mathcal{L}_{ren} = \|M_t \odot (\hat{C} - C)\|_p, \quad \mathcal{L}_{geo} = \|M_t \odot (\hat{D} - D)\|_p \tag{2}$$

Bundle adjustment optimizes both P_t and Θ using the following combined loss:

$$\text{Photo-BA:} \quad \min \sum_{t \in \mathcal{W}} \lambda_{ren} \mathcal{L}_{ren}(\hat{C}, C; M_t) + \lambda_{geo} \mathcal{L}_{geo}(\hat{D}, D; M_t) \tag{3}$$

where $\mathcal{W}$ denotes the set of frames used for the bundle adjustment and the hyperparameters λ_{ren} and λ_{geo} are the loss weights. Additionally, the objective function can include any other regularization terms, such as artifact suppressing [27] or opacity regularization [60].

During the Mapping stage, both Θ and P_t are optimized over a set of keyframes. The exact algorithm for keyframe selection, keyframe update and optimization strategies for tracking and mapping phase vary between different SLAM approaches and their specific objectives.

Implicit Neural Representations. Neural field-based SLAM methods [40, 59, 44, 62, 37] uses a set of neural networks $F(x, d; \Theta) \rightarrow (c_i, \sigma_i)$, optimized to estimate the color c_i and the volume density σ_i for an input 3D coordinate x and the view direction d. The the occupancy can be calculated from the volume density σ_i and the distance between adjacent samples δ_i as $\alpha_i = 1 - \exp(-\sigma_i \delta_i)$.

3D Gaussian Splatting. For 3D Gaussian Splatting [19] SLAM methods, the scene is represented by a set of Gaussians with mean μ^i, covariance Σ^i in world space, color c_i, and opacity α^i. The shape parameters and occupancy α^i of the *splatted* 2D Gaussians are computed as follows:

$$\bar{\mu}_t^i = \pi_t \mu^i, \quad \bar{\Sigma}_t^i = J_t R_t \Sigma^i R_t^T J_t^T, \quad \alpha_i = \alpha^i \exp(-\frac{1}{2}(p - \bar{\mu}_t^i)^\top \left(\bar{\Sigma}_t^i\right)^{-1} (p - \bar{\mu}_t^i)) \tag{4}$$

where J_t is the Jacobian of the projection π_t, p denotes a pixel coordinate, and $\bar{\mu}_t^i, \bar{\Sigma}_t^i$ are the splatted mean and covariance of Gaussian $\mathcal{G}^i$ in pixel space.

4 Near-Field Light Bundle Adjustment

We introduce a novel Near-Field Lighting based Bundle Adjustment loss, NFL-BA, that integrates near-field lighting with neural-rendering 3D scene representations to improve performance of existing SLAM systems on images captured with dynamic lighting co-located with the camera. Our proposed NFL-BA can replace commonly used Photometric Bundle adjustment loss, defined in Eq. 3, within neural-rendering based SLAM framework. Photo-BA typically optimizes scene appearance parameter as RGB color, which is suffient when the illumination on each scene point remains the constant throughout the capture. However, for scenes with a dynamic light co-located with a moving camera, the illumination received at each point varies per frame as the camera and the light moves through the scene. In this setting, the illumination received at each point depends on the relative distance and orientation between the point and the camera, as conceptualized in Fig. 3. Thus continuing to model scene appearance as simple RGB color is inaccurate for dynamic near-field lighting as it doesn't separate effects of illumination due to camera movement from the intrinsic view-independent color of the scene, i.e. albedo.

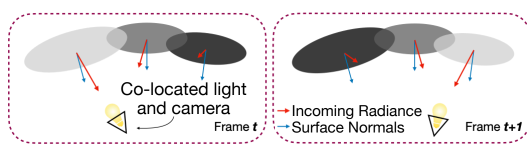

Figure 3: Illustration of our key idea. As the co-located light and camera, moves through the scene, different 3D Gaussians on the surface receive different intensities of light (red arrow), dependent on the relative distance and orientation between the 3D Gaussian and the camera.

Our goal is to explicitly model surface appearance as albedo and separate near-field lighting effects from it. To accurately model dynamic lighting we then represent near-field illumination effects with camera pose and scene geometry. In sec. 4.1 we describe our image formation model using neural rendering framework that will decompose the surface appearance into albedo and incoming lighting, which will be further represented as a function of scene geometry and camera pose. Then in sec. 4.2, we will use this image formation to create the Near-Field Bundle Adjustment loss and show how it can be easily integrated into neural rendering based SLAM framework.

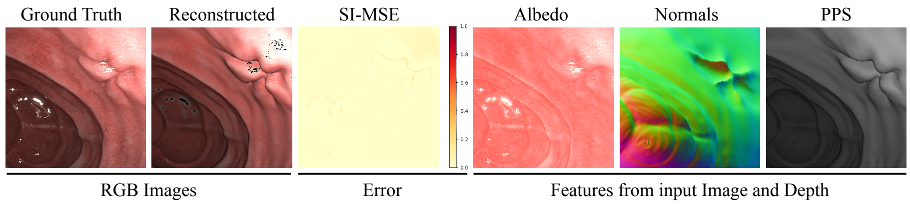

Figure 4: We show that C3VD images captured with a real endoscope conform to our co-located light-camera and zero attenuation β image formation model, as indicated by very low per-pixel scale-invariant MSE (col 3) between the original image (col 1) and the reconstructed image with masked-out specular regions (col 2).

4.1 Image Formation with Near-Field Lighting

We consider an image-formation model under near-field lighting for a single image following previous works [18, 34]. Each pixel p and the corresponding three-dimensional point x_p in the camera space receives different light intensities and directions, characterized by the light source to surface direction $L^d(\cdot)$ and attenuation term $L^a(\cdot)$, as follows:

$$L^d(x_p) = \frac{x_p - x_L}{\|x_p - x_L\|}, \quad L^a(x_p) = \frac{(L^d(x_p)^\top f)^\beta}{\|x_p - x_L\|^2}, \tag{5}$$

where x_L is the location of the light source, f is the forward (optical axis) vector. β is an angular attenuation coefficient, and will be discussed in sec. 4.2.

Assuming a diffuse reflectance model, which has proven effective for depth estimation in endoscopic scenes [34], we can approximate the rendered image at each pixel $\hat{C}(\cdot)$ as:

$$PPS(x_p) = L^a(x_p) \cdot (L^d(x_p)^\top n(x_p)), \qquad \hat{C}(p) = \rho(x_p) PPS(x_p), \tag{6}$$

where $\rho(\cdot)$ and $n(\cdot)$ are albedo and normal at position x_p of pixel p respectively. $PPS(\cdot)$ is a per-pixel shading term. Note that existing approaches that uses this near-field light image formation model [18, 34] uses pixel-based representation to predict depth map or surface geometry from images captured from a single viewpoint only. In this paper, we extend the Near-Field Image Formation model beyond single-view pixel-based representation to multi-view 3D representation.

Our key insight is that the standard volumetric rendering equation can be modified to incorporate the near-field lighting model described in eq. 6, while keeping the overall SLAM pipeline intact. In our framework, we reinterpret the direct color (c_i in eq. 3) as the product of the albedo $\rho(\cdot)$ and the shading term $PPS(\cdot)$, which models dynamic near-field lighting. Note that both albedo and shading is defined directly on the 3D neural representations, i.e. neural radiance field or 3D gaussians, and not in pixel-space. This leads to the modified rendering equation under near-field lighting:

$$\hat{C}_{pps}(p) = \sum_{i \in \mathcal{N}} \rho(x_i) PPS(x_i) \alpha_i \Pi_{j=1}^{i-1}(1 - \alpha_j) \tag{7}$$

Note that eq. 6 represents a special case of eq. 7 where a single sample is considered and the occupancy α_i equals one. Our image formation model assumes diffuse reflectance and no angular attenutation, to reduce the complexity of the modeling. While it is easy to extend the image formation model to handle specular reflectance and angular attenutation of lighting, this leads to additional parameters that needs to be optimized during the Bundle Adjustment.

Angular attenuation. Following previous works [21, 34], we simplify the near-field light image formation model by setting the attenuation coefficient β in eq. 5 to zero. This effectively ignores the directional fall-off component, reducing the light attenuation term to a simple inverse-square fall-off $L^a(x_p) = 1/\|x_p - x_L\|^2$. This simplification is justified because the angular attenuation in settings like endoscopy is often negligible compared to the inverse square law attenuation, and estimating β accurately can be challenging due to variations in endoscope designs. In future work, we plan determine the optimal value of β for different systems and incorporate the light direction vector r_t^e for more accurate modeling which can further improve camera rotation during Bundle Adjustment.

Empirical validation of our image-formation model on colonoscopy image. Fig. 4 provides an example colonoscopy image from the C3VD dataset [3], showing the accuracy of the near-light field model (Eq. 6) with β of 0. Albedo was estimated by converting each RGB image to HSV color space, setting the value channel to 1 across all pixels, then converting the modified image back to

RGB space. This standardizes pixel intensity variations, approximating a reflectance map where illumination effects are minimized, but does not strictly represent ground truth albedo. As shown, the image formulation model is sufficient to represent endoscopic scenes with low reconstruction errors.

4.2 Near-Field Light Bundle Adjustment Loss

Next, we will re-define the Photometric Bundle Adjustment loss of eq. 3 using the near-field lighting based image formation model defined in eq . 7 expressed as follows:

$$\text{NFL-BA:} \quad \min \sum_{t \in \mathcal{W}} \lambda_{ren} \mathcal{L}_{ren}(\hat{C}_{pps}, C; M_t) + \lambda_{geo} \mathcal{L}_{geo}(\hat{D}, D; M_t) \tag{8}$$

where $\hat{C}_{pps}$ denotes the rendered image with near-field lighting-incorporated volumetric rendering equation (Eq. 7). This reformulation seamlessly integrates near-field lighting cues into the neural rendering framework without altering the rest of the SLAM framework. Since our formulation is confined solely to the rendering process, and thus to the bundle adjustment, we do not modify or replace any other SLAM components for fair comparison. This design choice enables easier integration with existing neural rendering-based SLAM methods.

Choice of image space in optimization. Note that many settings, and especially endoscopy, frames are stored in standard sRGB color space, whereas our near-field shading term $PPS(\cdot)$ is computed in a linear space. To ensure consistency, we apply an inverse gamma correction of $\gamma = 2.2$ to the sRGB images before computing PPS, or equivalently, gamma-correct the linear PPS output by $\gamma = 1/2.2$ when rendering back to sRGB. This step aligns the lighting model with the true photometric intensities and prevents bias from the nonlinear sRGB transfer function.

Normal calculation during Bundle Adjustment. To calculate the normals $n(\cdot)$ from neural fields, we utilize the direction of the gradient of the occupancy with respect to the spatial coordinates as follows [4]: $n(x_i) = -\nabla\sigma(x_i)/\|\nabla\sigma(x_i)\|$. For Gaussian Splatting, we use the shortest axis of each Gaussian as its normal, following [51, 7, 47]. In both cases, we ensure the computed normal is oriented towards the camera by enforcing $n(x)^\top L^d(x)$ to be positive. Otherwise, we flip the normals by multiplying them by -1 for stability.

5 Evaluation

Our proposed method is a plug-in approach that can be applied to any existing neural-rendering-based SLAM framework. We first test our method on endoscopy videos using one neural implicit SLAM, NICE-SLAM [59], as well as two existing 3DGS-SLAM frameworks: the general-purpose MonoGS [27] and the endoscopy-specific EndoGSLAM [45]. In each case, we replace the standard Photometric Bundle Adjustment loss (3) with our proposed equation NFL-BA loss (8). Additionally, we also test MonoGS [27] on self-captured indoor scenes with a co-located light and camera.

Table 1: Quantitative Evaluation on the C3VD [3] dataset with oracle depth map. Replacing Photometric BA with NFL-BA significantly improves tracking quality of two state-of-the-art 3D Gaussian SLAMs, MonoGS [23] and EndoGS [60], and one neural implicit SLAM, NICE-SLAM [59].

		Tracking		Mapping
Method	BA	ATE$_t$ (mm)↓	ATE$_r$ (°)↓	Chamfer (mm)↓
NICE-SLAM,	Photo	4.16	2.68	1.95
CVPR'22	NFL	**2.88**	2.81	**1.70**
EndoGSLAM,	Photo	1.93	1.81	0.85
MICCAI'24	NFL	2.04	**1.13**	0.97
MonoGS,	Photo	2.90	1.11	1.16
CVPR'24	NFL	**1.60**	1.49	**0.79**

5.1 Evaluation Setting

Datasets. We evaluate our method on three datasets that reflect different challenges in handling near-field dynamic lighting: (1) a phantom endoscopy dataset, (2) a clinical endoscopy dataset, and (3) a dataset of indoor scenes captured with phone camera with flashlight turned on.

C3VD. The C3VD dataset [3] (CC BY-NC-SA 4.0) was created using a phantom colon with synthetic materials to simulate realistic tissue geometry. The endoscopy video was captured by a surgeon who performs different endoscopy procedures on the phantom colon with a real endoscope capturing RGB images coupled with corresponding depth maps. We focus on 8 sequences ranging from 70 to 800 frames from different regions of the colon, for more details, please see supplementary. We evalute using both ground truth and predicted depths.

Colon10K. To test generalization in real-world clinical endoscopy settings, we evaluate on Colon10K [24], a large-scale video dataset without depth or pose supervision. Videos are sampled from actual

Table 2: Quantitative Evaluation on the C3VD [3] dataset with depth maps estimated by SOTA techniques, PPSNet [34] and DA-Hybrid. Replacing Photometric BA with NFL-BA significantly improves tracking for both MonoGS [23] and EndoGS [60], and mapping and rendering quality for MonoGS [23]. Note that SOTA performance for each of the tracking, mapping, and rendering metrics is observed when NFL-BA is used.

		ATE_t (mm)↓		ATE_r (°)↓		Chamfer (mm)↓		LPIPS↓	
Method	BA	PPS-Net	DaHybrid	PPS-Net	DaHybrid	PPS-Net	DaHybrid	PPS-Net	DaHybrid
EndoGSLAM,	Photo	3.03	6.67	1.73	2.26	**1.23**	**2.12**	**0.39**	0.43
MICCAI'24	NFL	**2.62**	**3.91**	**1.24**	**1.58**	1.25	2.39	**0.39**	**0.42**
MonoGS,	Photo	3.48	4.63	1.70	1.69	1.59	1.34	0.56	0.52
CVPR'24	NFL	**2.18**	**2.35**	**1.65**	**1.14**	**0.99**	**1.13**	**0.53**	0.52

procedures and are typically around 300-600 frames. This setting is significantly more challenging than the phantom setting, since frames may contain motion blur, specular highlights, and fluid occlusions. Sequences are uniformly sampled and fisheye corrected.

Self-Captured Indoor Scenes. To study the role of near-field lighting in a controlled non-clinical setting, we capture a dataset of four indoor scenes (*Guitars*, *Porch*, *Pool*, and *Stairs*) using an RGB-D camera. Scenes include objects with varied geometry and reflectance (diffuse, specular), imaged under dynamic motion. Scenes are captured using a co-located point light source mounted to camera. Ground-truth camera trajectories were recorded via motion capture, but no reference point clouds are available; hence, we report only trajectory error (ATE_t) and perceptual quality (LPIPS), omitting Chamfer distance. Details and additional visualizations are included in the supplementary.

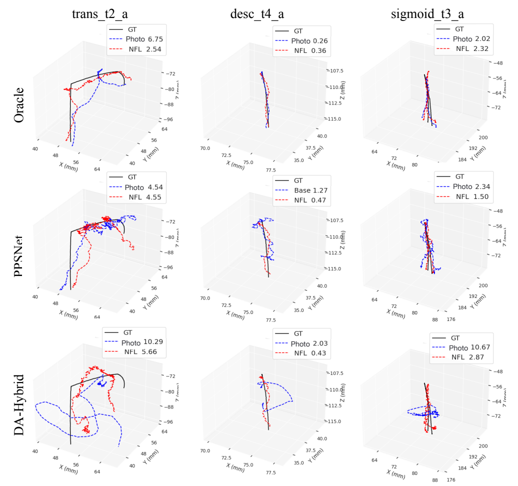

Figure 5: Camera tracking improvement over EndoGSLAM [45]. Replacing the Photo-BA loss (in blue) with NFL-BA loss (in red) significantly improves camera tracking for different depth initialization. Average tracking error ATE_t for each sequence is reported in the inset. (zoom for details)

Metrics. For evaluation, we basically followed other neural rendering SLAM algorithms [23, 45]. For tracking performance, we measure the root mean square error of the Absolute Trajectory Error (ATE) for both translation and rotation across all frames. Translation error ATE_t is in millimeters (mm) for the endoscopy scenes and meters (m) for the in-door scenes. and rotation error ATE_r is in degrees. To assess the mapping quality, we use the Chamfer distance from ground truth point clouds to the nearest points in the estimated point clouds [46], for more details please see supplementary. In addition, we evaluate rendering quality using the Learned Perceptual Image Patch Similarity (LPIPS) [54]. We note that for many endoscopic SLAM applications, tracking and mapping accuracies are more important than photorealism of the rendered images, unlike many indoor or outdoor scenes.

Computational costs We trained all models on a single NVIDIA RTX A6000 GPU. The per-scene optimization takes ∼1 FPS. For more information on runtime speed, please see supplemental.

5.2 Evaluation on C3VD Endoscopy Data

Because NFL-BA is designed as a drop-in replacement for photometric bundle adjustment, we only adjusted the two associated loss weights; all other hyperparameters remain identical between the Photo-BA and NFL-BA experiments. Please see supplemental for detailed hyperparameter settings.

SLAM with oracle depth map. In Tab. 1 we replace Photometric Bundle Adjustment loss with NFL-BA loss for depth is initialized with ground-truth or oracle. NFL-BA significantly improves camera localization (ATE_t) and mapping for NICE-SLAM and MonoGS, and only camera rotation (ATE_r) for EndoGSLAM. EndoGSLAM was specifically designed for synthetic data with an oracle depth map, and we will show later that for estimated depth maps or real endoscopy videos, it performs significantly worse than MonoGS Ground-truth depths are never available during endoscopy, and the majority of endoscopes hardly have any depth sensors.

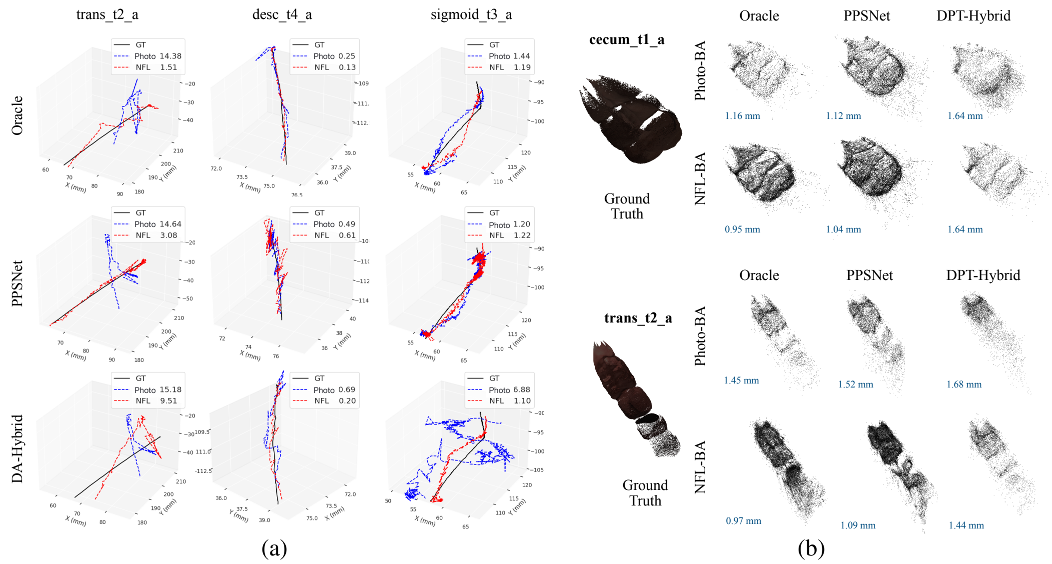

Figure 6: (a) Camera tracking improvement over MonoGS [23]. Replacing the Photo BA loss (in blue) with NFL-BA loss (in red) significantly improves camera tracking for different depth initialization. Average tracking error ATE$_t$ for each sequence is reported in the inset. (b) Reconstructed point clouds using MonoGS [23] show that NFL-BA improves coverage and density while reducing scatter compared to Photometric BA, as measured by Chamfer distance. (zoom for details)

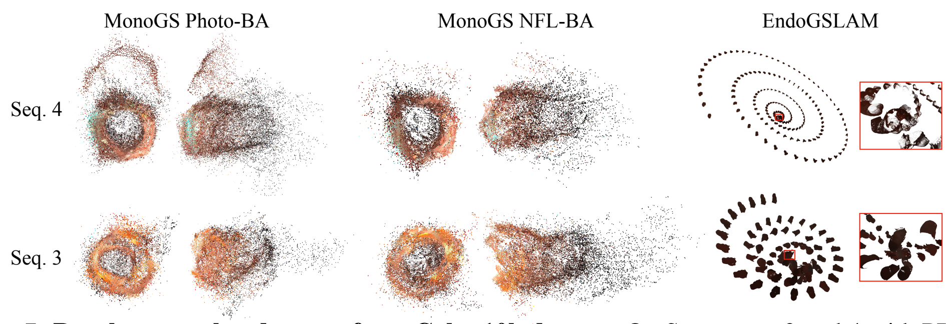

Figure 7: **Results on real endoscopy from Colon10k dataset.** On Sequences 3 and 4 with PPSNet depth, NFL-BA improves MonoGS tracking and mapping, yielding more coherent, elongated colon structures, while EndoGSLAM fails under sudden camera motion.

SLAM with predicted depth map. Under realistic conditions with estimated depths, NFL-BA's impact is even more pronounced. In Tab. **??** we replace Photometric BA loss with NFL-BA loss for MonoGS [23] and EndoGSLAM [45] for depth maps we use PPSNet [34], a state-of-the-art monocular depth estimation algorithm for endoscopy, and fine-turned general-purpose depth estimator, which we will call it as DA-Hybrid - DepthAnything[49] with DINOv2 encoder [31]. NFL-BA significantly improves camera localization (ATE_t) and camera rotation (ATE_r) for both MonoGS [23] and EndoGSLAM [45] while producing similar rendering quality. For example, camera localization for MonoGS is improved by 37% for PPSNet and 49% for DA-Hybrid depth initialization. Mapping accuracy of MonoGS also improves by 37% for PPSNet and 16% for DA-Hybrid depth maps. Overall, these results demonstrate that NFL-BA can compensate for noisy depth estimation and improves performance. Across all four metrics, for tracking, mapping, and rendering, the SOTA performance on the C3VD dataset is in fact achieved when NFL-BA loss is used in the SLAM framework.

5.3 Evaluation on Real Endoscopy Data

We show results on real endoscopy sequence from Colon10k sequence 3 and 4 in Fig. 7. EndoGSLAM fails to construct any real structure, with many disconnected regions along a spiral trajectory. EndoGSLAM assumes constant velocity and is not robust to the sudden motion common in endoscopy procedures, which is significantly more in real data than C3VD. This results extremely poor or failed reconstructions.

Table 3: Quantitative results on four self-captured indoor scenes under dynamic lighting, comparing MonoGS with standard Photo-BA versus NFL-BA. For each scene, the best of each metric is bold.

	Guitars		Porch		Pool		Stairs	
BA	ATE_t (m)↓	LPIPS ↓	ATE_t (m)↓	LPIPS ↓	ATE_t (m)↓	LPIPS ↓	ATE_t (m)↓	LPIPS ↓
Photo	0.30	0.39	0.50	**0.49**	0.41	0.46	0.36	0.40
NFL	**0.18**	**0.37**	**0.35**	0.50	**0.30**	**0.44**	**0.20**	**0.31**

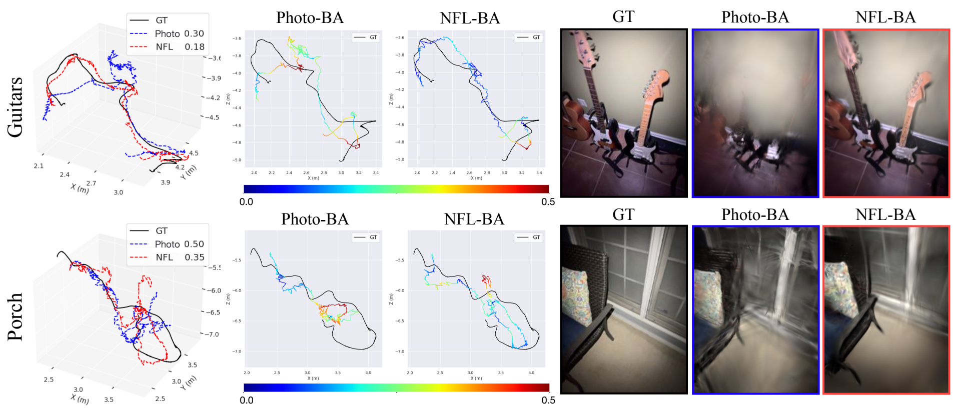

Figure 8: **Results on indoor scenes captured with co-located flashlight and phone camera.** Qualitative comparison on two self-captured indoor scenes using MonoGS with standard Photo-BA versus NFL-BA. (left) Estimated camera trajectories overlaid on ground truth. (center) Per-frame tracking error relative to ground truth. (right) Example re-rendered views, illustrating the sharper, more accurate reconstructions enabled by NFL-BA.

Sequence 4. This pull-back "down-the-barrel" sequence exposes a clear cylindrical lumen. With Photo-BA, MonoGS captures the overall shape but produces a broken segment due to trajectory drift. NFL-BA corrects this, yielding a continuous "hollow-center" reconstruction. Minor artifacts from extreme specular highlights remain (green points), as detailed in the supplement.

Sequence 3. In the extended traversal, both Photo-BA and NFL-BA recover the colon's general geometry, but NFL-BA produces a longer, tighter model with less point scatter. It also better preserves interior ridges (interactive point clouds in the supplement).

5.4 Evaluation on Indoor Scene

To validate NFL-BA in a non-medical setting, we evaluate on four indoor scenes. Table 3 shows that replacing standard Photometric BA with NFL-BA yields substantial reductions in ATE_t across all scenes: from 0.30m to 0.18m (40%) in *Guitar*, 0.50m to 0.35m (30%) in *Outdoor*, 0.41m to 0.30m (27%) in *Pool*, and 0.36m to 0.20m (44%) in *Stair*. On average, NFL-BA reduces tracking error by $\sim 35\%$, demonstrating that near-field shading cues greatly enhance pose estimation even in richly textured, well-lit indoor environments. While LPIPS remains largely comparable, with slight improvements in *Guitars* and *Stairs* and minor variations in *Porch* and *Pool*, the primary benefit of NFL-BA is clear in trajectory accuracy (see Fig. 8).

6 Conclusions

In this paper, we presented a novel bundle adjustment loss that explicitly models dynamic near-field lighting by incorporating light intensity fall-off based on the relative distance and orientation between the surface and the co-located light and camera. This formulation is especially effective for endoscopic scenes, where traditional geometric or photometric bundle adjustment losses struggle under dynamic near-field lighting conditions on textureless surfaces. We demonstrated the general applicability of our approach by integrating it into three different neural rendering-based SLAM methods, improving performance on a challenging endoscopy dataset and indoor scenes captured with a phone camera with a flashlight turned on.

Limitations. While our new formulation for SLAM effectively represents scenes with co-located and dynamic lighting environments, it is currently limited in handling specular reflections, sub-surface scattering, and inter-reflections. Incorporating a more complex image formulation is beyond the scope of the current work, and addressing these remains a promising direction for future research.

7 Acknowledgments

This work is supported by a National Institute of Health (NIH) project #R21EB035832 "Next-gen 3D Modeling of Endoscopy Videos" and #R21EB037440 “Gen-AI Airway Simulator for 3D Endoscopy”. We also thank Stephen M. Pizer and Dr. Sarah McGill for helpful discussions during the project.

References

[1] Hatem Alismail, Brett Browning, and Simon Lucey. Photometric bundle adjustment for vision-based slam, 2016.

[2] Víctor M Batlle, José MM Montiel, Pascal Fua, and Juan D Tardós. LightNeuS: Neural surface reconstruction in endoscopy using illumination decline. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2023.

[3] Taylor L Bobrow, Mayank Golhar, Rohan Vijayan, Venkata S Akshintala, Juan R Garcia, and Nicholas J Durr. Colonoscopy 3d video dataset with paired depth from 2d-3d registration. *Medical Image Analysis*, page 102956, 2023.

[4] Mark Boss, Raphael Braun, Varun Jampani, Jonathan T. Barron, Ce Liu, and Hendrik P.A. Lensch. Nerd: Neural reflectance decomposition from image collections. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12684–12694, October 2021.

[5] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. M. Montiel, and J. D. Tardós. Orb-slam3: An accurate open-source library for visual, visual-inertial, and multi-map slam. *IEEE Transactions on Robotics*, 37(6):1874–1890, 2021.

[6] Devendra Singh Chaplot, Ruslan Salakhutdinov, Abhinav Gupta, and Saurabh Gupta. Neural topological slam for visual navigation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12875–12884, 2020.

[7] Hanlin Chen, Fangyin Wei, Chen Li, Tianxin Huang, Yunsong Wang, and Gim Hee Lee. Vcr-gaus: View consistent depth-normal regularizer for gaussian surface reconstruction. *arXiv preprint arXiv:2406.05774*, 2024.

[8] B. Cui, H. Zhang, X. Li, W. Zhou, and T. Cheng. Endodac: Efficient adapting foundation model for self-supervised depth estimation from any endoscopic camera. In *Proceedings of the Medical Image Computing and Computer-Assisted Intervention Conference (MICCAI)*, 2024.

[9] Tianchen Deng, Yaohui Chen, Leyan Zhang, Jianfei Yang, Shenghai Yuan, Jiuming Liu, Danwei Wang, Hesheng Wang, and Weidong Chen. Compact 3d gaussian splatting for dense visual slam, 2024.

[10] J. Engel, V. Koltun, and D. Cremers. Direct sparse odometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(3):611–625, 2018.

[11] Oscar G. Grasa, Javier Civera, and J. M. M. Montiel. Ekf monocular slam with relocalization for laparoscopic sequences. In *2011 IEEE International Conference on Robotics and Automation*, pages 4816–4821, 2011.

[12] Jiaxin Guo, Jiangliu Wang, Di Kang, Wenzhen Dong, Wenting Wang, and Yun-hui Liu. Free-SurGS: SfM-Free 3D Gaussian Splatting for Surgical Scene Reconstruction . In *proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, volume LNCS 15007. Springer Nature Switzerland, October 2024.

[13] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.

[14] Michel Hayoz, Christopher Hahne, Thomas Kurmann, Max Allan, Guido Beldi, Daniel Candinas, Pablo Márquez-Neila, and Raphael Sznitman. Online 3D reconstruction and dense tracking in endoscopic videos . In *proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, volume LNCS 15006. Springer Nature Switzerland, October 2024.

[15] Huajian Huang, Longwei Li, Cheng Hui, and Sai-Kit Yeung. Photo-slam: Real-time simultaneous localization and photorealistic mapping for monocular, stereo, and rgb-d cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

[16] Yiming Huang, Beilei Cui, Long Bai, Ziqi Guo, Mengya Xu, Mobarakol Islam, and Hongliang Ren. Endo-4DGS: Endoscopic Monocular Scene Reconstruction with 4D Gaussian Splatting . In *proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, volume LNCS 15006. Springer Nature Switzerland, October 2024.

[17] Raúl Iranzo, Víctor M Batlle, Juan D Tardós, and José MM Montiel. Endometric: Near-light metric scale monocular slam. *arXiv preprint arXiv:2410.15065*, 2024.

[18] Y. Iwahori, H. Sugie, and N. Ishii. Reconstructing shape from shading images under point light source illumination. In *[1990] Proceedings. 10th International Conference on Pattern Recognition*, volume i, pages 83–87 vol.1, 1990.

[19] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkuehler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4), July 2023.

[20] C. Lichy and Other authors. Photometric stereo with near-field lighting using neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages zz–aa, 2022.

[21] Daniel Lichy, Soumyadip Sengupta, and David W. Jacobs. Fast light-weight near-field photometric stereo. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12612–12621, June 2022.

[22] Yifan Liu, Chenxin Li, Chen Yang, and Yixuan Yuan. Endogaussian: Gaussian splatting for deformable surgical scene reconstruction. *arXiv preprint arXiv:2401.12561*, 2024.

[23] J. Lu, Y. Zhang, and X. Chen. Monogs: Monocular gaussian splatting for robust slam. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2023.

[24] Ruibin Ma, Sarah K. McGill, Rui Wang, Julian Rosenman, Jan-Michael Frahm, Yubo Zhang, and Stephen Pizer. Colon10k: A benchmark for place recognition in colonoscopy. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 1279–1283, 2021.

[25] Ruibin Ma, Rui Wang, Stephen Pizer, Julian Rosenman, Sarah K McGill, and Jan-Michael Frahm. Real-time 3d reconstruction of colonoscopic surfaces for determining missing regions. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part V 22*, pages 573–582. Springer, 2019.

[26] R. Martínez-Cantín and J. D. Tardós. Unscented kalman filter for visual slam with photometric calibration. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1234–1241, 2016.

[27] Hidenobu Matsuki, Riku Murai, Paul H.J. Kelly, and Andrew J. Davison. Gaussian splatting slam. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 18039–18048, June 2024.

[28] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: representing scenes as neural radiance fields for view synthesis. *Commun. ACM*, 65(1):99–106, December 2021.

[29] Javier Morlana, Juan D Tardós, and José MM Montiel. Topological slam in colonoscopies leveraging deep features and topological priors. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 733–743. Springer, 2024.

[30] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos. Orb-slam: A versatile and accurate monocular slam system. In *IEEE Transactions on Robotics*, volume 31, pages 1147–1163, 2015.

[31] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy V. Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel HAZIZA, Francisco Massa, Alaaeldin El-Nouby, Mido Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Herve Jegou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. DINOv2: Learning robust visual features without supervision. *Transactions on Machine Learning Research*, 2024. Featured Certification.

[32] Kutsev Bengisu Ozyoruk, Guliz Irem Gokceler, Taylor L. Bobrow, Gulfize Coskun, Kagan Incetan, Yasin Almalioglu, Faisal Mahmood, Eva Curto, Luis Perdigoto, Marina Oliveira, Hasan Sahin, Helder Araujo, Henrique Alexandrino, Nicholas J. Durr, Hunter B. Gilbert, and Mehmet Turan. Endoslam dataset and an unsupervised monocular visual odometry and depth estimation approach for endoscopic videos. *Medical Image Analysis*, 71:102058, 2021.

[33] Albert Palomer, Pere Ridao, and David Ribas. Inspection of an underwater structure using point-cloud slam with an auv and a laser scanner. *Journal of field robotics*, 36(8):1333–1344, 2019.

[34] Akshay Paruchuri, Samuel Ehrenstein, Shuxian Wang, Inbar Fried, Stephen M Pizer, Marc Niethammer, and Roni Sengupta. Leveraging near-field lighting for monocular depth estimation from endoscopy videos. In *Computer Vision – ECCV 2024*, Cham, 2024. Springer Nature Switzerland.

[35] Juan J Gómez Rodríguez, José MM Montiel, and Juan D Tardós. Nr-slam: Non-rigid monocular slam. *IEEE Transactions on Robotics*, 2024.

[36] Javier Rodríguez-Puigvert*, Víctor M. Batlle*, José María M. Montiel, Rubén Martínez-Cantín, Pascal Fua, Juan D. Tardós, and Javier Civera. LightDepth: single-view depth self-supervision from illumination decline. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023.

[37] Erik Sandström, Yue Li, Luc Van Gool, and Martin R Oswald. Point-slam: Dense neural point cloud-based slam. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 18433–18444, 2023.

[38] Yufei Shi, Beijia Lu, Jia-Wei Liu, Ming Li, and Mike Zheng Shou. Colonnerf: Neural radiance fields for high-fidelity long-sequence colonoscopy reconstruction. *arXiv preprint arXiv:2312.02015*, 2023.

[39] D. Stoyanov, G. Mylonas, F. Deligianni, A. Darzi, and G.-Z. Yang. Soft-tissue motion tracking and structure estimation for robotic assisted mis procedures. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 139–146, 2005.

[40] E. Sucar, S. Liu, J. Ortiz, and A. J. Davison. imap: Implicit mapping and positioning in real-time. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 6229–6238, 2021.

[41] X. Sun, O. Kähler, and K. Sattler. Lightglue: Feature matching under adverse conditions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 500–509, 2023.

[42] Zachary Teed and Jia Deng. Droid-slam: Deep visual slam for monocular, stereo, and rgb-d cameras. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 16558–16569. Curran Associates, Inc., 2021.

[43] Fabio Tosi, Youmin Zhang, Ziren Gong, Erik Sandström, Stefano Mattoccia, Martin R Oswald, and Matteo Poggi. How nerfs and 3d gaussian splatting are reshaping slam: a survey. *arXiv preprint arXiv:2402.13255*, 4, 2024.

[44] Hengyi Wang, Jingwen Wang, and Lourdes Agapito. Co-slam: Joint coordinate and sparse parametric encodings for neural real-time slam. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13293–13302, June 2023.

[45] Kailing Wang, Chen Yang, Yuehao Wang, Sikuang Li, Yan Wang, Qi Dou, Xiaokang Yang, and Wei Shen. EndoGSLAM: Real-Time Dense Reconstruction and Tracking in Endoscopic Surgeries using Gaussian Splatting . In *proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, volume LNCS 15006. Springer Nature Switzerland, October 2024.

[46] Shuxian Wang, Yubo Zhang, Sarah K. McGill, Julian G. Rosenman, Jan-Michael Frahm, Soumyadip Sengupta, and Stephen M. Pizer. A surface-normal based neural framework for colonoscopy reconstruction. In Alejandro Frangi, Marleen de Bruijne, Demian Wassermann, and Nassir Navab, editors, *Information Processing in Medical Imaging*, pages 797–809, Cham, 2023. Springer Nature Switzerland.

[47] Qianyi Wu, Jianmin Zheng, and Jianfei Cai. Surface reconstruction from 3d gaussian splatting via local structural hints. In *Computer Vision – ECCV 2024*, Cham, 2024. Springer Nature Switzerland.

[48] Chi Yan, Delin Qu, Dan Xu, Bin Zhao, Zhigang Wang, Dong Wang, and Xuelong Li. Gs-slam: Dense visual slam with 3d gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19595–19604, 2024.

[49] Lihe Yang, Bingyi Kang, Zilong Huang, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth anything: Unleashing the power of large-scale unlabeled data. In *CVPR*, 2024.

[50] Shuojue Yang, Qian Li, Daiyun Shen, Bingchen Gong, Qi Dou, and Yueming Jin. Deform3DGS: Flexible Deformation for Fast Surgical Scene Reconstruction with Gaussian Splatting . In *proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, volume LNCS 15006. Springer Nature Switzerland, October 2024.

[51] Keyang Ye, Qiming Hou, and Kun Zhou. 3d gaussian splatting with deferred reflection. In *ACM SIGGRAPH 2024 Conference Papers*, SIGGRAPH '24, New York, NY, USA, 2024. Association for Computing Machinery.

[52] Vladimir Yugay, Yue Li, Theo Gevers, and Martin R. Oswald. Gaussian-slam: Photo-realistic dense slam with gaussian splatting, 2024.

[53] Carlton Zdanski, Stephanie Davis, Yi Hong, Di Miao, Cory Quammen, Sorin Mitran, Brad Davis, Marc Niethammer, Julia Kimbell, Elizabeth Pitkin, Jason Fine, Lynn Fordham, Brad Vaughn, and Richard Superfine. Quantitative assessment of the upper airway in infants and children with subglottic stenosis. *The Laryngoscope*, 126, 07 2015.

[54] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[55] Y. Zhang, M. Oswald, and D. Cremers. Airslam: Illumination-invariant hybrid slam. In *International Conference on Computer Vision (ICCV)*, pages 2345–2354, 2023.

[56] Yubo Zhang, Jan-Michael Frahm, Samuel Ehrenstein, Sarah K McGill, Julian G Rosenman, Shuxian Wang, and Stephen M Pizer. Colde: a depth estimation framework for colonoscopy reconstruction. *arXiv preprint arXiv:2111.10371*, 2021.

[57] S. Zhi, J. Lai, A. Kundu, M. Bloesch, A. Davison, and A. Zisserman. Nerf-slam: Real-time dense monocular slam with neural radiance fields. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2022.

[58] L. Zhou, R. Klette, and K. Scheibe. A multi-image shape-from-shading framework for near-lighting perspective endoscopes. *International Journal of Computer Vision*, 82:1–24, 2009.

[59] A. Zhu, Z. Zhang, H. Su, L. Li, G. Wang, and X. Li. Nice-slam: Neural implicit scalable encoding for slam. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[60] Lingting Zhu, Zhao Wang, Jiahao Cui, Zhenchao Jin, Guying Lin, and Lequan Yu. Endogs: Deformable endoscopic tissues reconstruction with gaussian splatting. In M. Emre Celebi, Mauricio Reyes, Zhen Chen, and Xiaoxiao Li, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024 Workshops*, pages 135–145, Cham, 2025. Springer Nature Switzerland.

[61] Z. Zhu, T. Yu, X. Zhang, J. Li, Y. Zhang, and Y. Fu. Neuralrgb-d: Neural representations for depth estimation and scene mapping. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2022.

[62] Zihan Zhu, Songyou Peng, Viktor Larsson, Zhaopeng Cui, Martin R Oswald, Andreas Geiger, and Marc Pollefeys. Nicer-slam: Neural implicit scene encoding for rgb slam. In *International Conference on 3D Vision (3DV)*, March 2024.