Gaussian Pancakes: Geometrically-Regularized 3D Gaussian Splatting for Realistic Endoscopic Reconstruction

Sierra Bonilla[1,2], Shuai Zhang[1,3], Dimitrios Psychogyios[1,2], Danail Stoyanov[1,2], Francisco Vasconcelos[1,2], and Sophia Bano[1,2]

[1] Wellcome/EPSRC Centre for Interventional and Surgical Sciences (WEISS), UK
{sierra.bonilla.21,shuai.z}@ucl.ac.uk
[2] Department of Computer Science, University College London, London, UK
[3] Department of Medical Physics and Biomedical Engineering, University College London, UK

Abstract. Within colorectal cancer diagnostics, conventional colonoscopy techniques face critical limitations, including a limited field of view and a lack of depth information, which can impede the detection of precancerous lesions. Current methods struggle to provide comprehensive and accurate 3D reconstructions of the colonic surface which can help minimize the missing regions and reinspection for pre-cancerous polyps. Addressing this, we introduce "Gaussian Pancakes", a method that leverages 3D Gaussian Splatting (3D GS) combined with a Recurrent Neural Network-based Simultaneous Localization and Mapping (RNNSLAM) system. By introducing geometric and depth regularization into the 3D GS framework, our approach ensures more accurate alignment of Gaussians with the colon surface, resulting in smoother 3D reconstructions with novel viewing of detailed textures and structures. Evaluations across three diverse datasets show that Gaussian Pancakes enhances novel view synthesis quality, surpassing current leading methods with a 18% boost in PSNR and a 16% improvement in SSIM. It also delivers over $100\times$ faster rendering and more than $10\times$ shorter training times, making it a practical tool for real-time applications. Hence, this holds promise for achieving clinical translation for better detection and diagnosis of colorectal cancer. Code: https://github.com/smbonilla/GaussianPancakes.

Keywords: Gaussian Splatting · 3D Reconstruction · Robotic Surgery

1 Introduction

Colorectal cancer remains a major global health challenge, consistently ranking among the top three cancers in prevalence and mortality [1]. The slow progression of colorectal cancer offers detecting pre-cancerous growth through colonoscopy. Despite several AI-based methods proposed in the literature for polyp detection during colonoscopy [6], challenges persist, particularly in identifying polyps hidden behind folds, which account for up to three-quarters of all missed cases [19].

Colonoscopists encounter difficulties in thorough surface inspection, due to the use of a monocular camera with limited field of view and lack of 3D information while navigating the colon's complex structure. Generating full 3D reconstructions with high-quality textures from endoscopic images in near real-time would enable improved diagnosis and treatment through downstream tasks like automated surgical path planning [2], AR/VR training environments[25], AI-based polyp detection [6], and for determining missing regions for re-inspection [17].

The exploration of novel view synthesis and 3D reconstruction in endoscopy has significantly advanced with the advent of Neural Radiance Fields (NeRF) [18] and Neural Implicit Surfaces (NeuS) [26] which learn a continuous function that implicitly represents the 3D scene trained from 2D images and paired camera poses. However, NeRF and NeuS, including their colonoscopy-specific adaptations ColonNeRF[24], REIM-NeRF[20] and LightNeuS[7], are limited by lengthy training times, slow rendering speeds, and a lack of explicit geometry, which complicates integration into practical workflows. Both NeRF and NeuS rely on pre-known camera poses and depths, further restricting their versatility. To address these challenges, the field of computer vision and surgical vision[11,16,29] has recently shifted its attention to 3D Gaussian Splatting (3D GS)[13]. This technique represents 3D scenes through populations of 3D Gaussians, offering real-time rendering, explicit manipulable geometry for easier integration, and significantly reduced training times. Nonetheless, it faces challenges, including a lack of geometry constraints leading to artifacts in untrained views and the necessity of a robust point cloud and known camera poses as starting points.

In response to these limitations, we propose Gaussian Pancakes, an approach that synergizes a geometrically-regularized 3D GS with a colonoscopy-tailored Simultaneous Localization and Mapping (SLAM) system, RNNSLAM [17], to enhance texture rendering and anatomical accuracy while boosting training and rendering speeds. Our method capitalizes on the strengths of RNNSLAM, which generates camera poses, depth maps, and rudimentary surface reconstruction from only the monocular video inputs in real-time, providing a robust alternative to traditional Structure-from-Motion (SfM) methods. Although RNNSLAM alone yields surface reconstructions, the lack of photorealism and anatomical detail limits its use in novel view synthesis tasks. Gaussian Pancakes overcomes these drawbacks by replacing surfels with 3D Gaussians and incorporating surface-based regularization, resulting in significantly accurate radiance fields and vivid textures. Our key contributions are: **1)** A pipeline that integrates 3D GS with SLAM for photorealistic 3D colonoscopic reconstructions, demonstrating robustness over using a SfM method seen in Fig. 1 (C-D); **2)** Improved 3D GS method by incorporating geometric and depth regularizations, aligning Gaussians with the surfaces by effectively "pancaking" them. This enhancement reduces geometric error and artifacts in novel view synthesis, seen in Fig. 1; **3)** Optimized training and rendering process for radiance fields in surgical scenes, reducing training times to roughly 2 minutes and improving image rendering speeds by more than 100 times, making it more applicable in clinical settings.
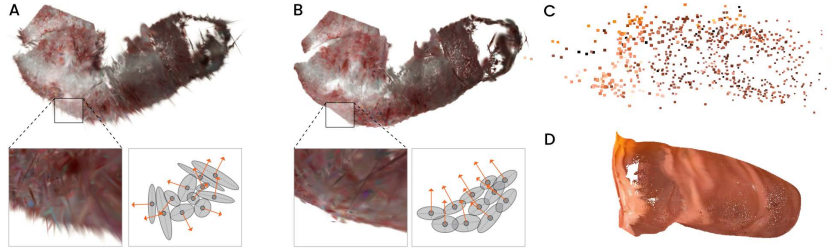
Fig. 1. Illustrating the benefit of Gaussian Pancakes, arrows indicating Gaussians' principal normal direction, A) Gaussian Splatting without Pancaking (PSNR = 37.54) and B) Gaussian Splatting with Pancaking (PSNR = 38.50), on a synthetic video sequence generated by a colonoscopy simulator [28]; C) Sparse point cloud from SfM and D) point cloud from RNNSLAM on the In-Vivo dataset [17].

2   Method

Gaussian Pancakes employs 3D GS for realistic texture rendering and RNNSLAM for 3D reconstruction. As shown in Fig . 2, it starts with RNNSLAM-generated camera poses, depth maps, and meshes. This mesh is sampled, using stratified sampling, to create a sparse RGB point cloud [3]. Around these points, Gaussians with corresponding attributes are initialized (Sec. 2.2). Supervision is twofold: photometric and geometric regularization, with the addition of geometric regularization ensuring Gaussians align with the surface and limit floating artifacts.

2.1   Preliminaries of RNNSLAM

RNNSLAM [17] uses sparse depth estimation from a SfM method as a proxy for ground truth to train a recurrent neural network for depth and camera pose prediction. Bundle-adjusted direct sparse odometry is used to jointly optimize the predicted poses and depth by minimizing the intensity difference over a window of recent frames. The fusion pipeline is then used to reconstruct colon meshes with textures, enabling the detection of missing regions. While this approach has shown effectiveness in reconstructing dense 3D colon maps from visible sections in real colonoscopy, RNNSLAM suffers from pose estimation drift, resulting in texture misalignment and structural collapse in the fused meshes, which is a common problem among similar methods [22]. Our proposed method uses the refined depth maps, camera poses and sparsely sampled points on converged meshes of RNNSLAM to provide high-quality texture renderings of the structures.

2.2   Preliminaries 3D Gaussian Splatting

3D GS employs Gaussian distributions for scene representation and leverages GPU-accelerated rendering for efficient optimization [13], [15], [9]. 3D GS method models scenes using 3D Gaussians around the input RGB point cloud, represented by position $\boldsymbol{\mu}$, scale s, orientation as quaternion $\mathbf{q}$, spherical coefficients
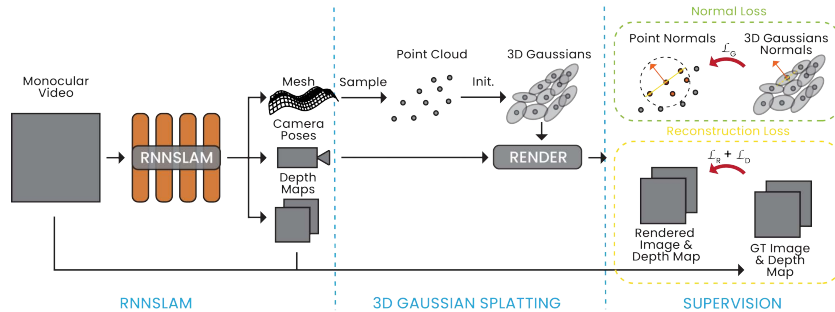
Fig. 2. Proposed Gaussian Pancakes' pipeline highlighting our contributions: A) RNNSLAM for mesh, camera poses, & depth maps; B) 3D GS initialization, C) Geometric & depth regularizations, distinguishing our approach from traditional 3D GS.

SH, and opacity σ. Each Gaussian has a covariance matrix Σ describing the skew and variance of the distribution in 3 dimensions, calculated using the scaling matrix S and rotation matrix R by

$$\Sigma = RSS^T R^T \tag{1}$$

A Gaussian can be described by the multivariate Gaussian probability density:

$$G(\mathbf{x}) = e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})} \tag{2}$$

These Gaussians are expressed in 3D space but are projected onto 2D for image rendering given by a viewing transformation W and the covariance matrix Σ. The covariance matrix in the camera coordinate system is given by $\Sigma' = JW\Sigma W^T J^T$, where J is the Jacobian matrix of the affine approximation of W. The splatting process involves differentiable point-based alpha blending. The final image pixel color is accumulated from contributions along a ray, considering the densities and transmittances. The color of each Gaussian is represented using spherical harmonic coefficients, capturing view-dependent effects [21]. The Gaussian attributes are optimized using the Adam optimizer [14] through a rendering process, combined with steps to adjust Gaussian density.

2.3   Proposed Gaussian Pancakes: Extending 3D GS for Endoscopy

Instead of relying on the sparse point cloud output from an SfM method, our approach utilizes RNNSLAM for pose estimation, depth maps, and RGB point cloud generation [23]. This substitution addresses SfM's limitations in low-variation environments where uniform textures can impede accurate feature matching and depth perception, thoroughly described in the RNNSLAM paper [17].

Depth regularization: Endoscopic images have minimal diversity in image directions, causing 3D GS to overfit. This results in artifacts where depth is misrepresented, representing images onto a 2D plane [10]. To mitigate this problem,

we introduce a depth loss based on the Huber loss. For the depth maps $\{D_i\}_{i=1}^M$, with M is the number of viewpoints, the depth loss $\mathcal{L}_D(i)$ is defined as:

$$\mathcal{L}_D(i) = \begin{cases} 0.5\Delta D_i^2, & \text{if } |\Delta D_i| < \delta \\ \delta(\Delta D_i - 0.5\delta), & \text{otherwise} \end{cases} \tag{3}$$

$\Delta D_i = |D_i - \hat{D}_i|$ denotes the loss at i-th frame, and $\hat{D}_i$, the rendered depth.

Geometric regularization: Despite the depth loss, challenges persist due to the lack of surface constraints. Artifacts which present as floating objects often emerge as the rendered view gets further away from original camera positions. To mitigate this, we incorporate a geometric loss that constrains the Gaussians in accordance with the surface's principal curvature, effectively "pancaking" the Gaussians along the surface. For a point cloud $P = \{\mathbf{p}_m\}$ where $\{\mathbf{p}_m\} = (x_m, y_m, z_m)$ represents the m-th point, we compute the normal vector $\mathbf{n}_p$ for each point $\mathbf{p}_m$ in P. This computation assumes the local neighborhood of any point can be approximated by a tangent plane inferred from its 10 nearest neighbors. We denote these neighbors as $\{\mathbf{n}_j\}_{j=1}^k$, where $k = 10$. Centering the nearest neighbors by $\mathbf{n}'_j = \mathbf{n}_j - \frac{1}{k}\sum_{j=1}^k \mathbf{n}_j$. The covariance of $\mathbf{n}'$, C, is calculated

$$C = \frac{1}{k-1}\sum_{j=1}^k \mathbf{n}'_j \mathbf{n}'^T_j \tag{4}$$

Eigen decomposition of C yields eigenvalues and eigenvectors $\mathbf{v}_i$. The normal vector, $\mathbf{n}_p$, for each point, $\mathbf{p}_m$, is the eigenvector corresponding to the smallest eigenvalue. The concatenated matrix of normal vectors and their points, N, is indexed using a GPU-accelerated structure, enhancing search efficiency [12].

Normal vectors for each 3D Gaussian are efficiently recalculated per iteration with a custom CUDA kernel for the dynamic scene. For each Gaussian's rotation matrix R, we find the axis of least spread by locating the minimum index of its scale vector s. This axis, selected from the standard Cartesian set e_x, e_y, e_z as $\mathbf{e}_{min}$, is the primary basis for computing the normal vector $\mathbf{n}_{GP} = R\mathbf{e}_{min}$. Using normals A from the nearest point cloud points and Gaussian normals B, we calculate geometric loss with cosine similarity as follows:

$$\mathcal{L}_G = 1 - \left| \frac{A \cdot B}{||A|| \, ||B||} \right| \tag{5}$$

Loss Function: Our model's total loss $\mathcal{L}$ merges image reconstruction, depth, and geometric losses, balanced by weights $\lambda_1, \lambda_2, \lambda_3$. This compact formulation:

$$\mathcal{L} = (1 - \lambda_1)\mathcal{L}_{Image} + \lambda_1\mathcal{L}_{D-SSIM} + \lambda_2\mathcal{L}_D + \lambda_3\mathcal{L}_G \tag{6}$$

with $\mathcal{L}_{Image}$ an L1 loss and $\mathcal{L}_{D-SSIM}$, a structural similarity term, between rendered and ground truth images. This loss optimizes trade-offs between visual fidelity, depth error, and geometric alignment with empirically set weights.

Table 1. Assessment of photometric, reconstruction errors, training and rendering times against top methods across Simulation, Phantom, In-vivo datasets.

Model	Dataset	PSNR ↑	SSIM ↑ MS-SSIM*	LPIPS ↓	Depth ↓ MSE	FPS↑	GPU ↓ min
ColonNerf[24]	Phantom	25.54	0.86*	0.40	-	-	-
NeRF[18]	Simulation	35.29	0.92	0.14	0.007	<2	167
	Phantom	32.10	0.81	0.39	4.263	<2	88
	In-Vivo	18.93	0.67	0.43	0.109	<2	89
REIM-NeRF [20]	Simulation	32.22	0.82	0.33	**0.001**	<2	170
	Phantom	31.66	0.78	0.22	**0.013**	<2	120
	In-Vivo	18.94	0.65	0.45	**0.006**	<2	97
Gaussian Pancakes (ours)	Simulation	**40.34**	**0.97**	**0.05**	0.007	**> 100**	**0.83**
	Phantom	**32.31**	**0.90/1.00∗**	**0.20**	0.498	**> 100**	**1.70**
	In-Vivo	**26.25**	**0.83**	**0.21**	0.156	**> 100**	**1.25**

Table 2. Ablation study on In-Vivo data.

Method	PSNR ↑	Depth ↑ SSIM	Depth ↓ MSE	GPU ↓ min
GS	26.116	0.444	0.158	1.211
GS+Depth	26.062	0.446	**0.145**	**1.206**
GS+Pancaking	26.211	**0.460**	0.154	1.329
GS+Pancaking+Depth(ours)	**26.248**	0.458	0.156	1.249

3   Experiments

Dataset and Evaluation Metrics: Our evaluation encompasses three distinct datasets: Simulation [28], In-Vivo [17], and Phantom [8]. The Simulation dataset, featuring sequences of the cecum, rectum, and sigmoid regions at a resolution of 320×240, is derived from software simulations in Unity using the graphics engine[28] for image rendering. Depth and pose references, however, are obtained from RNNSLAM. The In-Vivo dataset comprises real colonoscopy sequences each at 270×216 resolution from the RNNSLAM study[17]. We selected three distinct sequences (`"cecum_t4_b"`, `"desc_t4_a"`, and `"transt_t1_a"`) covering the cecum, descending colon and transcending sections from the public Phantom dataset (C3VD[8]), with 1350×1080 resolution colonoscopy videos. C3VD's simple camera trajectories caused large gaps in reconstructions due to insufficient scene coverage for RNNSLAM's mesh surface fusion step, limiting us to choose three datasets with satisfactory RNNSLAM performance for evaluation.

Frame data for each scene is divided into training and testing sets with an 8:1 ratio, following the approach recommended in prior research [5]. For a comprehensive quality assessment of comparative view synthesis, we use several common metrics: Peak Signal-To-Noise Ratio (PSNR), Learned Perceptual Image Patch Similarity (LPIPS), specifically the VGG loss [27], and Structural Simi-

larity Index Measure (SSIM)/Multi-Scale Structural Similarity Index Measure (MS-SSIM) - for direct comparison with ColonNeRF. We measure Depth Mean Squared Error (MSE), frames per second (FPS) to evaluate rendering speed, and GPU minutes for training time (aggregated minutes across all GPUs used).

Implementation: For a fair comparison with REIM-NeRF and NeRF we downsampled the resolution of the phantom dataset (C3VD) to 338×270, similar to what was tested in [20]. We keep the original resolution of simulation and in-vivo datasets, since they are already close to [20]. Our methodology initiates with RNNSLAM, rendering 10 FPS on an Intel i7 CPU and Nvidia GeForce GTX 1080 GPU. Gaussian Pancakes uses inferred depth maps for supervision, along with inferred camera poses and points for initial setup, across all training sequences. Subsequently, Gaussian Pancakes models trained following [13]'s hyperparameters, but are trained for 7,000 iterations instead of 30,000, densifying Gaussians until the 4,000th iteration. We set δ at 0.2 in Eq. 3, with hyperparameters in Eq. 6 $\lambda_1, \lambda_2,$ and λ_3 finely tuned to 0.2, 0.6, and 0.2 (empirically set), respectively, balancing photometric, reconstruction, and geometric results. Geometric loss integration commences after the first 1000 iterations for non-uniform scaling. All models are trained for 1-2 minutes on an NVIDIA RTX A6000.

4   Results and Discussion

In Table 1, Gaussian Pancakes demonstrates performance improvements in photometric errors PSNR, SSIM, and LPIPS, and achieves over 100 FPS rendering speeds and training times less than 2% that of the other models for all datasets. Qualitative comparison is in the supplementary video from minute 3:46-4:43 and supplementary Fig. A.1, where it can be observed that NeRF and REIM-NeRF exhibit poorer performance on the In-Vivo dataset, even when compared to their own results on the other datasets, likely due to the reliance on ground truth camera poses and, for REIM-NeRF, accurate depth maps. The heightened noise in In-Vivo dataset's camera poses and depth maps highlights Gaussian Pancakes' superior ability to manage data noise. However, Gaussian Pancakes exhibited a higher Depth MSE across all datasets compared to REIM-NeRF, a result of using inferred depth maps for both training and evaluation; making Depth MSE a measure of residual training error rather than a comparison with actual ground truth. Our method effectively handles data noise, prioritizing high-quality images and smooth reconstructions even with suboptimal depth references. A comparison with ColonNeRF for all datasets was not possible due to unavailable code.

Ablation Analysis: Table 2 shows the incremental benefits of depth and geometric regularizations added to the foundational 3D GS method, initialized with RNNSLAM outputs, to reduce artifacts in unseen viewpoints. However, standard metrics fall short in capturing these improvements, where test views are too close to training views to reveal the artifacts. Nevertheless, the quantitative data reveals nuanced improvements: introducing depth regularization marginally reduces Depth MSE, and normal regularization (GS+Pancaking) im-

proves Depth SSIM, suggesting better surface consistency. The combination of both (GS+Pancaking+Depth) achieves the best overall balance, as evidenced by modest gains in PSNR and Depth SSIM, alongside a negligible increase in processing time. Phantom and Simulation datasets show metric-based improvements that are more pronounced, further detailed in the Appendix Table A.1. For a comprehensive understanding, see the depth image comparison in supplementary Fig. A.3 and video from 4:43-5:40, which qualitatively shows artifact mitigation. Additionally, surfaces are visually smoother, seen in Fig. A.2.

5   Conclusion

Our evaluations across three varied datasets show that it achieves superior rendering quality, smoother reconstructions, fewer artifacts, and greatly reduces computational cost. Additionally, the method outputs explicit geometry, which sets it apart from the other leading methods with its potential for seamless integration into clinical practices. However, it's important to note that the method's success partly hinges on the quality of RNNSLAM's surface fusion, facing occasional challenges in less consistent environments. Moving forward, we plan to integrate SuperPoint [4] for better point cloud creation and to refine depth map precision. This direction aims to not only bolster our method's accuracy but also its applicability in a broader range of clinical scenarios.

References

1. Cancer (Feb 2022), https://www.who.int/news-room/fact-sheets/detail/cancer
2. Adamkiewicz, M., Chen, T., Caccavale, A., Gardner, R., Culbertson, P., Bohg, J., Schwager, M.: Vision-only robot navigation in a neural radiance world. IEEE Robotics and Automation Letters **7**(2), 4606–4613 (2022). https://doi.org/10.1109/LRA.2022.3150497
3. Alexa, M., Rusinkiewicz, S., Nehab, D., Shilane, P.: Stratified point sampling of 3d models. In: Proc. Eurographics Symp. on Point-Based Graphics. pp. 49–56 (2004)
4. Barbed, O.L., Chadebecq, F., Morlana, J., Montiel, J.M., Murillo, A.C.: Superpoint features in endoscopy. In: MICCAI Workshop on Imaging Systems for GI Endoscopy. pp. 45–55. Springer (2022)

5. Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P., Hedman, P.: Mip-nerf 360: Unbounded anti-aliased neural radiance fields (2022)
6. Barua, I., Vinsard, D.G., Jodal, H.C., Løberg, M., Kalager, M., Holme, Ø., Misawa, M., Bretthauer, M., Mori, Y.: Artificial intelligence for polyp detection during colonoscopy: a systematic review and meta-analysis. Endoscopy **53**(03), 277–284 (2020)
7. Batlle, V.M., Montiel, J.M., Fua, P., Tardós, J.D.: Lightneus: Neural surface reconstruction in endoscopy using illumination decline. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 502–512. Springer (2023)
8. 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 **90**, 102956 (2023)
9. Chen, G., Wang, W.: A survey on 3d gaussian splatting. arXiv preprint arXiv:2401.03890 (2024)
10. Chung, J., Oh, J., Lee, K.M.: Depth-regularized optimization for 3d gaussian splatting in few-shot images (2024)
11. Huang, Y., Cui, B., Bai, L., Guo, Z., Xu, M., Ren, H.: Endo-4dgs: Distilling depth ranking for endoscopic monocular scene reconstruction with 4d gaussian splatting. arXiv preprint arXiv:2401.16416 (2024)
12. Johnson, J., Douze, M., Jégou, H.: Billion-scale similarity search with gpus (2017)
13. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics (ToG) **42**(4), 1–14 (2023)
14. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization (2017)
15. Lassner, C., Zollhofer, M.: Pulsar: Efficient sphere-based neural rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1440–1449 (2021)
16. Liu, Y., Li, C., Yang, C., Yuan, Y.: Endogaussian: Gaussian splatting for deformable surgical scene reconstruction. arXiv preprint arXiv:2401.12561 (2024)
17. Ma, R., Wang, R., Zhang, Y., Pizer, S., McGill, S.K., Rosenman, J., Frahm, J.M.: Rnnslam: Reconstructing the 3d colon to visualize missing regions during a colonoscopy. Medical image analysis **72**, 102100 (2021)
18. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis (2020)
19. Pickhardt, P.J., Nugent, P.A., Mysliwiec, P.A., Choi, J.R., Schindler, W.R.: Location of adenomas missed by optical colonoscopy. Annals of internal medicine **141**(5), 352–359 (2004)
20. Psychogyios, D., Vasconcelos, F., Stoyanov, D.: Realistic endoscopic illumination modeling for nerf-based data generation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 535–544. Springer (2023)
21. Ramamoorthi, R., Hanrahan, P.: An efficient representation for irradiance environment maps. In: Proceedings of the 28th annual conference on Computer graphics and interactive techniques. pp. 497–500 (2001)
22. Rau, A., Bano, S., Jin, Y., Azagra, P., Morlana, J., Sanderson, E., Matuszewski, B.J., Lee, J.Y., Lee, D.J., Posner, E., Frank, N., Elangovan, V., Raviteja, S., Li, Z., Liu, J., Lalithkumar, S., Islam, M., Ren, H., Montiel, J.M.M., Stoyanov, D.: Simcol3d – 3d reconstruction during colonoscopy challenge (2023)
23. Schönberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2016)

24. Shi, Y., Lu, B., Liu, J.W., Li, M., Shou, M.Z.: Colonnerf: Neural radiance fields for high-fidelity long-sequence colonoscopy reconstruction. arXiv preprint arXiv:2312.02015 (2023)
25. Sun, X., Wang, F., Ma, Z., Su, H.: Dynamic surface reconstruction in robot-assisted minimally invasive surgery based on neural radiance fields. International Journal of Computer Assisted Radiology and Surgery pp. 1–12 (2023)
26. Wang, P., Liu, L., Liu, Y., Theobalt, C., Komura, T., Wang, W.: Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction (2023)
27. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: CVPR (2018)
28. Zhang, S., Zhao, L., Huang, S., Ye, M., Hao, Q.: A template-based 3d reconstruction of colon structures and textures from stereo colonoscopic images. IEEE Transactions on Medical Robotics and Bionics **3**(1), 85–95 (2020)
29. Zhu, L., Wang, Z., Cui, J., Jin, Z., Lin, G., Yu, L.: Endogs: Deformable endoscopic tissues reconstruction with gaussian splatting