Single-View Seafloor Recovery from Imaging Sonar via Differentiable Rendering

Supplementary Material

Table 1. In-distribution synthetic sampling parameters. Empirical min/max over 10,000 train, 200 val, and 200 test samples.

Parameter	Min	Max
Sensor parameters		
Azimuth spread (°)	10.0	19.0
Start range (m)	1.53	4.97
End range (m)	3.80	7.50
Range coverage (m)	2.03	5.49
Range bins	380	512
Azimuth bins	36	64
Elevation spread (°)	10.0	19.0
Seafloor parameters		
Amplitude (cm)	2.01	9.98
Frequency (m^{-1})	2.00	15.0
Ground tilt (°)	6.87	38.0

Table 2. HoloOcean standard sampling parameters. Empirical min/max over 149 samples.

Parameter	Min	Max
Sensor parameters		
Azimuth spread (°)	10.0	29.0
Start range (m)	1.01	5.99
End range (m)	4.16	8.50
Range coverage (m)	2.51	5.50
Range bins	512	512
Azimuth bins	48	48
Elevation spread (°)	10.0	29.0
Seafloor parameters		
Amplitude (cm)	1.03	7.97
Frequency (m^{-1})	1.00	4.00
Ground tilt (°)	6.59	47.23

Table 3. HoloOcean rough-terrain sampling parameters. Empirical min/max over 35 samples.

Parameter	Min	Max
Sensor parameters		
Azimuth spread (°)	10.0	29.0
Start range (m)	5.14	7.98
End range (m)	8.56	12.62
Range coverage (m)	3.12	4.97
Range bins	512	512
Azimuth bins	48	48
Elevation spread (°)	11.0	28.0
Seafloor parameters		
Amplitude (cm)	10.0	15.9
Frequency (m^{-1})	1.00	3.50
Ground tilt (°)	14.4	51.1

1. Synthetic Dataset Parameters

Tabs. 1 to 3 summarize the empirical sampling ranges observed in each dataset. We report the minimum and maximum values across the full split for each setting. Note that the amplitude refers to the peak-to-trough height (maximum minus minimum elevation). Although we targeted matching configuration ranges between the synthetic and HoloOcean standard datasets, integration with HoloOcean resulted in slightly different realized parameter bounds. We present samples from each dataset in Fig. 5.

We refer to the synthetic dataset generated with our own forward model as "in-distribution," since the CNN is trained and evaluated on samples drawn from the same simulator and parameter ranges. The HoloOcean datasets differ in simulator, sensor model, and realized parameter ranges, and are therefore treated as out of distribution, even though the underlying seafloor prior (Perlin height fields) is shared.

2. Hyperparameters and Configs

Table 4 provides the key hyperparameter ranges for our differentiable renderer and inverse reconstructions. We choose the elevation sampling density, n_{el}, so that each beam approximates a continuous vertical fan. The near-range padding adds out-of-view bins to the front of the height field to prevent artifacts from offscreen occlusions. The ranges listed here include values that we found stable in practice. We did not tune these heavily for each dataset.

3. Additional Ablations

To further identify the optimal configurations of our system, and the benefits of the generic viewpoint approach, we perform additional ablations over the learning rate, optimization steps, and generic viewpoint coverage.

3.1. Learning Rate Ablation

To study the tradeoff between optimization budget and reconstruction accuracy we sweep the geometry learning rate across several orders of magnitude and vary the number

Table 4. Key hyperparameters for differentiable rendering and inverse reconstruction.

Component	Parameter	Default / range
Forward model / renderer	Elevation samples n_{el}	$6\,n_{\mathrm{bins}}$ per azimuth beam
	Gaussian bin spread σ_{bins}	0.5 - 1.0 bins
	Near-range padding	4 bins in front of $r_{\min}$
	Two-way geometric spreading	$1/r^4$ correction (TVG undo optional)
	TVG exponent	3.2 - 4.0
	Diffuse exponent γ	1.0 - 2.0
	Specular spread σ_{spec}	$5^\circ - 10^\circ$
Seafloor / intersections	Collision sharpness α	2500 - 4500
	HT plane coverage	$\approx 90\%$ of elevation span
	GV sampling range	60–97.5% of elevation span
Image processing	Azimuth gains	One scalar per beam
	Gain learning rate	$0 \rightarrow 10^{-1}$ (after warmup)
Optimization	Steps per frame	100-300
	TV weight λ_{TV}	0.0 - 1.0
	Optimizer	AdamW
	Geometry learning rate	10^{-4}
	Warmup steps n_{warmup}	30 (freeze gains)

of gradient steps. For each combination we fit the polar height field on the HoloOcean rough terrain dataset using the generic viewpoint (GV) approach and a fixed TV weight of 0.1. We report the 3D MSE over the test set.

As shown in Tab. 5, very small learning rates converge slowly and leave significant error even after many steps, while very large learning rates overshoot and degrade performance. The optimization remains stable up to a learning rate around 5×10^{-4}, where it begins to create extreme geometry resulting in large error. A mid–range value around 10^{-4} provides a stable compromise. It typically achieves optimal accuracy around 150-200 steps.

3.2. Generic Viewpoint

The generic–viewpoint prior is implemented by sampling base–plane orientations that cover a specified fraction of the sonar elevation span. At each optimization step, we sample a random plane within the specified range (from min. coverage to full coverage) and apply the corresponding tilt to the latent seafloor. This encourages reconstructions that remain consistent under small changes to the supporting plane.

To test how this prior performs under different sampling ranges, we vary the minimum coverage of the elevation span and evaluate the 3D MSE on all three datasets. Low coverage allows shallow planes that only intersect part of the range. High coverage forces the plane to span most of the elevation arc, which restricts the feasible orientations.

As shown in Tab. 6, the generic viewpoint prior remains stable across all sampling schemes. However, it clearly benefits from limiting very shallow planes, likely due to the sparse sampling and extreme geometry which occurs in this zone. Additionally, near very high coverage restrictions (around 90%) the model performs similarly to the high-tilt approach, since we are no longer sampling over a broad range of small tilt variations. The best performance in-distribution occurs when we sample from a minimum coverage of 70%, and for both HoloOcean datasets the best performance occurs when the minimum coverage is 40%.

4. Runtime and Scaling

We profile the runtime and GPU memory usage of our differentiable sonar renderer and single-frame inversion on a desktop RTX 5090 GPU. We vary the number of range and azimuth bins around typical imaging sonar settings (Range $\in \{300, 400, 500, 600\}$, Azimuth $\in \{48, 96, 128\}$). For each configuration we report the average forward render time, the wall-clock time required to complete a full optimization of a single frame with 150 gradient steps, and the peak GPU memory recorded by PyTorch during the run.

We provide results at two elevation sampling levels. In the first level we cast 1500 elevation rays per azimuth beam. This configuration prioritizes speed and is sufficient for most of our experiments. In the second level we use 3000 rays per beam, which provides higher fidelity results, particularly when the number of range bins is large. Runtime and memory scale approximately linearly with the number of azimuth beams, the number of range bins, and the number of elevation rays.

At a realistic ARIS-like resolution of 96×500 bins,

Table 5. HoloOcean Rough Terrain 3D MSE (cm^2, ↓) scores across optimization steps and learning rates. Evaluated using the generic viewpoint (GV) approach with $\lambda_{TV} = 0.1$. For each number of steps, the best learning rate is shown in bold. The overall optimal performance is achieved at 200 steps with a learning rate of 1×10^{-4}.

Steps / LR	1×10^{-6}	1×10^{-5}	5×10^{-5}	1×10^{-4}	5×10^{-4}
50	7.034	6.423	5.169	**4.834**	5.984
75	6.998	6.147	4.808	**4.529**	6.160
100	6.960	5.916	4.595	**4.364**	6.750
150	6.886	5.556	4.370	**4.219**	9.056
200	6.813	5.292	4.279	***4.200***	12.398
300	6.673	4.938	**4.286**	4.325	21.012
400	6.540	4.721	**4.399**	4.568	30.553

Table 6. 3D MSE (cm^2, ↓) across different generic viewpoint sampling schemes. Evaluated with $\lambda_{TV} = 0.1$.

GV Min. Coverage	In-Distribution	HoloOcean	HO Rough
0%	0.821	1.036	4.958
10%	0.792	0.963	4.633
20%	0.683	0.908	4.325
30%	0.632	0.890	4.223
40%	0.601	**0.885**	**4.165**
50%	0.583	0.894	4.174
60%	0.578	0.907	4.213
70%	**0.577**	0.917	4.285
80%	0.581	0.933	4.384
90%	0.599	0.960	4.532

the 1500-ray rendering requires 83 ms per forward render, 12.5 s for a full 150-step reconstruction, and a peak of 9.7 GiB of GPU memory. With 3000 rays the same configuration requires 165 ms per render, 24.8 s for reconstruction, and 19.4 GiB of peak memory. The largest configuration with 3000 rays and 600×128 bins exceeds 30 GiB and does not fit in GPU memory.

5. 3D Reconstructions of Real Riverbeds

In this section we present several examples of full 3D reconstructions produced by our differentiable renderer. These fits use a higher learning rate (3×10^{-3}) and stronger TV regularization ($\lambda_{TV} = 0.5$) than was used in the main paper. This choice increases contrast and occlusions in the rendered images, but also makes the underlying 3D structure easier to visualize. The full raw seafloor mesh for two real images (from Kenai Rightbank and Kenai Channel) are shown in Figs. 1 and 3 without modification. We show additional orthographic renders after applying one level of Catmull–Clark subdivision in Blender to smooth high–frequency noise in Figs. 2 and 4.

Table 7. Runtime and peak GPU memory for 1500 and 3000 elevation rays per azimuth beam on an RTX 5090 GPU. We report average forward render time, total optimization time for 150 steps, and peak GPU memory. The configuration with 3000 elevation rays, 600 range bins, and 128 azimuth bins exceeds 30 GiB of memory and does not fit.

	$n_{\mathrm{el}} = 1500$				$n_{\mathrm{el}} = 3000$		
Range bins	Render (ms)	Optim. (s)	Peak VRAM (GiB)	Range bins	Render (ms)	Optim. (s)	Peak VRAM (GiB)
48 Azimuth Beams							
300	31	4.6	2.9	300	56	8.3	5.8
400	37	5.6	3.9	400	70	10.5	7.7
500	45	6.7	4.8	500	84	12.6	9.7
600	52	7.8	5.8	600	98	14.8	11.6
96 Azimuth Beams							
300	57	8.6	5.8	300	107	16.1	11.6
400	71	10.6	7.7	400	134	20.1	15.5
500	83	12.5	9.7	500	165	24.8	19.4
600	100	15.0	11.6	600	197	29.5	23.2
128 Azimuth Beams							
300	73	10.9	7.8	300	148	22.2	15.5
400	94	14.2	10.3	400	185	27.7	20.6
500	113	17.0	12.9	500	223	33.5	25.8
600	132	19.8	15.5	600			>30

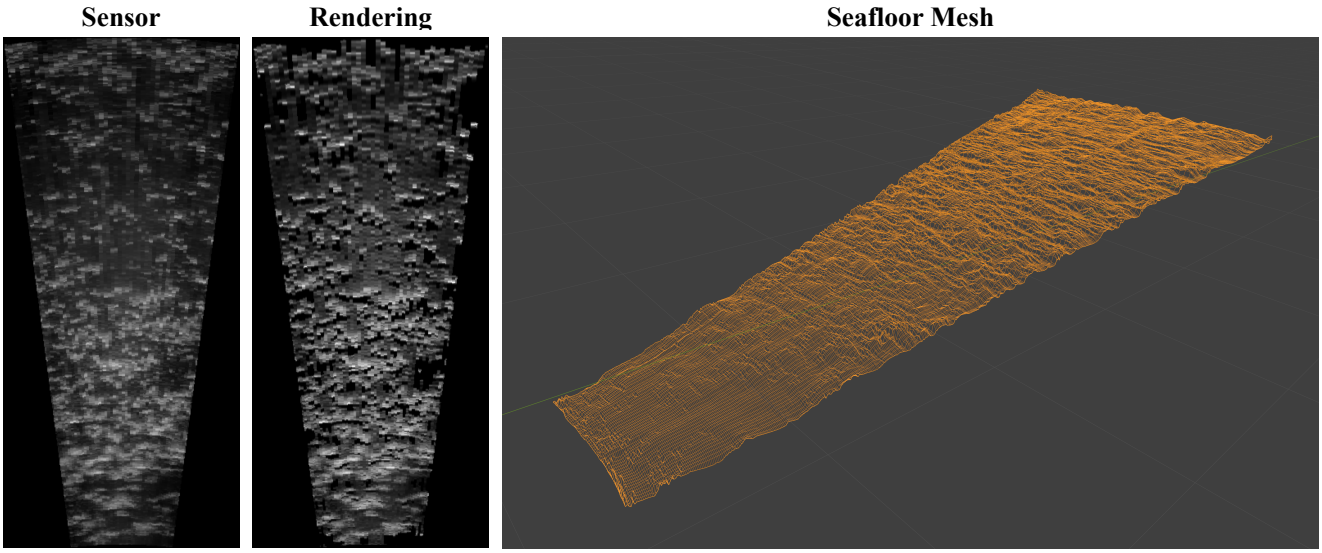

Figure 1. **Predicted seafloor mesh on Kenai Rightbank from a real sensor reading** (*cf.* Sec. 5).

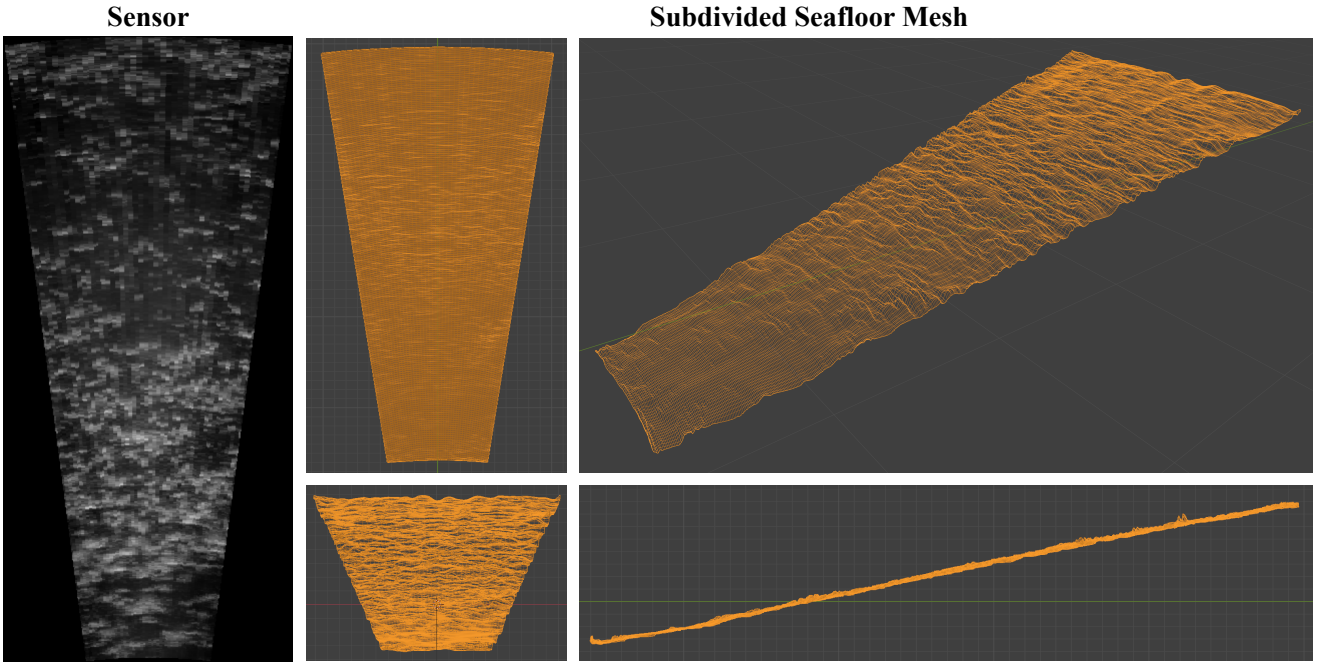

Figure 2. **Additional views of the recovered seafloor geometry on Kenai Rightbank** (*cf.* Sec. 5). On the left we show the target again, with additional orthographic views of the recovered height field on the right after one level of Catmull–Clark subdivision in Blender (applied only for visualization).

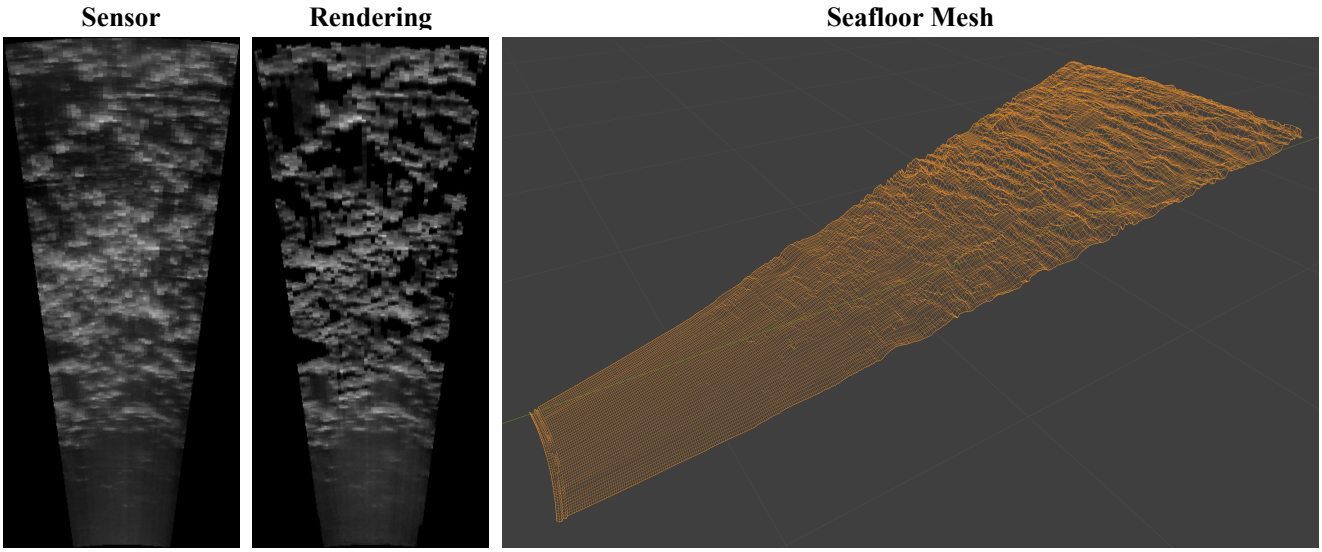

Figure 3. **Predicted seafloor mesh on Kenai Channel from a real sensor reading** (*cf.* Sec. 5).

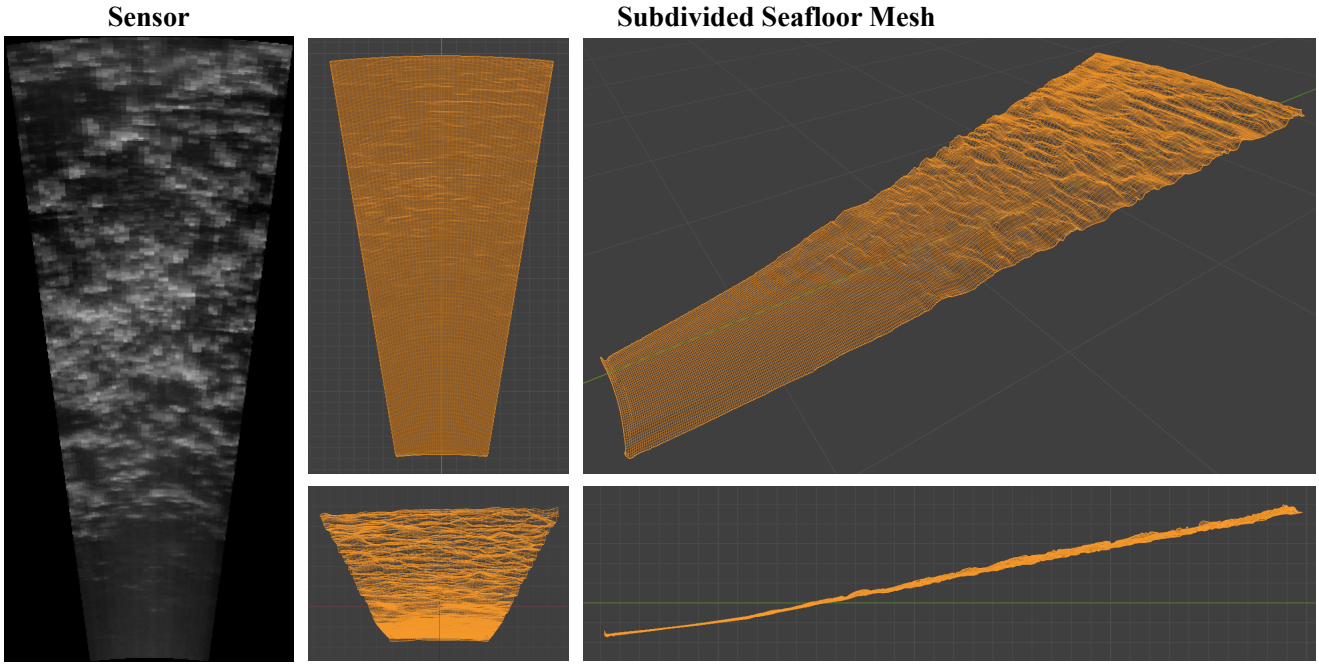

Figure 4. **Additional views of the recovered seafloor geometry on Kenai Channel** (*cf.* Sec. 5). On the left we show the target again, with additional orthographic views of the recovered height field on the right after one level of Catmull–Clark subdivision in Blender (applied only for visualization).

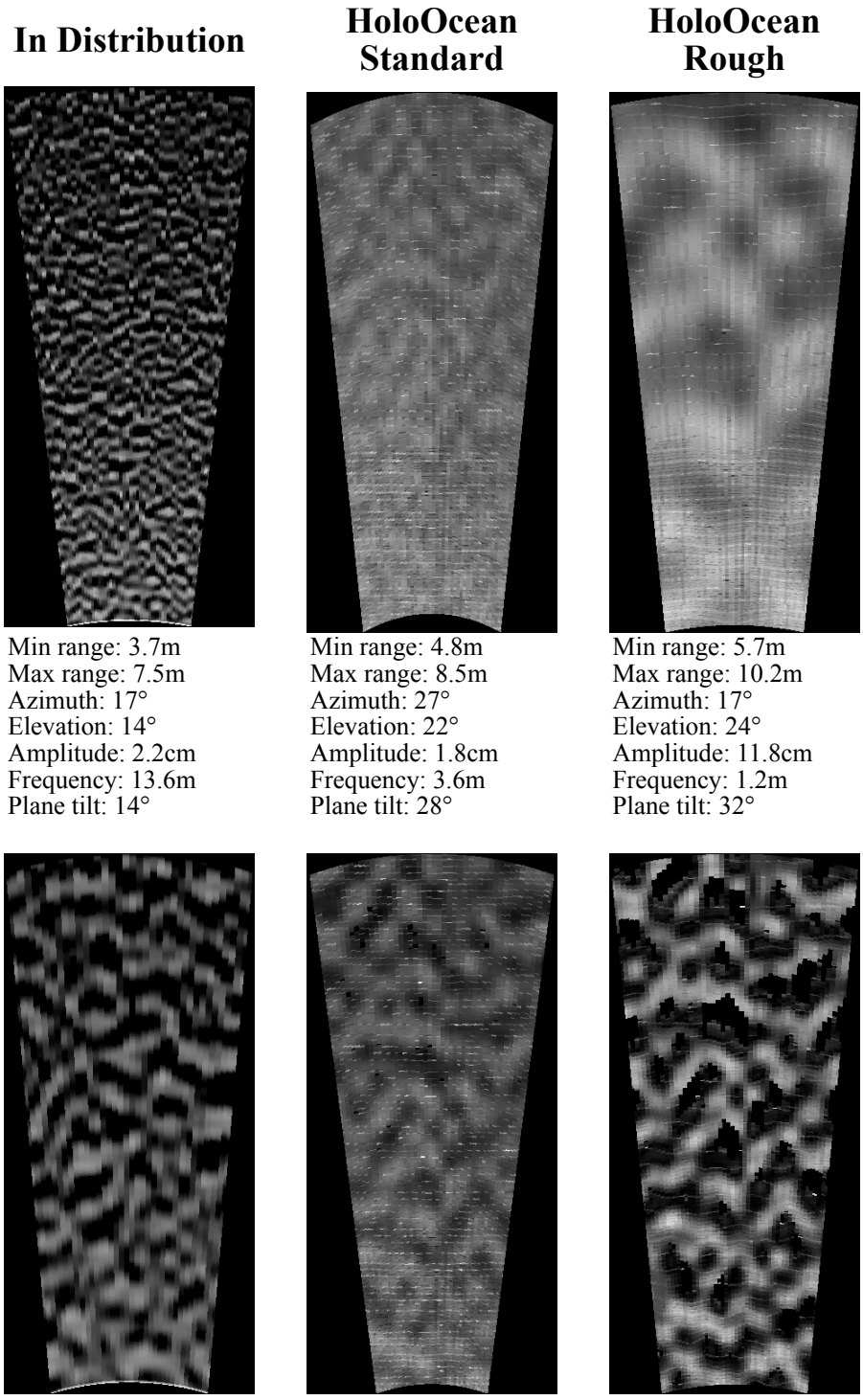

Figure 5. **Example target images from each of our three datasets**. Sampling details can be found in Sec. 1