Bridging the Domain Gap: Enhancing Underwater Laser Stripe Segmentation with Synthetic Data

Anonymous CVPR submission

Paper ID *****

Abstract

Underwater 3D mapping using low-cost scanning systems relies on accurate laser stripes segmentation. However, the scarcity of annotated data and the variability of underwater environments limit the model's generalization and scalability. To address this issue, we introduce a synthetic dataset specifically designed for laser stripe segmentation. Created using a custom laser scanner module integrated into Blender and the Infinigen procedural generator. The dataset contains over 1,200 high-resolution images across 23 diverse terrains, each with ground truth. We evaluate the impact of synthetic data using a segmentation network trained under different field-to-synthetic data ratios. Our results show that augmenting field datasets with synthetic images significantly improves performance on unseen domains—achieving up to 10% higher recall and 7% higher precision on deep-sea imagery from the Salas y Gómez Ridge, a location with different lighting, seafloor composition, and depth. Our findings highlights the value of synthetic data for domain diversity, reducing annotation costs and enhancing model generalization, supporting broader and more robust deployment of underwater laser mapping systems.

1. Introduction

High-resolution seafloor mapping is essential for understanding ocean dynamics and the ecosystems that inhabit the seabed [13]. This task typically involves deploying remotely operated vehicles (ROVs) or autonomous underwater vehicles (AUVs) equipped with sensors to capture detailed information about underwater terrains. Commonly used sensors include single-beam and multi-beam echo sounders or underwater LIDARs. However, these technologies are often bulky, expensive, and energy-intensive.

An emerging alternative for underwater 3D reconstruction is the use of optical-based structured light systems, specifically one-stripe laser scanners. These systems are characterized by their high precision, energy efficiency, and relatively low cost [1, 5]. The method consists of projecting a laser stripe onto the seafloor and capturing its deformation using a monocular camera. With proper system calibration, 3D coordinates of the seabed surface can be computed from the observed laser stripe. Figure 1 illustrates the scanning process.

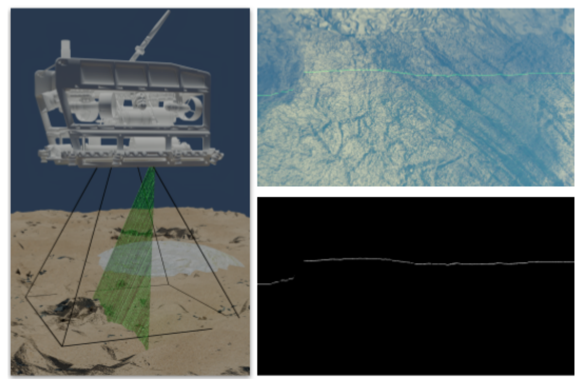


Figure 1. The image shows (left) a simulation of MBARI's MiniROV performing laser scanning, (top right) an simulated image captured during the survey, and (bottom right) the corresponding ground truth mask

A crucial part of the process is the segmentation of the laser stripe in the captured image. Existing methods for underwater laser stripe segmentation can be broadly divided into two categories: classical approaches based on intensity thresholding [7, 12] or color cues [1, 3]; and more recent approaches using deep learning. Among these, a novel transformer-based architecture has been proposed [4], significantly outperforming classical methods—achieving up to 71% improvement in accuracy under illuminated conditions.

Despite these advancements, the performance of deep learning models—like in other domains—is constrained by the size and diversity of the training datasets. In the case of underwater environments, this limitation is even more severe due to the well-known lack of publicly available annotated imagery [6]. This scarcity is largely due to the high operational cost of collecting field data, as well as the time-consuming and labor-intensive process of manually labeling high-resolution images.

Moreover, underwater imagery exhibits high variability caused by differences in terrain composition, water clarity, laser and camera characteristics, and lighting conditions—often influenced by depth and the vehicle's illumination system. For instance, a model trained on field data collected in Monterey Bay, California, and controlled experiments in a test tank performed well in those domains. However, its segmentation accuracy significantly degraded when applied to data from Easter Island, Chile, where the seafloor composition, lighting, and ecological context differ markedly. This lack of generalization restricts the usability and scalability of the method across diverse marine regions.

To address these limitations and to complement scarce annotated field data, we propose a synthetic dataset designed to improve the robustness and adaptability of underwater laser segmentation systems.

Our main contributions are:

(i) A novel *Synthetic Underwater Laser Scanner Module*, integrated with Blender and Infinigen, enabling rapid generation of synthetic underwater scanning data;

(ii) A dataset of over 1,200 simulated images from 23 diverse underwater terrains, each paired with pixel-level ground truth masks;

(iii) A comprehensive evaluation quantifying the impact of synthetic data on segmentation performance in underwater scenarios.

The remainder of this paper is organized as follows: Section 2 describes the synthetic data generation pipeline; Section 3 presents the experimental setup; Section 4 discusses the results; and Section 5 concludes the paper.

2. Synthetic Underwater Laser Scanner Module

The system module was develop in Blender 3.6 [2] and implemented in Python 3.10.

2.1. System

The synthetic data generation pipeline is designed to simulate the core components of an underwater laser scanning system: a monocular camera, a laser stripe, and an illumination setup.

The camera is encapsulated within a rigid camera rig, with a resolution of 1920×1080. It is spawned at a random position above the terrain at altitudes ranging between 0.5 and 3 meters, always oriented downward (along the negative Z-axis). The camera's focal length and sensor size are user-configurable, allowing realistic replication of specific optical setups.

The laser is modeled using a `SPOT` light object in Blender. It is rigidly mounted relative to the camera with a configurable baseline (between 0.4 m and 0.6 m). The laser orientation can be set at a fixed incidence angle or dynamically adjusted based on the camera's altitude to simulate realistic scanning geometries. The beam intensity is randomly sampled between 500 and 1000 lumens to emulate realistic laser output variation. The spot size is constrained to control the spread of the projected laser stripe across the scene - between $30°$ and $45°$.

To render a physically plausible thin laser stripe, we implement a custom node-based shader in Blender using Python. This setup isolates the Y-component of the light's surface normal using a `Separate XYZ` node, followed by a `Math > Compare` operation with a narrow threshold (e.g., 1×10^{-4}) to ensure a sharp-edge beam with minimal soft falloff, effectively restricting emission to a single narrow plane.

Two lighting configurations were implemented to emulate different AUV/ROV setups:

- A dual-spotlight system with left and right lights mounted on the center camera rig.
- A quad-light system with four spotlights (forward-left, forward-right, rear-left, rear-right).

Light intensities are randomly selected in the range of 500–1000 W to simulate operational variability and environmental lighting conditions.

2.2. Terrain Generation

The simulated images were generated using two complementary approaches for underwater terrain formation: field 3D models and procedurally generated environments.

2.2.1. MBARI's 3D Survey Models

We used pre-existing 3D models from previous MBARI (Monterey Bay Aquarium Research Institute) seafloor surveys, publicly available on Sketchfab [9]. Three distinct models were selected, each representing geological features observed in the Monterey Bay area. These meshes were imported into Blender and placed over a flat sandy plane to provide environmental continuity and context. Additional props such as rocks were added to increase scene complexity.

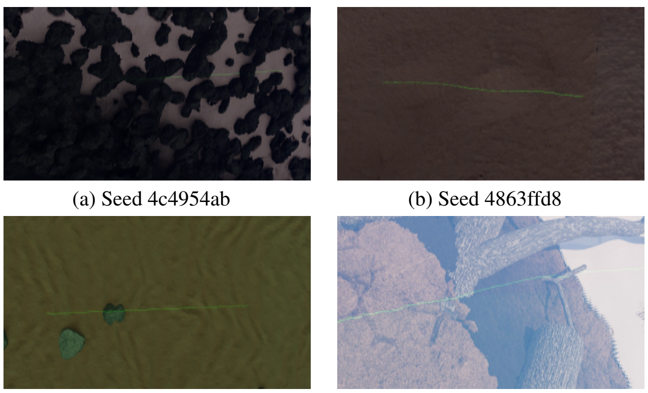
2.2.2. Procedural Generation with Infinigen

To diversify terrain geometry, we used Infinigen [11], a photorealistic procedural scene generator built on Blender. We extended Infinigen's source code to integrate our custom laser scanner module, enabling automatic placement of the camera-laser rig within procedurally generated scenes. A total of 20 unique terrains were generated using the coral reef configuration, each derived from a different random seed to ensure variation in structural features (Figure 2).

Underwater Visual Effects

Two common optical effects in underwater environments are *absorption* and *scattering* [8]. To simulate these effects, we applied a volumetric absorption shader to the background and surrounding water using Blender's `Principled Volume` node. The volume was tinted with cool blue to approximate the spectral absorption of light underwater. The density of the volume was randomly varied between 0.08 and 0.2 to simulate different water clarity conditions.

To further enhance visual realism, a scattering component was added to the volumetric shader to emulate the diffusion of light caused by suspended particles in the water column.



(a) Seed 4c4954ab (b) Seed 4863ffd8 (c) Seed 161a2dbc (d) Seed 68067fa

Figure 2. Samples from Infinigen-based terrains

2.3. Data Collection

To simulate realistic AUV/ROV surveys, a data collection script was developed to capture multiple frames across each terrain. A standard survey grid of $5\,\mathrm{m} \times 5\,\mathrm{m}$ was defined, and a step size parameter controlled the spatial resolution of the survey.

Each survey begins with the system positioned at the bottom-left corner of the grid. It then follows a *mowing-the-lawn* trajectory, commonly used in underwater surveys. At each step, one image is captured, with Gaussian noise applied independently to the position in all three spatial axes to simulate underwater vehicle motion and localization uncertainty.

For each scene configuration, two renders are produced per frame: one with the system lights and laser active (RGB image) and one with all non-laser lights disabled, isolating the laser contribution. The latter is used to generate the ground truth binary segmentation mask for the laser stripe.

In total, 23 unique survey scenes were generated, resulting in 1277 images with corresponding ground truth masks.

3. Evaluation Methodology

3.1. Dataset Composition

The simulated dataset consists of 1,277 images generated across 24 distinct underwater survey scenarios. To evaluate the impact of synthetic data on segmentation performance, we also incorporated 564 manually annotated field images. These field annotations were created using Roboflow.

The field dataset is divided as follows:

- **Monterey Canyon, California:** 207 images collected during field deployments using an ROV equipped with a laser scanner.
- **MBARI's Test Tank:** 187 images captured under controlled conditions with varying lighting configurations.

These two subsets form the **training set**, representing environments with similar sensor setups and moderate lighting variability.

To assess the generalization ability of the segmentation model, a separate **test set** was curated using 156 images collected during a scientific expedition to the Salas y Gómez Ridge in Chile. These images were taken at greater depth (approximately 500 meters deeper than Monterey deployments) and with a different vehicle platform featuring a distinct lighting system. This evaluation setup enables us to rigorously test the model's ability to adapt to previously unseen underwater environments.

Figure 3 shows representative examples of the field annotated images used in both the training and test sets.

3.2. Network Implementation Details

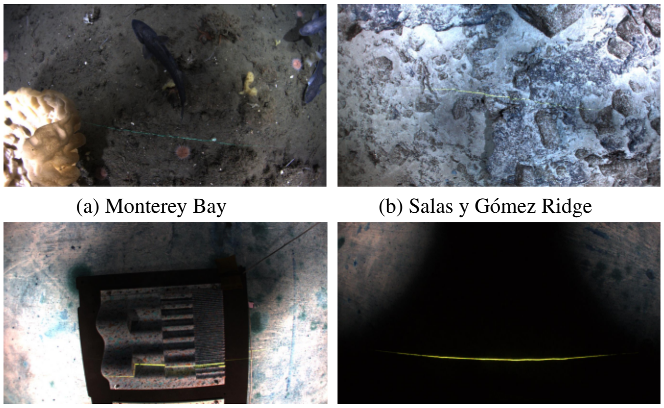
To assess the impact of the simulated dataset on segmentation performance, we used the same deep learning architecture across all experiments: **SPLASH-SegFormer** [4] . This network was trained using different proportions of synthetic and field data to systematically analyze the contribution of the simulated dataset to the target task.

All experiments were conducted on a workstation equipped with an NVIDIA GeForce RTX 3090 GPU. The input images were all in Full HD resolution (1920 × 1080).

For each training run, 10% of the dataset was reserved for validation, and the remaining 90% was used for training. The training process employed a OneCycle learning rate scheduler and early stopping to avoid overfitting. The best-performing model from each run, as determined by the validation loss, was saved and subsequently evaluated on the held-out test set.

3.3. Training Modes

To establish a baseline, we trained three models using only field data: one with the full dataset (Monterey Canyon + test tank), one with only test tank images, and one with only Monterey Canyon images. This allowed us to assess how the training domain affects generalization.



(a) Monterey Bay (b) Salas y Gómez Ridge

(c) MBARI's test tank (illuminated) (d) MBARI's test tank (non-illuminated)

Figure 3. Samples from field images

To evaluate the contribution of synthetic data, we trained three additional models: one using only synthetic data, one with a 1:1 ratio of synthetic to field data, and one with a 2:1 ratio. Finally, for reference, we trained a model including the test set (Salas y Gómez Ridge) in the training data. This result is shown as an upper-bound comparison, not used for evaluation.

3.4. Metrics

We evaluate the performance over the test set of our model using the following metrics: recall, precision, F1-score and Mean Intersection Over Union (mIoU) [10].

4. Results

Table 1 presents the evaluation metrics for each training mode, measured on the test set composed of images from the Salas y Gómez Ridge.

As expected, the lowest performance was observed in the models trained exclusively with test tank data or only with Monterey Canyon field data. This highlights the limitations of domain-specific training and the need for diverse data to ensure generalization.

Training with both Monterey and test tank datasets significantly improved the results, as the combined data provided greater variability in lighting conditions and seafloor textures.

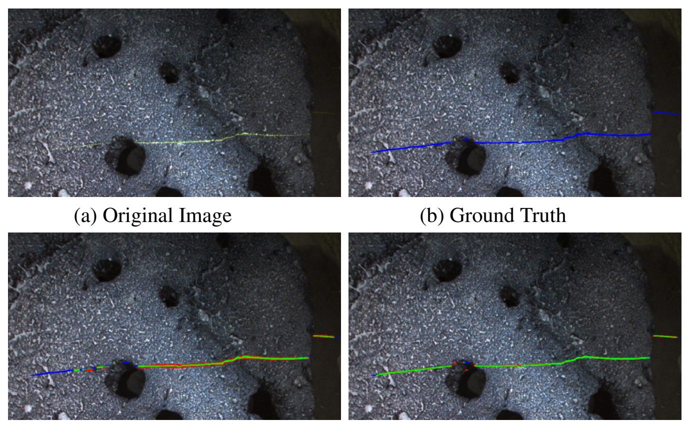
Building upon this baseline, incorporating synthetic images yielded further improvements. Training with a 1:1 ratio of synthetic to field data led to a 3% increase in recall and a 4% increase in precision compared to the baseline model.

Notably, training with a 2:1 ratio of synthetic to field data resulted in a 10% increase in recall and a 7% improvement in precision over the baseline. These results suggest that synthetic data enhances the model's ability to generalize and makes it more robust for deployment in previously unseen underwater environments.

Table 1. Model comparison across key metrics: recall, precision, mIoU and F1-score. The best result is bolded.

Model	Recall % ↑	Precision % ↑	F1-score ↑	mIoU↑
Test tank	27.31	66.93	34.79	24.53
Monterey	38.57	50.17	40.09	28.16
Previous Field	61.55	66.78	61.23	47.43
Sim only	64.46	65.04	61.71	45.80
Ratio 1:1	64.52	70.03	63.80	49.65
Ratio 2:1	**71.35**	**71.89**	**68.04**	**53.90**
*Baseline**	*84.28*	*80.53*	*81.91*	*69.98*

In Figure 4, we illustrate the results. The ground truth is shown in blue, the intersection between the prediction and ground truth (true positives) is shown in green, and false positives are highlighted in red. As the examples show, the network trained with synthetic data produces predictions that more closely align with the ground truth.



(a) Original Image (b) Ground Truth

(c) Prediction with Previous Field (d) Prediction with ratio 2:1

Figure 4. Qualitative Comparison Showing the Impact of Synthetic Data

5. Conclusion

By integrating a custom laser scanner module into Blender and Infinigen, we generated over 1,200 high-resolution synthetic images with ground truth masks across 23 procedurally generated terrains. Our experiments show that incorporating synthetic data into the training pipeline significantly improves performance—especially in recall and precision—when applied to previously unseen domains with different depths, lighting, and ecosystems.

These results validate synthetic data as a scalable, cost-effective solution for training robust underwater segmentation models. This paves the way for broader adoption of laser-based mapping systems and contributes to making detailed seafloor exploration more accessible and scalable.

References

[1] Michael Bleier, Joschka van der Lucht, and A. Nüchter. SCOUT3D – An Underwater Laser Scanning System for Mobile Mapping. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W18:13–18, 2019. 1

[2] Blender Online Community. Blender - a 3d modelling and rendering package, 2023. Blender Foundation. Stichting Blender Foundation, Amsterdam. 2

[3] Adrian Bodenmann, Blair Thornton, and Tamaki Ura. Generation of High-resolution Three-dimensional Reconstructions of the Seafloor in Color using a Single Camera and Structured Light. *Journal of Field Robotics*, 34(5):833–851, 2017. 1

[4] Javiera Fuentes-Guíñez, Giancarlo Troni, and Hans Lobel. Splash-segformer pipeline: A transformer-based approach for high-resolution and low-cost laser scanner seafloor mapping. Under revision at IEEE Robotics and Automation Letters (RA-L), 2025. 1, 3

[5] Li Kong, Li Ma, Kai Wang, Xiong Peng, and Nan Geng. Three-Dimensional-Scanning of Pipe Inner Walls Based on Line Laser. *Sensors*, 24(11):3554, 2024. 1

[6] Chongyi Li, Chunle Guo, Wenqi Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao. An underwater image enhancement benchmark dataset and beyond. *IEEE Transactions on Image Processing*, 29:4376–4389, 2020. 2

[7] Haitao Lin, Hua Zhang, Yonglong Li, Jianwen Huo, Hao Deng, and Huan Zhang. Method of 3D reconstruction of underwater concrete by laser line scanning. *Optics and Lasers in Engineering*, 183:108468, 2024. 1

[8] Risheng Liu, Xin Fan, Ming Zhu, Minjun Hou, and Zhongxuan Luo. Real-world underwater enhancement: Challenges, benchmarks, and solutions, 2019. 3

[9] Monterey Bay Aquarium Research Institute. Sketchfab profile - mbari. `https://sketchfab.com/mbari`, 2017. Organization / Scientific Organization. Member since June 5th, 2017. 2

[10] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830, 2011. 4

[11] Alexander Raistrick, Lahav Lipson, Zeyu Ma, Lingjie Mei, Mingzhe Wang, Yiming Zuo, Karhan Kayan, Hongyu Wen, Beining Han, Yihan Wang, Alejandro Newell, Hei Law, Ankit Goyal, Kaiyu Yang, and Jia Deng. Infinite photorealistic worlds using procedural generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12630–12641, 2023. 2

[12] Chris Roman, Gabrielle Inglis, and James Rutter. Application of structured light imaging for high resolution mapping of underwater archaeological sites. In *OCEANS'10 IEEE SYDNEY*, pages 1–9, 2010. 1

[13] Anne-Catherine Wölfl, Henry Snaith, Sam Amirebrahimi, Colin W Devey, Boris Dorschel, Vicki Ferrini, Veerle A Huvenne, Martin Jakobsson, James Jencks, Gary Johnston, et al. Seafloor mapping–the challenge of a truly global ocean bathymetry. *Frontiers in Marine Science*, page 283, 2019. 1