Closing the Loop Between Multimodal SLAM and 3D Gaussian Splatting for Large-Scale Underwater Reconstruction

Daniel Yang*[1,2], Jungseok Hong*[1], Yogesh Girdhar[2] John J. Leonard[1]

***Abstract*— 3D Gaussian splatting is a powerful visual representation, providing high quality 3D scene reconstruction, but is crucially dependent on accurate estimation of camera poses, typically from computationally intensive process like structure from motion unsuitable for field robot applications. However in these domains, multimodal sensor data including depth, DVL, pressure, and monocular images is available and suitable for pose-graph optimization based SLAM methods that can estimate the vehicle's trajectory and provide an estimate of our needed camera poses. We initialize a 3DGS model using noisy poses and images observing a known landmark. Incrementally, we refine these camera poses and use our scene representation to track new image observations and continue reconstructing the scene. As our scene is further optimized, we can refine previously observed images and feed such poses back into the pose-graph. We show a COLMAP-free 3D reconstruction of a 20 m by 20 m underwater reef with complex geometry as well as more accurate global pose estimation of our underwater vehicle during its 800 m trajectory surveying the reef.**

I. Introduction

Recent advances in autonomous underwater vehicles (AUVs) and underwater mapping have enabled the mapping of challenging underwater environments. However, essential scientific applications, such as coral reef monitoring and large-scale spatiotemporal environmental surveys, are still constrained. These tasks require dense, geometrically consistent scene representations over long trajectories, where the quality of dense reconstruction depends on the input images and the accuracy of camera pose estimation. Achieving such reconstructions for underwater scenes remains challenging due to degraded visual conditions (e.g., light attenuation, turbidity, color distortion, and low-texture scenes) and limited sensing capabilities, resulting in persistent inaccuracies in pose estimation.

To estimate poses in batch, structure-from-motion and multi-view stereo methods have been used, but they require substantial computational resources. Visual simultaneous localization and mapping (SLAM) methods can be used to estimate poses incrementally and build maps in real time, but they are fundamentally unreliable underwater due to the challenges mentioned above. To address this issue, state-of-the-art underwater SLAM systems integrate multimodal sensors, combining inertial, velocity, and depth measurements with visual and/or acoustic cues in pose-graph optimization frameworks. While these systems effectively estimate vehicle trajectories over long distances, the resulting maps are typically sparse and optimized for navigation rather than dense scene representation. Consequently, existing systems are generally inadequate for downstream scientific analysis that requires detailed information about the target environments. This fundamental mismatch between existing systems and scalable, dense reconstruction remains a key bottleneck for using maps across a wide range of tasks.

Recently, 3D Gaussian Splatting (3DGS) [1] has emerged as an efficient and differentiable representation that produces high-fidelity, dense reconstructions while allowing for pose refinement via gradient-based optimization. This advancement has motivated the development of 3DGS-based SLAM systems [2]–[5]. However, these systems are primarily tested in terrestrial environments and rely on reliable visual characteristics or RGB-D sensing, assumptions that do not hold in underwater contexts where sensor measurements are noisier.

We address this gap by tightly coupling multimodal pose-graph SLAM with incremental updates to a 3DGS model. Rather than treating dense reconstruction as a post-processing step or relying on batch-based structure-from-motion, we propose a closed loop between pose estimation from SLAM and scene representation. Starting from regions of low pose uncertainty near a known landmark, we incrementally expand the dense scene representation. As the 3DGS model improves, it is used to refine camera poses via differentiable rendering, and these refined poses are added to the pose graph to improve global trajectory estimation.

This bidirectional refinement allows pose estimation and dense reconstruction to mutually improve each other over extended trajectories, even when initial pose estimates are noisy and input images are distorted. In contrast to previous 3DGS-based SLAM systems, our approach explicitly leverages the pattern of underwater surveys and the complementary strengths of multimodal SLAM and differentiable scene representations. The resulting pipeline is bundle-adjustment-free, incremental, and capable of dense reconstruction at scales relevant to real-world underwater surveys, spanning hundreds of meters while simultaneously improving global pose accuracy.

Our contributions are as follows:

1) Framework for integrating 3D Gaussian Splatting with pose-graph SLAM to achieve both increased 3D reconstruction quality and less error in pose estimation, suitable for field robotics applications.
2) less time to build 3dgs model using slam pose estimation
3) real-world coral reef surveys spanning an 800 m vehi-

[1]D. Yang, J. Hong, J. Leonard are with the Computer Science and Artificial Intelligence Laboratory (CSAIL) at the Massachusetts Institute of Technology (MIT), 32 Vassar St, Cambridge, MA 02139, USA.

[2]D. Yang and Y. Girdhar are with Applied Ocean Physics and Engineering Department at the Woods Hole Oceanographic Institute (WHOI)

cle trajectory and reconstructing a 20 m × 20 m reef area with complex geometry.
4) finetune depth estimation model with synthetic data?

II. Related works

A. Bundle Adjustment-free 3D reconstruction

Recent feed-forward 3D models reduce reliance on classical SfM by predicting point maps directly from pairs of images, from which camera parameters can be estimated. DUSt3R [6] introduced a framework for dense per-pixel 3D pointmap regression, enabling end-to-end reconstruction from arbitrary image pairs without camera calibrations. Many works build upon DUSt3R to improve dense feature matching [7] and introduce strategies for handling multiple views [8] or a stream of images [9], [10]. Other works like VGGT [11] utilize feed-forward transformers to simultaneously attend to and jointly reason over all input views. These models rely on learned geometric priors and dense correspondences that can initialize or regularize poses. However, when texture is scarce and viewpoints are suboptimal (e.g. limited field of view, minimal overlap) as is common in underwater scenes, they tend to struggle.

B. Underwater SLAM

Underwater SLAM is substantially more challenging than terrestrial or aerial SLAM due to degraded visual conditions, limited sensing bandwidth, and the absence of GPS [12], [13]. Optical imagery suffers from attenuation, color distortion, turbidity, and low-texture scenes, causing vision-only and visual-inertial SLAM systems to fail or drift even under moderate underwater conditions [14]. To achieve robust long-horizon navigation, underwater SLAM systems therefore rely on multi-modal sensor fusion, commonly integrating IMU, DVL, and depth measurements within factor-graph or pose-graph optimization frameworks [15]. Visual or acoustic sensing is typically used to detect features and provide relative pose constraints. Underwater SLAM systems such as [16] focus on robust trajectory estimation using sparse map representations (e.g., keyframes). In contrast, recent works including [17], [18] yield dense maps, but typically evaluate on relatively short trajectories rather than long-horizon mapping.

C. Dense SLAM

Dense underwater reconstruction has therefore been dominated by SfM and MVS pipelines, COLMAP [19] and its commercial equivalent Metashape [20]. These pipelines provide high-quality models but require computationally intensive bundle adjustment and are unsuitable for incremental field deployment. Camera pose estimation remains a key bottleneck for scaling dense reconstruction to large underwater environments. One recent method [21] eschews these pipelines for learning based ego-motion estimation from images [22] to build underwater scene maps, but relies on masking out most of the scene where water degradation effects are visible and is limited in the fidelity of its output.

Recent advances in dense scene representations, particularly radiance field methods like NeRF [23] and 3D Gaussian Splatting (3DGS) [24], offer new opportunities to bridge the gap between sparse SLAM and dense reconstruction. 3D Gaussian Splatting (3DGS) [24] has emerged as a computationally efficient explicit representation capable of producing high-fidelity dense geometry and appearance while supporting differentiable optimization. Several recent SLAM systems [2]–[5] incorporate 3D Gaussian representations to jointly optimize camera poses and dense scene models, demonstrating real-time or near-real-time photo-realistic reconstruction in terrestrial environments. However, these methods typically assume strong visual observability, controlled illumination, or RGB-D sensing, and are not designed to operate under the degraded visual conditions and multi-modal sensing constraints characteristic of underwater domains.

In contrast, this work bridges sparse multimodal underwater SLAM and dense 3D Gaussian Splatting by closing the loop between pose-graph optimization and dense scene reconstruction, enabling bundle-adjustment-free large-scale underwater reconstruction while achieving reasonable global vehicle pose estimation.

III. Method

A. Problem Formulation

We aim to construct a dense 3D representation model of an underwater environment by estimating robot poses by fusing multimodal sensor data. We assumed that an underwater environment is surveyed by a robot navigating the environment in a flower-petal pattern with a fixed, known landmark at the center. Let $\mathcal{X} = \{\mathbf{X}_t\}_{i=1}^{N}$ denote the sequence of robot poses, where each pose $\mathbf{X}_t \in SE(3)$. The sensor measurements $\mathcal{Z}$ include monocular images $\mathcal{I} = \{I_t\}_{t=t_0}$ from a downward-facing calibrated camera, linear velocities $\mathcal{V} = \{v_t\}_{t=t_0}^{t_N}$, $v_t \in \mathbb{R}^3$ from a DVL, angular velocity $\Omega = \{\omega_t\}_{t=t_0}^{t_N}$, $\omega_t \in \mathbb{R}^3$ and linear acceleration $\mathcal{A} = \{a_t\}_{t=t_0}^{t_N}$, $a_t \in \mathbb{R}^3$ from an IMU, and depth measurements $\mathcal{D} = \{d_t\}_{t=t_0}^{t_N}$, $d_t \in \mathbb{R}$ from a pressure sensor. Note that depth here refers to distance from the sea surface, not depth from the camera sensor.

B. Pose-Graph Optimization for Pose Estimation

We formulate the pose estimation problem as a factor graph optimization problem, where we estimate the robot's state $\mathcal{X}$ and landmark $\mathcal{L}$ via Maximum a Posteriori (MAP) inference given sensor measurements $\mathcal{Z}$.

$$\hat{\mathcal{X}}, \hat{\mathcal{L}} = \underset{\mathcal{X}, \mathcal{L} \in SE(3)}{\arg\max}\ p(\mathcal{X}, \mathcal{L} \mid \mathcal{Z}). \tag{1}$$

To ensure scalability to long-duration missions, which are typically required in underwater survey missions, we first estimate odometry using an Extended Kalman Filter (EKF) rather than creating separate factors for each sensor measurement. Prior to odometry estimation, we apply an IMU complementary filter [25] to preprocess raw IMU measurements, which are typically affected by noise and sensor bias. This filter employs quaternion decomposition to yield

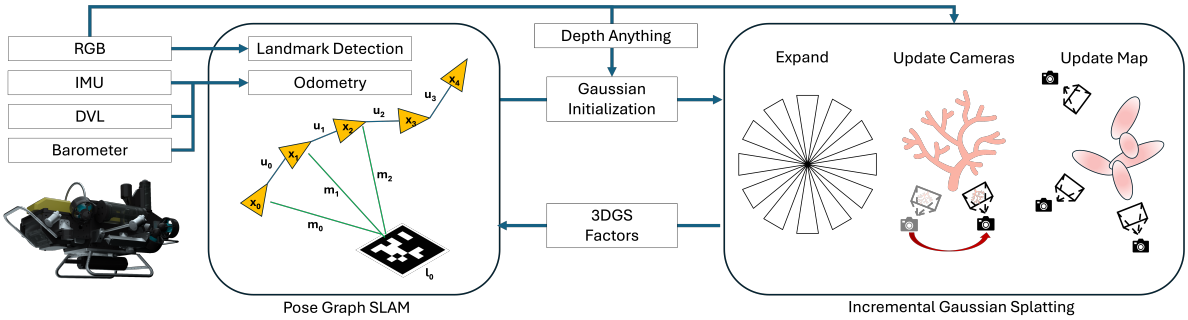

Fig. 1: overview figure

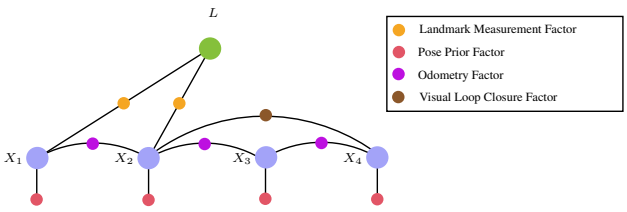

Fig. 2: Information obtained from objects and odometry is used for building a factor graph (X_i: vehicle states; L_i: landmark states; $\theta_{Li}, \phi_{Li}, z_{Li}$: bearing, elevation, and ranges to objects; z_v, ϕ_v, ψ_v: depth, pitch, and roll of the vehicle; x_v, y_v, θ_v: location and yaw of the vehicle).

a stable orientation estimate in real-time. The filtered IMU data, combined with depth and DVL velocity measurements, are then fused using an EKF [26] to obtain odometry data.

To build a factor graph, we incorporate prior knowledge of a static landmark located at the center of the survey site. We add both odometry and landmark measurement factors into the graph, each associated with robust noise models (*e.g.,* Huber) to handle potential outliers. The final pose trajectory is estimated by solving the factor graph using Levenberg–Marquardt (LM) optimization.

C. Incremental 3D Gaussian Splatting with Uncertainty-Aware Pose Refinement

We utilize a 3D Gaussian Splatting (3DGS) representation within our framework, parameterizing the scene with a set of isotropic 3D Gaussians and using zero order spherical harmonics. Thus, each 3D Gaussian has a set of 8 parameters, consisting of the mean $\boldsymbol{\mu} \in \mathbb{R}^3$, scale $\boldsymbol{S} \in \mathbb{R}^1$, opacity $\boldsymbol{o} \in \mathbb{R}^1$, and color $\boldsymbol{c} \in \mathbb{R}^3$. This reduced set of parameters requires less space and simplifies the optimization problem. Typically, a 3DGS model is trained given a dataset of RGB images, corresponding poses, and a point cloud to initialize the 3DGS model, the latter two of which are typically obtained from structure-from-motion (e.g. COLMAP [19]). All views of the model are jointly optimized and available at the start of optimization.

In contrast to this paradigm of precise camera poses and an accurate, sparse point cloud to initialize the full scene, our underwater field environment context has much noisier, sparse estimated data. Camera poses are estimated by fusing multimodal sensor data while a monocular RGB sensor is used with a monocular depth estimator (e.g. DepthAnythingV2 [27] to generate point clouds to initialize 3D Gaussians. Similarly, our pose-graph incorporating a fixed static landmark does not have error uniformly distributed across all poses, with poses near the landmark having much lower uncertainty than poses further away from the landmark.

Thus, we propose incrementally building the scene representation, starting from the region of least noise, near the landmark, and gradually introducing and optimizing new observations on the frontier of what has already been incorporated into the scene. In other words, we are radially outwardly optimizing the scene representation. Notably, before incorporating new noisy observations we can refine the pose estimates by fitting the observation against the model optimized thus far. Given an 3DGS model of the partial scene, a new image observation, and a noisy estimate of the new image pose, we can refine the pose by minimizing reconstruction error and propagating the gradients back to the pose as implemented in libraries like `gsplat` [28]. This requires visual overlap between the rendered image at the pre-refinement pose and can fail if the pre-refinement pose is too far away from the actual pose. As the scene representation is gradually refined, we interleave this pose refinement process, hoping that improvements in scene representation lead to more accurate pose estimation.

D. Iterative, Global Pose Refinement from 3D Gaussian Splats

Given a 3DGS model that has refined a subset of camera poses where the scene representation accurately captures the scene, we can incorporate these refined pose estimates as factors into our factor graph. These factors not only improve pose estimation locally around the camera frame but also much further away from the landmark, as error is mitigated before it causes the pose to drift significantly. These further away improved poses can then be used with optimization of the 3DGS model and enable reconstruction of areas where previously, there was too much error to be corrected with the pose refinement process.

IV. Experiments

A. Dataset Collection

We evaluate our method on coral reef benthic survey data collected at two sites in the US Virgin Islands. An AprilTag [29] is placed at the center of the reef, and we

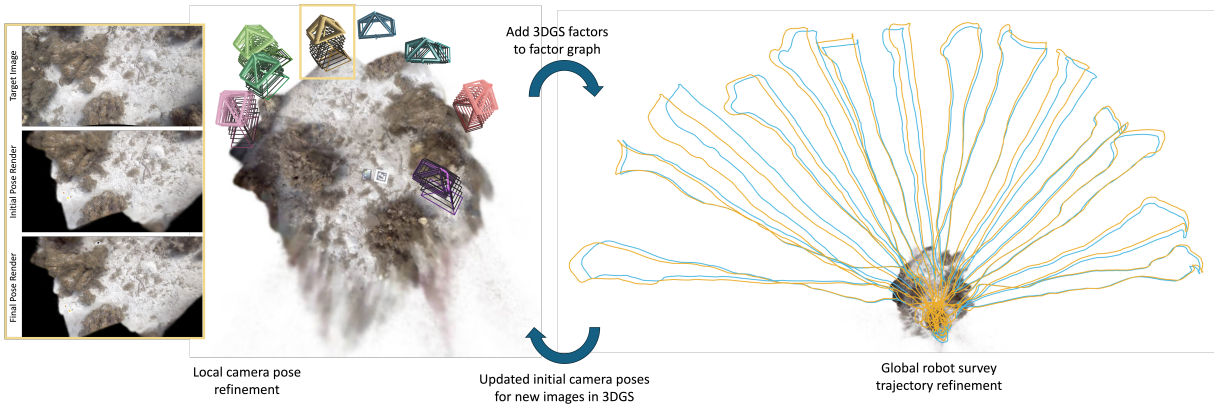

Fig. 3: overview figure

utilize CUREE [30], a flexible low-cost underwater vehicle platform, to collect the dataset. CUREE performs a flower-petal shaped survey, where each petal is 10 m long and has a sensor suite including a downward facing camera recording benthic imagery, a Waterlinked DVL A50, an ICM-20602 IMU from a BlueRobotics Navigator, and a BlueRobotics Bar30 pressure sensor.

B. Evaluation Metrics

C. Comparative Baselines

V. Results

A. Pose Estimation Accuracy

B. 3D Reconstruction Quality

C. Ablations

D. Running time

VI. Conclusions

References

[1] B. Kerbl, G. Kopanas, T. Leimkuehler, and G. Drettakis, "3D Gaussian Splatting for Real-Time Radiance Field Rendering," *ACM Transactions on Graphics*, vol. 42, no. 4, pp. 1–14, Aug. 2023. [Online]. Available: https://dl.acm.org/doi/10.1145/3592433

[2] N. Keetha, J. Karhade, K. M. Jatavallabhula, G. Yang, S. Scherer, D. Ramanan, and J. Luiten, "SplaTAM: Splat, Track & Map 3D Gaussians for Dense RGB-D SLAM," Apr. 2024, arXiv:2312.02126 [cs]. [Online]. Available: http://arxiv.org/abs/2312.02126

[3] J. Zheng, Z. Zhu, V. Bieri, M. Pollefeys, S. Peng, and I. Armeni, "WildGS-SLAM: Monocular Gaussian Splatting SLAM in Dynamic Environments," Apr. 2025, arXiv:2504.03886 [cs]. [Online]. Available: http://arxiv.org/abs/2504.03886

[4] E. Sandström, K. Tateno, M. Oechsle, M. Niemeyer, L. V. Gool, M. R. Oswald, and F. Tombari, "Splat-SLAM: Globally Optimized RGB-only SLAM with 3D Gaussians," May 2024, arXiv:2405.16544 [cs]. [Online]. Available: http://arxiv.org/abs/2405.16544

[5] H. Matsuki, R. Murai, P. H. J. Kelly, and A. J. Davison, "Gaussian Splatting SLAM," Apr. 2024, arXiv:2312.06741. [Online]. Available: http://arxiv.org/abs/2312.06741

[6] S. Wang, V. Leroy, Y. Cabon, B. Chidlovskii, and J. Revaud, "DUSt3R: Geometric 3D Vision Made Easy," Dec. 2024, arXiv:2312.14132 [cs]. [Online]. Available: http://arxiv.org/abs/2312.14132

[7] Y. Cabon, L. Stoffl, L. Antsfeld, G. Csurka, B. Chidlovskii, J. Revaud, and V. Leroy, "Must3r: Multi-view network for stereo 3d reconstruction," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 1050–1060.

[8] B. Duisterhof, L. Zust, P. Weinzaepfel, V. Leroy, Y. Cabon, and J. Revaud, "MASt3R-SfM: a Fully-Integrated Solution for Unconstrained Structure-from-Motion," Sep. 2024, arXiv:2409.19152 [cs]. [Online]. Available: http://arxiv.org/abs/2409.19152

[9] H. Wang and L. Agapito, "3D Reconstruction with Spatial Memory," Aug. 2024, arXiv:2408.16061 [cs]. [Online]. Available: http://arxiv.org/abs/2408.16061

[10] Q. Wang, Y. Zhang, A. Holynski, A. A. Efros, and A. Kanazawa, "Continuous 3D Perception Model with Persistent State," Jan. 2025, arXiv:2501.12387 [cs]. [Online]. Available: http://arxiv.org/abs/2501.12387

[11] J. Wang, M. Chen, N. Karaev, A. Vedaldi, C. Rupprecht, and D. Novotny, "VGGT: Visual Geometry Grounded Transformer," Mar. 2025, arXiv:2503.11651 [cs]. [Online]. Available: http://arxiv.org/abs/2503.11651

[12] J. C. Kinsey, R. M. Eustice, and L. L. Whitcomb, "A survey of underwater vehicle navigation: Recent advances and new challenges," in *IFAC conference of manoeuvering and control of marine craft*, vol. 88. Lisbon, 2006, pp. 1–12.

[13] J. J. Leonard and A. Bahr, "Autonomous underwater vehicle navigation," *Springer handbook of ocean engineering*, pp. 341–358.

[14] B. Joshi, S. Rahman, M. Kalaitzakis, B. Cain, J. Johnson, M. Xanthidis, N. Karapetyan, A. Hernandez, A. Q. Li, N. Vitzilaios *et al.*, "Experimental comparison of open source visual-inertial-based state estimation algorithms in the underwater domain," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 7227–7233.

[15] F. F. R. Merveille, B. Jia, Z. Xu, and B. Fred, "Advancements in sensor fusion for underwater slam: A review on enhanced navigation and environmental perception," *Sensors (Basel, Switzerland)*, vol. 24, no. 23, p. 7490, 2024.

[16] S. Rahman, A. Q. Li, and I. Rekleitis, "Svin2: An underwater slam system using sonar, visual, inertial, and depth sensor," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 1861–1868.

[17] W. Wang, B. Joshi, N. Burgdorfer, K. Batsosc, A. Q. Lid, P. Mordohaia, and I. Rekleitisb, "Real-time dense 3d mapping of underwater environments," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 5184–5191.

[18] S. Xu, K. Zhang, and S. Wang, "Aqua-slam: Tightly-coupled underwater acoustic-visual-inertial slam with sensor calibration," *IEEE Transactions on Robotics*, 2025.

[19] J. L. Schönberger and J.-M. Frahm, "Structure-from-motion revisited," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[20] "Agisoft Metashape: Professional Edition." [Online]. Available: https://www.agisoft.com/features/professional-edition/

[21] J. Sauder, G. Banc-Prandi, A. Meibom, and D. Tuia, "Scalable semantic 3D mapping of coral reefs with deep learning," *Methods in Ecology and Evolution*, vol. 15, no. 5, pp. 916–934, 2024, _eprint: https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.14307. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.14307

[22] J. Bian, Z. Li, N. Wang, H. Zhan, C. Shen, M.-M. Cheng, and I. Reid,

Method	Tektite				Yawzi			
	PSNR	LPIPS	SSIM	RMSE (m)	PSNR	LPIPS	SSIM	RMSE (m)
Metashape+sfm+batch	21.9285	0.5558	0.7499	0.000000	21.1643	0.5341	0.8189	0.000000
Metashape+md+batch	21.5406	0.5517	0.7457	0.000000	20.8738	0.5282	0.8255	0.000000
GTSAM+md+batch	18.6762	0.4971	0.8225	0.333700	18.2283	0.4871	0.8900	0.283686
Metashape+sfm+inc	23.4572	0.6080	0.6257	0.204176	23.2464	0.6237	0.6155	0.254395
Metashape+md+inc	22.9804	0.5958	0.6536	0.227291	22.7186	0.6081	0.6443	0.263354
GTSAM+md+inc	22.7708	0.5814	0.7193	0.364013	21.9614	0.5750	0.7399	0.323584
Ours (GTSAM+md+inc+reopt)	23.1433	0.5900	0.6949	0.143776	22.1597	0.5779	0.7289	0.255172

TABLE I: **Evaluation metrics across different configurations.** sfm means Gaussians in the splat are initialized from the point cloud byproduct of structure from motion, while md means Gaussians are initialized from monodepth estimator outputs. Batch means the splat is batch optimized while inc means the splat is incrementally optimized following our method. Higher is better for PSNR and SSIM while lower is better for LPIPS and RMSE (e.g. PSNR ↑, SSIM ↑, LPIPS ↓, RMSE ↓).

TABLE II: Evaluation metrics across different configurations.

Metric	O+L	O+L+LC	O+L+LC+GS1	O+L+LC+GS2
max	1.220883	1.191861	1.003618	
mean	0.261338	0.258766	0.210663	
median	0.221612	0.219229	0.178597	
min	0.015266	0.002618	0.008797	
rmse	0.312291	0.308885	0.244823	

"Unsupervised Scale-consistent Depth and Ego-motion Learning from Monocular Video," in *Advances in Neural Information Processing Systems*, vol. 32. Curran Associates, Inc., 2019. [Online]. Available: https://papers.nips.cc/paper$_files/paper/2019/hash/6364d3f0f495b6ab9dcf8d3b5c6e0b01-Abstract.html$

[23] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," in *ECCV*, 2020.

[24] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3d gaussian splatting for real-time radiance field rendering," *ACM Transactions on Graphics*, vol. 42, no. 4, July 2023. [Online]. Available: https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/

[25] R. G. Valenti, I. Dryanovski, and J. Xiao, "Keeping a good attitude: A quaternion-based orientation filter for imus and margs," *Sensors*, vol. 15, no. 8, pp. 19 302–19 330, 2015.

[26] T. Moore and D. Stouch, "A generalized extended kalman filter implementation for the robot operating system," in *Intelligent Autonomous Systems 13: Proceedings of the 13th International Conference IAS-13*. Springer, 2016, pp. 335–348.

[27] L. Yang, B. Kang, Z. Huang, Z. Zhao, X. Xu, J. Feng, and H. Zhao, "Depth Anything V2," Oct. 2024, arXiv:2406.09414 [cs]. [Online]. Available: http://arxiv.org/abs/2406.09414

[28] V. Ye, R. Li, J. Kerr, M. Turkulainen, B. Yi, Z. Pan, O. Seiskari, J. Ye, J. Hu, M. Tancik *et al.*, "gsplat: An open-source library for gaussian splatting," *Journal of Machine Learning Research*, vol. 26, no. 34, pp. 1–17, 2025.

[29] E. Olson, "AprilTag: A robust and flexible visual fiducial system," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, May 2011, pp. 3400–3407.

[30] Y. Girdhar, N. McGuire, L. Cai, S. Jamieson, S. McCammon, B. Claus, J. E. S. Soucie, J. E. Todd, and T. A. Mooney, "CUREE: A Curious Underwater Robot for Ecosystem Exploration," Apr. 2023, arXiv:2303.03943 [cs]. [Online]. Available: http://arxiv.org/abs/2303.03943