WaterDescatterGS: Diverse Underwater 3D Scene Reconstruction Using Gaussian Splatting

Shijun Zhou[2,3]
Yuqing Xu[1]
Xing Xie[2,3]
Baojie Fan[1*]
Jiandong Tian[2]

[1]Nanjing University of Posts and Telecommunications,
[2]Shenyang Institute of Automation, Chinese Academy of Sciences,
[3]University of Chinese Academy of Sciences

Abstract

Existing physics-based approaches for underwater 3D restoration and reconstruction commonly model backscatter as a function of scene depth and attenuation coefficients. However, they often overlook the view-dependent nature of underwater scattering across camera poses, making it difficult to separate medium degradation from intrinsic scene radiance and thereby limiting both visual and geometric accuracy. To address this, we present WaterDescatterGS, an efficient physics-integrated underwater 3DGS framework that explicitly models view-dependent water degradation at the pixel level. Leveraging the multi-view geometry inherent to Structure-from-Motion, our framework dynamically derives base per-view illumination directions from known camera rotations and a single globally optimized sun direction. To accommodate complex real-world photometric variations beyond this physical prior, a Physics-Residual Appearance Embedding is introduced to absorb unmodeled degradations via view-specific angular corrections. Ultimately, these refined per-view parameters are naturally integrated into the physical model to compute precise per-pixel scattering angles coupled with the rendered depth. During inference, a cross-view attention mechanism dynamically aggregates adjacent features to enhance robust novel view synthesis in turbid water. Extensive experiments demonstrate that, despite operating entirely without clean reference supervision, WaterDescatterGS outperforms established baselines in novel view synthesis, remains competitive with concurrent methods, and improves novel-view-consistent scene restoration and geometric reconstruction accuracy on real and synthetic scenes.

1 Introduction

High-fidelity 3D scene reconstruction in underwater environments serves as the foundation for marine science, archaeology, and autonomous navigation. While 3D Gaussian Splatting (3DGS) [15] has recently transformed novel view synthesis for in-air scenes, extending it below the surface remains a formidable challenge. Water introduces wavelength-dependent attenuation and complex backscatter that vary significantly with viewpoint and spatial position [39], which cannot be adequately captured by the standard emission-absorption formulation used in 3DGS.

*Corresponding author.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

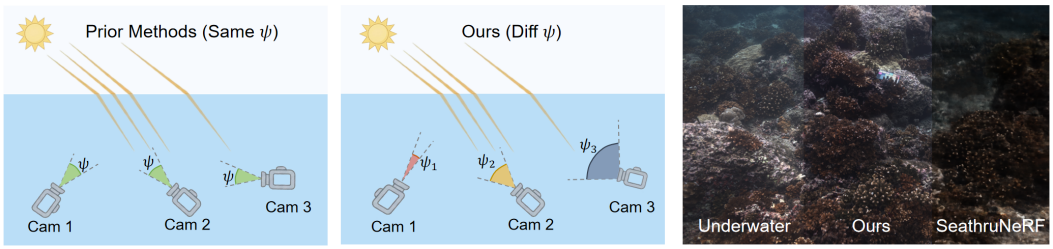

Figure 1: **Motivation and result.** (a) Prior physics-based methods treat the scattering angle ψ as a scene-level constant; for clarity we depict ψ at a single representative point along each ray, though in-scattering integrates along the full optical path. (b) Our method derives ψ from the relative geometry between each camera and a learned global sun direction, naturally producing view-dependent backscatter. (c) Qualitative restoration comparison on a real underwater scene.

A natural response is to embed a physical scattering model into the rendering pipeline. Underwater image formation is primarily governed by *attenuation*, which absorbs scene radiance along the optical path, and *backscatter*, which redirects ambient light toward the camera. Backscatter depends strongly on the *scattering angle* ψ between the incident sunlight and the viewing direction [39]. Because ψ is determined jointly by camera orientation and pixel position, it varies both across views and spatially within a single image. The seawater phase function $P(\psi)$ is non-monotonic, exhibiting a strong forward-scattering peak near $\psi \to 0°$ and a secondary backscattering peak near $\psi \to 180°$ [27], so even modest variations in the sun-view geometry translate into measurable changes in backscatter intensity. Modeling this angular structure is therefore essential for physically accurate rendering.

However, existing physics-based underwater reconstruction methods [46, 20] typically approximate underwater backscatter solely as a function of scene depth and medium attenuation coefficients. As illustrated in Figure 1(a), they fundamentally collapse the scattering angle ψ to a scene-level constant. The root cause is an *identifiability* problem: blindly estimating the view-dependent ψ for every camera pose without strong priors is highly ill-posed. While the angular dependence of scattering has long been recognized in classical models [13] and exploited in laboratory inverse rendering [8] and in-situ color restoration [30], recent single-image restoration methods [39] that rigorously formulate this dependence rely on complex numerical solvers. Directly integrating such heavy computations into a multi-view 3D pipeline is computationally prohibitive, fundamentally negating the real-time rendering advantage inherent to 3DGS.

To break this ambiguity, our key insight is that the multi-view geometry inherent to Structure-from-Motion (SfM) [35] provides precisely the constraints needed (Figure 1b). Since obtaining accurate camera poses is a prerequisite for 3D reconstruction, a single learned global sun direction suffices to derive a base illumination direction for every camera view. No per-image light estimation is required, and the number of additional parameters is minimal (two angles for the sun). In other words, the multi-view geometry inherited from SfM is the *missing constraint* that resolves the physical ambiguity, while the 3DGS framework turns this solution into a real-time, per-pixel rendering engine.

Building on this, we present **WaterDescatterGS**, a framework where physics and geometry mutually reinforce each other: accurate geometry enables precise scattering estimation, while physical constraints regularize the geometry in return. We explicitly model view-dependent degradation via a dual-parameterization strategy: SfM-derived base illumination is refined by appearance embeddings, which absorb unmodeled photometric variations (e.g., refraction-induced illumination shifts) as *bounded* residuals on the light direction rather than unconstrained latent codes. These optimized per-view parameters then naturally drive the physical model to compute precise per-pixel scattering angles. Furthermore, we introduce position-aware per-Gaussian attenuation coefficients to capture spatially varying turbidity, alongside a leave-one-out cross-view attention mechanism for robust novel view synthesis.

Our contributions are summarized as follows:

- **Physics-geometry synergy.** We show that coupling SfM camera rotations with a learned global sun direction yields per-pixel scattering angles for every view without per-image light estimation.

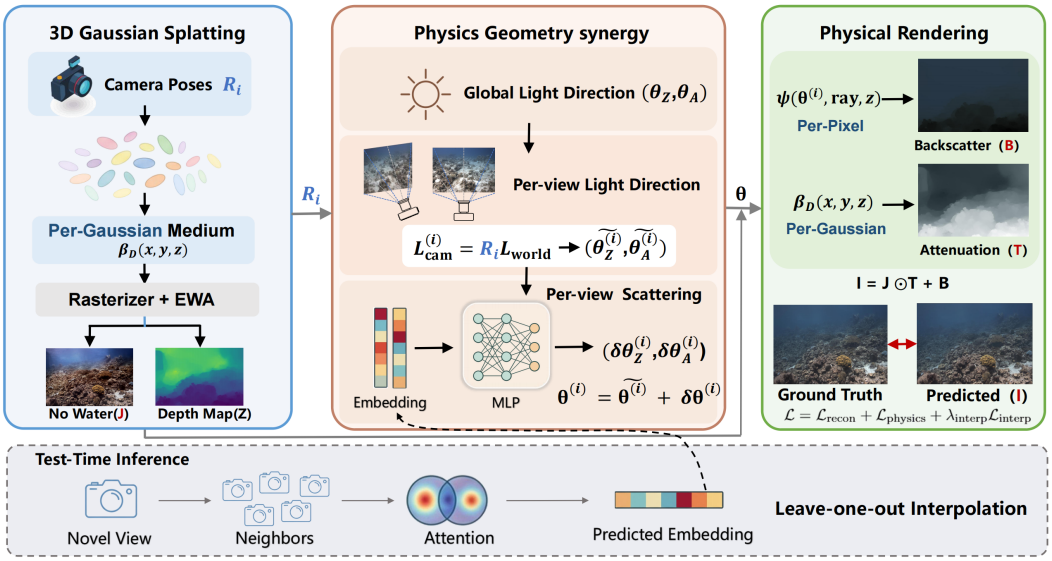

Figure 2: **Framework overview. (Left)** 3DGS reconstructs clean radiance J and depth Z, enhanced by per-Gaussian medium attenuation β_D and EWA splatting. **(Middle)** The physics-geometry module derives base per-view illumination from a global sun direction via camera poses, refined by learned appearance embeddings. **(Right)** Per-view parameters θ drive the physical model to compute per-pixel scattering angles ψ, composing $I = J \odot T + B$. **(Bottom)** A leave-one-out attention network interpolates embeddings for novel views.

- **Bounded residual medium modeling.** Appearance embeddings constrained as bounded angle residuals on the light direction absorb photometric variations, and position-aware per-Gaussian attenuation captures spatially varying turbidity.
- **Cross-view attention.** A leave-one-out training strategy enables robust, artifact-free generalization to novel viewpoints at test time.
- **Strong performance** on real-world underwater benchmarks in novel view synthesis, novel-view-consistent scene restoration, and geometric reconstruction.

2 Related Work

Neural rendering for underwater scenes. Neural rendering provides a unified framework for jointly modeling geometry and appearance [25, 15, 29, 12, 16, 38, 49, 45]. SeaThru-NeRF [18] jointly optimizes medium parameters with a NeRF radiance field but its dense volumetric sampling limits efficiency; related media-aware NeRF variants further extend this paradigm to general scattering [26, 22, 51]. Recent 3DGS-based approaches offer efficient rasterization: SeaSplat explicitly models attenuation and scattering in 3DGS [46], WaterSplatting [20] introduces a learned neural water field, UW-GS [41] suppresses distractors for robustness, and concurrent work Gaussian splashing [28] performs direct volumetric rendering underwater. Most underwater 3DGS methods, however, still treat degradation as appearance fitting or rely on scene-level/spatially uniform medium assumptions, limiting physical interpretability across diverse water conditions. A complementary line of single-image enhancement methods [43, 7, 54, 9, 52, 21, 53, 37, 31, 19] restores individual frames in 2D but cannot enforce multi-view 3D consistency.

Physics-based and view-dependent underwater modeling. Several methods integrate underwater image formation into explicit 3D representations [46, 1, 49], and some condition medium parameters on the viewing ray: WaterNeRF [36], WaterHE-NeRF [55], and RestorGS [33]. These approaches model backscatter as a function of depth and *viewing direction* rather than the scattering angle ψ between sunlight and the viewing ray. The two are different physical quantities: viewing direction governs surface appearance, whereas ψ drives the phase function that modulates volumetric scattering. Tian et al. [39] show that scattering varies spatially even in homogeneous media due to imaging geometry. We bring this insight into 3DGS: since the sun direction is fixed within a single dive, camera rotation alone produces sufficient angular variation in ψ to drive our formulation (Figure 3).

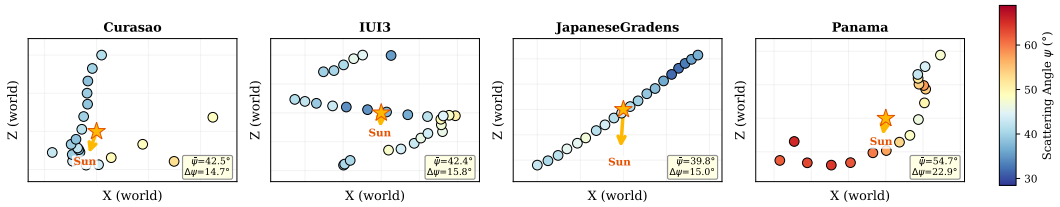

Figure 3: Top-down view of camera positions across four scenes, colored by scattering angle ψ. With the sun direction (star) fixed in world coordinates, camera rotation alone produces 14.7°–60.9° variation in ψ, motivating per-view light angle extraction from SfM rotations (Equation (5)).

From static light priors to per-pixel scattering angles. SeaThru-NeRF [18] introduced per-scene medium coefficients into the ray-marching integral. Its medium term depends on the distance along the ray, while viewing direction is handled by the radiance field, which keeps the medium compact and treats the illumination as static within a scene. We build on this formulation and add the angle between illumination and the line of sight: ψ is computed per pixel from the SfM camera rotation and a globally optimized sun direction, and it drives the phase function inside the backscatter term. The medium response therefore varies across views and across pixels. Within our framework, replacing this angular field with one learned per-scene constant lowers PSNR by 1.23 dB (Table 3a). Concurrent work 3D-UIR [48] decouples appearance and medium within 3DGS, and Plenodium [44] learns a plenoptic medium conditioned on position and viewing direction. Our work focuses on the sun–view scattering angle, a complementary aspect; we compare against both in Section 4.2 and Section V.

3 Method

Given posed underwater images $\{I_i\}_{i=1}^N$, we jointly optimize a 3D Gaussian scene representation and recover physical parameters of underwater light transport without clean reference images. Our framework bridges SfM camera poses with a physics-based formation model (Sections 3.1 and 3.2), and further integrates latent appearance embeddings for view-dependent variations (Sections 3.3 and 3.4) and multi-scale medium modeling for localized turbidity (Section 3.5), all optimized end-to-end.

3.1 Underwater Image Formation Model

Following the Jaffe-McGlamery model [13, 24], the underwater image at pixel $\mathbf{x}$ and channel $c \in \{R, G, B\}$ is:

$$I^c(\mathbf{x}) = \underbrace{J^c(\mathbf{x}) \cdot t^c(\mathbf{x})}_{\text{attenuated scene}} + \underbrace{B^c(\mathbf{x})}_{\text{backscatter}}, \tag{1}$$

where J^c is the clean scene radiance, t^c is the transmission map, and B^c is the backscattered light from suspended particles.

Attenuation. The transmission map follows Beer-Lambert and the parameterizations of attenuation coefficients used in modern underwater imaging [3, 1]:

$$t^c(\mathbf{x}) = \exp\big(-\beta_D^c \cdot \eta(\mathbf{x}) \cdot z(\mathbf{x})\big), \tag{2}$$

where β_D^c is the wavelength-dependent attenuation coefficient, $z(\mathbf{x})$ is the scene depth, and $\eta(\mathbf{x}) = \sqrt{x_n^2 + y_n^2 + f^2}/f$ corrects for oblique ray paths.

Backscatter. Suspended particles scatter ambient illumination toward the camera:

$$B^c(\mathbf{x}) = B_s^c \cdot f \cdot \tilde{P}(\psi) \cdot \frac{\eta(\mathbf{x})}{f_\eta(\mathbf{x})} \cdot \Big(1 - \exp\big(-\beta_B^c \cdot z(\mathbf{x}) \cdot f_\eta(\mathbf{x})/f\big)\Big), \tag{3}$$

The Fournier-Forand parameters $n \in [1.33, 1.47]$ and $\mu \in [3, 5]$ are jointly optimized within physically plausible ranges [27, 6] and remain stable, indicating that ψ, not phase-function shape, is the primary driver.

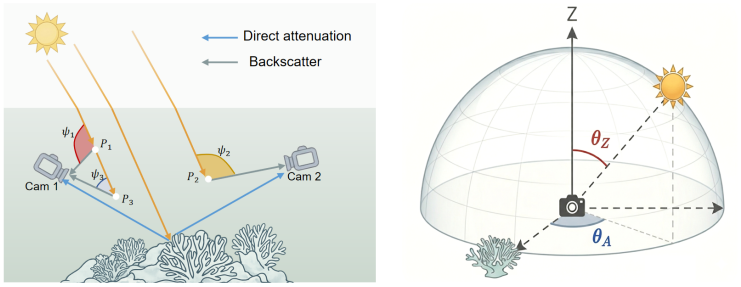

Figure 4: **Geometric formulation of view-dependent underwater light transport. (Left)** Our model captures multi-level angular scattering variance, including both inter-camera ($\psi_1 \neq \psi_2$) and intra-camera ($\psi_1 \neq \psi_3$) dependencies, which directly modulate backscatter intensity via the phase function $\tilde{P}(\psi)$. **(Right)** The global sun direction is parameterized using zenith (θ_Z) and azimuth (θ_A) angles.

3.2 Physics-Geometry Synergy

The central insight is deceptively simple: *the sun does not move between frames; only the camera does*. Following the directional-illumination assumption adopted in widely used underwater image formation models [1, 39], we represent the absolute sun direction in world coordinates using just two global, learnable angles: zenith θ_Z^w and azimuth θ_A^w. As illustrated in Figure 4, our geometric formulation captures two levels of angular variation: (1) **inter-camera**, where different poses observe different scattering angles (e.g., $\psi_1 \neq \psi_2$); and (2) **intra-camera**, where different pixels within the same field of view possess distinct scattering angles (e.g., $\psi_1 \neq \psi_3$).

This multi-level angular variation modulates the phase function $\tilde{P}(\psi)$ in Equation (3). Independent per-image light estimation is severely ill-posed without geometric priors; we instead exploit the camera rotation matrices $\mathbf{R}_i$ that 3DGS already provides from SfM.

Global sun parameterization. The sun direction in world coordinates is parameterized by two learnable angles, zenith θ_Z^w and azimuth θ_A^w, shared across all images:

$$\mathbf{L}_{\text{world}} = \begin{pmatrix} \sin\theta_Z^w \sin\theta_A^w \\ \cos\theta_Z^w \\ \sin\theta_Z^w \cos\theta_A^w \end{pmatrix}. \tag{4}$$

Camera-space light direction. For each image i with camera rotation $\mathbf{R}_i$ (world-to-camera), the base light direction in camera coordinates is:

$$\mathbf{L}_{\text{cam}}^{(i)} = \mathbf{R}_i \, \mathbf{L}_{\text{world}}. \tag{5}$$

Via inverse spherical coordinates, we extract the base per-camera light angles $(\tilde{\theta}_Z^{(i)}, \tilde{\theta}_A^{(i)})$ from $\mathbf{L}_{\text{cam}}^{(i)}$. This grounds per-view illumination in the global sun position (θ_Z^w, θ_A^w) and SfM camera geometry, building multi-view consistency directly into the parameterization rather than enforcing it via soft regularization.

Rendering pipeline. We render clean radiance J and depth z via standard 3DGS alpha compositing [15]:

$$J^c(\mathbf{x}) = \sum_j C_j^c \alpha_j \prod_{k<j}(1-\alpha_k), \quad z(\mathbf{x}) = \sum_j z_j \alpha_j \prod_{k<j}(1-\alpha_k), \tag{6}$$

where C_j^c, z_j, α_j are the color, depth, and opacity of Gaussian j. The physical layer synthesizes: $\hat{I}^c = J^c \odot t^c + B^c$, with $J^c \odot t^c$ rendered by applying attenuation to each Gaussian *prior* to splatting (Section 3.5), and B^c computed from Equation (3). All components are differentiable, enabling end-to-end optimization without clean reference images.

3.3 Physics-Residual Appearance Embedding

The analytical Jaffe-McGlamery model relies on idealized scattering assumptions, whereas real underwater captures often exhibit per-image photometric variations that cannot be explained by

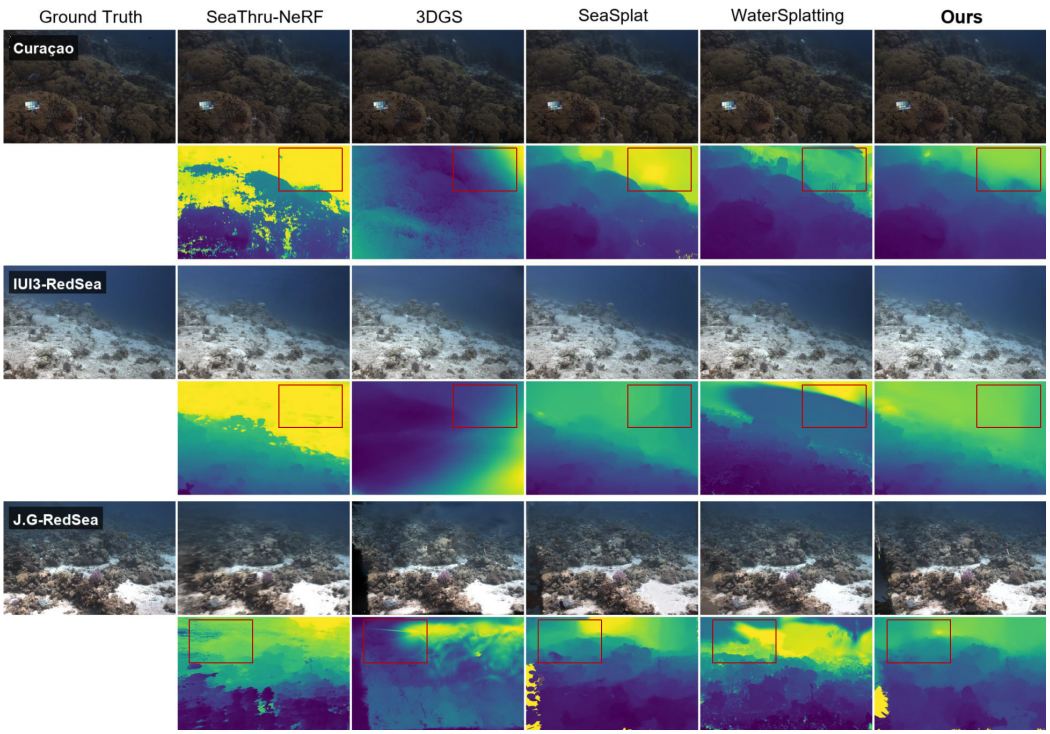

Figure 5: **Novel view synthesis comparison.** Top: rendered RGB. Bottom: depth. Our method produces sharper geometry and more accurate depth, particularly for complex coral structures.

static scene geometry alone. For instance, dynamic surface waves refract sunlight and introduce small fluctuations in the effective illumination direction that are difficult to capture with a static analytical model. To bridge this gap between theoretical physics and real observations, we introduce a hybrid compensation mechanism. Inspired by NeRF-W [23], each training image i is assigned a learnable embedding $\mathbf{e}_i \in \mathbb{R}^d$ (d=32). Crucially, instead of producing unconstrained latent corrections that entangle geometry and appearance, the embedding is decoded exclusively into *physically meaningful angular residuals* $(\delta\theta_Z, \delta\theta_A)$. This constrains the network to act as a low-dimensional angular corrector, compensating directional shifts in the phase function $\tilde{P}(\psi)$ (see Equation (3)) while preserving the physical structure of the scattering model and maintaining stable medium optimization:

$$(\delta\theta_Z^{(i)},\ \delta\theta_A^{(i)}) = g_\phi\left(\mathbf{e}_i,\ \bar{z},\ \frac{\tilde{\theta}_Z^{(i)}}{\pi/2},\ \frac{\tilde{\theta}_A^{(i)}}{2\pi}\right), \tag{7}$$

where g_ϕ is a shared MLP and $\bar{z}$ is the mean rendered depth, detached from the computation graph. Denoting the raw network outputs as (r_Z, r_A), we smoothly bound them via sigmoid clamping:

$$\delta\theta_Z = S_Z \cdot \big(2\sigma(r_Z) - 1\big) \in [-S_Z, S_Z], \quad \delta\theta_A = S_A \cdot \big(2\sigma(r_A) - 1\big) \in [-S_A, S_A], \tag{8}$$

with S_Z=20° and S_A=90°, reflecting the physical prior that sun elevation varies less than azimuth. The final light direction combines the geometry-derived base with the learned residual: $\theta^{(i)} = \tilde{\theta}^{(i)} + \delta\theta^{(i)}$. The last layer of g_ϕ is initialized with near-zero weights, ensuring the global sun direction converges first before per-image refinements take effect. The per-pixel scattering angle ψ is then computed as the geometric angle between $\theta^{(i)}$ and each viewing ray, driving the phase function in Equation (3).

3.4 Leave-One-Out Learned Interpolation

Per-image embeddings (Section 3.3) absorb temporal photometric drift but are tied to specific training images; a novel test viewpoint therefore has no embedding of its own. To predict a plausible embedding $\hat{\mathbf{e}}_q$ for an unseen query view, we propose a lightweight interpolation network that learns

which neighbors are most informative via attention. Given query camera position $\mathbf{T}_q$, we find $K{=}5$ nearest training cameras and construct features:

$$\mathbf{f}_j = \left[\mathbf{e}_j;\ \frac{\mathbf{T}_j - \mathbf{T}_q}{\|\mathbf{T}_j - \mathbf{T}_q\| + \epsilon};\ \|\mathbf{T}_j - \mathbf{T}_q\|\right] \in \mathbb{R}^{d+4}, \tag{9}$$

separating direction from distance. Each feature is encoded by a shared MLP and aggregated via multi-head attention ($H{=}4$ heads), then decoded to the predicted embedding $\hat{\mathbf{e}}_q$. Training uses leave-one-out supervision: $\mathcal{L}_{\text{interp}} = \|\hat{\mathbf{e}}_i - \mathbf{e}_i\|_2^2$.

3.5 Multi-Scale Medium Modeling

The image-level model assumes spatially uniform water, which is a first-order approximation that cannot capture **local** turbidity variations. We address this at three scales.

Per-Gaussian medium attenuation (PGMNetwork). A position-conditioned MLP predicts spatially varying attenuation: $\beta_D^{(j)} = \text{PGM}(\mathbf{x}_j) \in \mathbb{R}_+^3$. Each Gaussian is attenuated *prior* to splatting: $\tilde{c}_j = c_j \cdot \exp(-\beta_{D,c}^{(j)} \cdot \hat{z}_j \cdot Z_s)$, where $\hat{z}_j \in [0,1]$ is normalized depth and Z_s is a learnable scene depth scale. This 3D pre-splatting formulation naturally resolves overlapping structures at different depths with different local water properties. While the appearance embedding (Section 3.3) absorbs per-frame illumination drift, PGMNetwork captures per-location attenuation heterogeneity; the two are orthogonal by construction (image-indexed vs. position-indexed).

Anti-aliased EWA splatting. Standard Gaussian splatting can produce aliasing when projected Gaussians become sub-pixel, an effect amplified in underwater scenes where depth ambiguity encourages smaller Gaussians. Inspired by [4, 5], we apply an Elliptical Weighted Average (EWA) low-pass filter [47, 56]: $\tilde{\mathbf{\Sigma}}^{2D} = \mathbf{\Sigma}^{2D} + s_{\text{blur}} \cdot \mathbf{I}$, with density compensation $\tilde{\alpha}_j = \alpha_j \cdot \sqrt{\det(\mathbf{\Sigma}^{2D}) / \det(\tilde{\mathbf{\Sigma}}^{2D})}$ to preserve total energy. Empirically, EWA shows positive synergy with PGMNetwork (Section 4.3), as smoother rendering gradients stabilize the per-Gaussian attenuation learning. By mitigating high-frequency structural artifacts, it prevents the network from incorrectly baking spatial aliasing into the local medium properties.

Higher-order scattering modulation. While the Fournier-Forand phase function models primary single-scattering events, real turbid environments exhibit higher-order multiple scattering [32]. At the depth ranges of our benchmarks ($<$10 m, clear to moderate water), single scattering remains dominant, but residual multi-scattering effects introduce view-dependent deviations. A view-dependent modulation driven by degree-3 spherical harmonic (SH) representations [40] dynamically scales the base backscatter: $B^c \leftarrow B^c \cdot \Delta B^c$. By constraining ΔB tightly around 1.0, the single-scattering prior remains the dominant anchor, while the SH parameterization absorbs bounded higher-order residuals rather than replacing the physics.

3.6 Training Objectives

Our total loss combines photometric reconstruction with physics-based regularization:

$$\mathcal{L} = \mathcal{L}_{\text{recon}} + \mathcal{L}_{\text{physics}} + \lambda_{\text{interp}} \mathcal{L}_{\text{interp}}. \tag{10}$$

Reconstruction loss. We combine pixel-wise ℓ_1 with D-SSIM [42] and VGG-based perceptual loss [14]:

$$\mathcal{L}_{\text{recon}} = (1 - \lambda_{\text{dssim}})\, \mathcal{L}_1 + \lambda_{\text{dssim}}\, \mathcal{L}_{\text{D-SSIM}} + \lambda_{\text{percep}}\, \mathcal{L}_{\text{VGG}}, \tag{11}$$

with $\lambda_{\text{dssim}}{=}0.4$ and $\lambda_{\text{percep}}{=}0.02$ determined by ablation (Section G).

Physics regularization. The decomposition is inherently underdetermined. We prevent degenerate solutions through complementary priors on the recovered clean radiance $\hat{J}$: Dark Channel Prior [10], Gray World and White Balance losses for color neutrality, saturation bounds, and edge-aware depth smoothness (details in Section F).

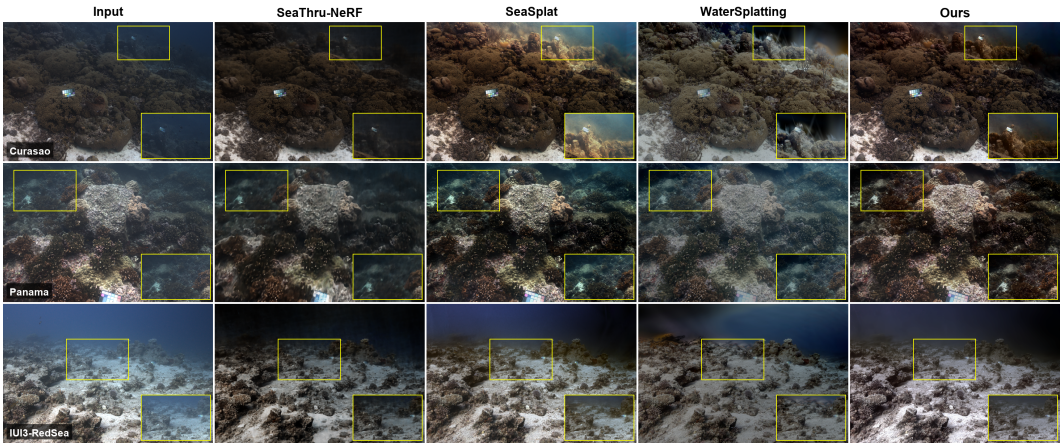

Figure 6: **Qualitative comparison of scene restoration.** Restored clean radiance J recovered by each method on three scenes (Curasao, Panama, IUI3-RedSea). SeaThru-NeRF [18] retains significant haze and color cast. SeaSplat [46] reduces backscatter but leaves residual fog. WaterSplatting [20] produces desaturated, low-contrast outputs. Our method removes more backscatter and preserves higher contrast than competing approaches.

Table 1: Quantitative comparison on the SeaThru-NeRF dataset. Cells shaded by rank: 1st , 2nd , 3rd .

	Curasao			IUI3			Japanese			Panama			Average			Efficiency	
Method	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓	FPS↑	VRAM↓
3DGS	0.873	28.31	0.221	0.843	22.98	0.246	0.854	21.49	0.216	0.893	29.20	0.152	0.866	25.50	0.209	150.28	3.7
SeaSplat	0.898	29.70	0.189	0.869	26.85	0.213	0.872	22.57	0.180	0.895	28.16	0.148	0.884	26.82	0.183	90.05	6.0
SeaThru-NeRF	0.824	23.54	0.352	0.784	22.65	0.341	0.734	18.82	0.359	0.851	22.78	0.290	0.798	21.95	0.336	0.10	37.7
WaterSplatting	0.922	29.49	0.148	0.889	29.38	0.203	0.876	23.32	0.152	0.943	31.15	0.104	0.908	28.34	0.152	88.19	5.8
3D-UIR [48]	0.907	30.98	0.187	0.841	28.30	0.252	0.857	23.37	0.187	0.900	29.82	0.181	0.876	28.12	0.202	–	–
Ours	0.954	33.22	0.139	0.913	29.52	0.195	0.883	23.85	0.169	0.945	31.20	0.116	0.924	29.45	0.155	90.05	6.0

4 Experiments

4.1 Experimental Setup

Datasets. We evaluate on the SeaThru-NeRF dataset [18] (four scenes: **Curasao, IUI3-RedSea, JapaneseGardens-RedSea, Panama**), the **SaltPond** scene [46], and the **D3** scene from the original SeaThru dataset [2]. All evaluated datasets share standard underwater-capture preprocessing (stable exposure, color-balanced raw data). We follow the standard llffhold=8 protocol [25]: images are sorted by filename and every eighth image is held out for testing. The scenes contain 21 (Curasao), 29 (IUI3), 20 (JapaneseGardens), 18 (Panama), 68 (D3), and 260 (SaltPond) images, giving 18/3, 25/4, 17/3, 15/3, 59/9, and 227/33 train/test views.

Baselines and Metrics. We compare with 3DGS [15] (no underwater modeling), SeaSplat [46] (physics-based 3DGS), SeaThru-NeRF [18] (physics-based NeRF), and WaterSplatting [20] (learning-based 3DGS with neural water field). We also list the reported results of the concurrent 3D-UIR [48], and rerun the concurrent Plenodium [44] on our splits (Section V). We report SSIM [42], PSNR, and LPIPS [50] on held-out test views.

Implementation. Built on 3DGS, we train for 30k iterations with a two-stage strategy (physics activates at 5k). Key hyperparameters: d=32 embedding dim, K=5 neighbors with 4-head attention, PGM hidden dim 16 (lr 10^{-3}), EWA blur strength 0.3. One model is optimized per scene, following the standard 3DGS protocol. The four benchmark scenes and D3 share one configuration of modules, loss weights, and physics settings. Hyperparameters were chosen on the four benchmark scenes, since no standard validation split exists; the sensitivity studies (Sections C and H) show stable results across broad settings. SaltPond, with over 200 training images, uses 60k iterations and a larger λ_{dssim} (Section M). Single NVIDIA RTX 4090 GPU.

Table 2: Results on SaltPond [46] and D3 [2]: generalization to unseen water types.

	SaltPond			D3			Efficiency	
Method	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓	FPS↑	VRAM↓
3DGS	0.723	27.43	0.300	0.844	27.99	0.289	149.85	4.3
SeaSplat	0.730	27.42	0.271	0.851	28.27	0.256	89.68	7.1
SeaThru-NeRF	0.512	12.03	0.578	0.562	20.34	0.555	0.11	43.2
WaterSplatting	0.646	26.02	0.418	0.852	28.03	0.219	88.45	6.6
Ours	0.747	27.48	0.247	0.895	33.05	0.207	90.31	7.4

4.2 Comparison with State-of-the-Art

Quantitative results. Table 1 presents results on the four-scene benchmark. Among the methods in the table, our method ranks first on average SSIM (0.924) and PSNR (29.45), outperforming WaterSplatting in SSIM by 1.8%. Average LPIPS is comparable (0.155 vs. 0.152), and per-scene LPIPS is mixed: WaterSplatting leads on JapaneseGardens and Panama. The reported numbers of 3D-UIR average 0.876 SSIM and 28.12 dB PSNR. The concurrent Plenodium is a strong competitor on this benchmark; its same-split results are reported in Section V. On Curasao, our method is best on all three metrics; on Panama, it leads in SSIM and PSNR. The PSNR gains over WaterSplatting are +3.73 dB and +0.05 dB, respectively. On the challenging JapaneseGardens scene (high turbidity), we achieve the best SSIM (0.883) and PSNR (23.85), demonstrating that physics-based constraints are particularly effective in degraded conditions. Table 2 shows results on two additional datasets. On D3, our method achieves SSIM 0.895 and PSNR 33.05, substantially outperforming all baselines and confirming the cross-dataset generalization of our physics prior. On SaltPond (200+ images, 60k iterations), we surpass SeaSplat in SSIM (0.747 vs. 0.730) and LPIPS (0.247 vs. 0.271).

Qualitative results. Figure 5 shows novel view synthesis and depth estimation results. While baseline methods achieve reasonable RGB rendering, their depth maps struggle to decouple the scattering medium from solid structures, resulting in noisy depth and spurious floating artifacts. Our method recovers precise geometry with smooth background transitions and sharp foreground boundaries. Figure 6 further compares the restored clean radiance J. Strictly coupling rendering with physical principles prevents baking scene color into the medium—a failure mode of WaterSplatting that causes desaturated outputs.

4.3 Ablation Studies

Table 3 presents the component ablation. Removing the physics model causes the largest drop (-2.6% SSIM, -3.33 dB PSNR), confirming explicit physical modeling is the primary driver. Critically, the no-physics baseline (t=1, B=0) *still retains* all photometric priors (DCP, Gray World, White Balance, saturation bounds) and all auxiliary modules (embedding, PGM, EWA, SH, perceptual loss); the -2.6% SSIM drop therefore isolates the physical model's contribution from these engineering components. Appearance embedding (-1.9% SSIM) is the second most important component. Camera rotation encoding (-0.4%) is more impactful than learned interpolation (-0.2%), confirming that exploiting SfM geometry provides a strong prior for view-dependent scattering. To isolate the angular field, the *w/o per-pixel* ψ row replaces the per-pixel scattering angle with one learnable per-scene constant, keeping every other module unchanged. The constant converges to $47.4°$–$50.5°$ from a $45°$ initialization, yet PSNR drops by 1.23 dB. This setting corresponds to a scene-level light prior.

To isolate the analytic physical core, a physics-only variant retaining only ψ+FF, the UIFM, and photometric priors (DCP/GW/WB/Sat) already exceeds the no-physics baseline by +1.8% SSIM and +1.02 dB PSNR, showing the physical model itself, not the auxiliary modules, drives the gains. Table 3b shows that the full combination of multi-scale components gives the best result.

4.4 Analysis

Component synergy. Appearance embeddings stabilize physical parameters against photometric variations; SH captures residual directional scattering, while PGMNetwork resolves spatially varying turbidity. Anti-aliased EWA provides smooth gradients that stabilize PGMNetwork, yielding positive synergy when combined (Table 3).

Table 3: Ablation study (four-scene average). **(a)** Core components. **(b)** Multi-scale medium modeling. The physics-only row strips all auxiliary modules, retaining only ψ+FF and photometric priors.

(a) Core components

Configuration	SSIM↑	PSNR↑	LPIPS↓
Full model	**0.924**	**29.45**	**0.155**
w/o physics	0.898	26.12	0.185
w/o embedding	0.905	26.80	0.172
w/o camera rot.	0.920	28.19	0.157
w/o per-pixel ψ	0.919	28.22	0.159
w/o SH medium	0.919	27.65	0.163
w/o perceptual	0.919	27.54	0.164
w/o interpolation	0.922	28.27	0.157
physics-only	0.916	27.14	0.168

(b) Multi-scale medium

Configuration	SSIM↑	PSNR↑	LPIPS↓
Base (img-level)	0.914	26.62	0.175
+ SH medium	0.919	27.54	0.163
+ EWA splatting	0.916	27.40	0.163
+ SH + EWA	0.919	28.29	0.164
+ SH + Perceptual	0.921	28.20	0.162
+ PGMNetwork	0.922	28.37	0.158
Full model	**0.924**	**29.45**	**0.155**

5 Limitations

Our method assumes a single directional sunlight source, restricting applicability to natural-light water regimes; artificial-lit deep-sea or multi-source environments are not addressed, where polarimetric techniques [34] remain complementary alternatives. The Fournier-Forand phase function models single-scattering events; while our SH modulation absorbs bounded multi-scattering residuals, the approach may degrade in highly turbid waters where cumulative multiple scattering becomes isotropic [32]. The observed scattering angle range does not probe the near-forward ($\psi<5^\circ$) or near-backward ($\psi>170^\circ$) peaks of the phase function, so FF identifiability in those regimes remains untested. On synthetic scenes, the absolute sun direction is recovered only approximately and is sensitive to initialization; attenuation magnitudes are identifiable only up to a global per-channel scale; and the inversion degrades measurably beyond an optical thickness of about 2.8 (Section T). Finally, the pipeline depends on accurate COLMAP poses: rotation noise of 0.5° on all training cameras lowers PSNR on Curasao from 33.22 to 25.47 dB (Section U). Integrating SfM-free 3DGS is orthogonal future work.

6 Conclusion

We presented WaterDescatterGS, a physics-integrated 3DGS framework that bridges multi-view geometry with rigorous underwater light transport. By coupling SfM camera rotations with a learned global sun direction, our method derives per-pixel scattering angles that drive a physically grounded formation model, outperforming established baselines and remaining competitive with concurrent methods on diverse real-world scenes without clean reference supervision. The framework provides a reliable geometric and visual foundation for downstream underwater applications.

Acknowledgments and Disclosure of Funding

This work was supported in part by the National Natural Science Foundation of China under Grant 62473205, the Natural Science Foundation of Jiangsu Province under Grant BK20241893, and the Guangdong Basic and Applied Basic Research Foundation under Grant 2024B1515250008, and was sponsored by the Young Academic Leader program of the Qing Lan Project in Jiangsu Province. This work was also supported by the Natural Science Foundation of China under Grant 2024YFE0115500. The authors declare no competing interests.

References

[1] Derya Akkaynak and Tali Treibitz. A revised underwater image formation model. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6723–6732, 2018.

[2] Derya Akkaynak and Tali Treibitz. Sea-thru: A method for removing water from underwater images. In *CVPR*, pages 1682–1691, 2019.

[3] Derya Akkaynak, Tali Treibitz, Tom Shlesinger, Yossi Loya, Raz Tamir, and David Iluz. What is the space of attenuation coefficients in underwater computer vision? In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4931–4940, 2017.

[4] Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5470–5479, June 2022.

[5] Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. Zip-nerf: Anti-aliased grid-based neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 19697–19705, October 2023.

[6] Georges R Fournier and J Luc Forand. Analytic phase function for ocean water. In *Ocean Optics XII*, volume 2258, pages 194–201. SPIE, 1994.

[7] Zhenqi Fu, Wu Wang, Yue Huang, Xinghao Ding, and Kai-Kuang Ma. Uncertainty inspired underwater image enhancement. In *European conference on computer vision*, pages 465–482. Springer, 2022.

[8] Ioannis Gkioulekas, Shuang Zhao, Kavita Bala, Todd Zickler, and Anat Levin. Inverse volume rendering with material dictionaries. *ACM Transactions on Graphics (TOG)*, 32(6):1–13, 2013.

[9] Chunle Guo, Ruiqi Wu, Xin Jin, Linghao Han, Weidong Zhang, Zhi Chai, and Chongyi Li. Underwater ranker: Learn which is better and how to be better. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 702–709, 2023.

[10] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE TPAMI*, 33(12):2341–2353, 2010.

[11] Peter Hedman, Julien Philip, True Price, Jan-Michael Frahm, George Drettakis, and Gabriel Brostow. Deep blending for free-viewpoint image-based rendering. *ACM Transactions on Graphics*, 37(6):1–15, 2018.

[12] Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, and Shenghua Gao. 2d gaussian splatting for geometrically accurate radiance fields. In *ACM SIGGRAPH 2024 Conference Papers*, SIGGRAPH '24, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400705250. doi: 10.1145/3641519.3657428. URL `https://doi.org/10.1145/3641519.3657428`.

[13] Jules S Jaffe. Computer modeling and the design of optimal underwater imaging systems. *IEEE Journal of Oceanic Engineering*, 15(2):101–111, 1990.

[14] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016.

[15] Bernhard Kerbl, Georgios Kopanas, Laura Leal-Taixé, and George Drettakis. 3D gaussian splatting for real-time radiance field rendering. *ACM TOG*, 42(4):1–14, 2023.

[16] Bernhard Kerbl, Andreas Meuleman, Georgios Kopanas, Michael Wimmer, Alexandre Lanvin, and George Drettakis. A hierarchical 3d gaussian representation for real-time rendering of very large datasets. *ACM Transactions On Graphics (TOG)*, 43(4):1–15, 2024.

[17] Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics*, 36(4):1–13, 2017.

[18] Deborah Levy, Amit Peleg, Naama Pearl, Dan Rosenbaum, Derya Akkaynak, Simon Korman, and Tali Treibitz. Seathru-NeRF: Neural radiance fields in scattering media. In *CVPR*, pages 56–65, 2023.

[19] 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, 2019.

[20] Huapeng Li, Wenxuan Song, Tianao Xu, Alexandre Elsig, and Jonas Kulhanek. WaterSplatting: Fast underwater 3D scene reconstruction using gaussian splatting. In *3DV*, pages 969–978, 2025.

[21] Wei-Tung Lin, Yong-Xiang Lin, Jyun-Wei Chen, and Kai-Lung Hua. Pixmamba: Leveraging state space models in a dual-level architecture for underwater image enhancement. In *Proceedings of the Asian Conference on Computer Vision*, pages 3622–3637, 2024.

[22] Shuhong Liu, Lin Gu, Ziteng Cui, Xuangeng Chu, and Tatsuya Harada. I2-nerf: Learning neural radiance fields under physically-grounded media interactions. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

[23] Ricardo Martin-Brualla, Noha Radwan, Mehdi S. M. Sajjadi, Jonathan T. Barron, Alexey Dosovitskiy, and Daniel Duckworth. NeRF in the wild: Neural radiance fields for unconstrained photo collections. In *CVPR*, pages 7210–7219, 2021.

[24] Benjamin L McGlamery. A computer model for underwater camera systems. In *Ocean Optics VI*, volume 208, pages 221–231. SPIE, 1980.

[25] 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, 2021.

[26] Ben Mildenhall, Peter Hedman, Ricardo Martin-Brualla, Pratul P Srinivasan, and Jonathan T Barron. Nerf in the dark: High dynamic range view synthesis from noisy raw images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16190–16199, 2022.

[27] Curtis D. Mobley. *Light and Water: Radiative Transfer in Natural Waters*. Academic Press, San Diego, 1994.

[28] Nir Mualem, Roy Amoyal, Oren Freifeld, and Derya Akkaynak. Gaussian splashing: Direct volumetric rendering underwater. *arXiv preprint arXiv:2411.19588*, 2024.

[29] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)*, 41(4): 1–15, 2022.

[30] David Nakath, Mengkun She, Yifan Song, and Kevin Köser. In-situ joint light and medium estimation for underwater color restoration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3731–3740, 2021.

[31] Opher Bar Nathan, Deborah Levy, Tali Treibitz, and Dan Rosenbaum. Osmosis: Rgbd diffusion prior for underwater image restoration. In *European conference on computer vision*, pages 302–319. Springer, 2024.

[32] Nick Pfeiffer and Glenn H Chapman. Successive order, multiple scattering of two-term henyey-greenstein phase functions. *Optics express*, 16(18):13637–13642, 2008.

[33] Yuanjian Qiao, Mingwen Shao, Lingzhuang Meng, and Kai Xu. Restorgs: Depth-aware gaussian splatting for efficient 3d scene restoration. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 11177–11186, 2025.

[34] Yoav Y Schechner and Nir Karpel. Recovery of underwater visibility and structure by polarization analysis. *IEEE Journal of oceanic engineering*, 30(3):570–587, 2006.

[35] Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *CVPR*, pages 4104–4113, 2016.

[36] Advaith Venkatramanan Sethuraman, Manikandasriram Srinivasan Ramanagopal, and Katherine A. Skinner. WaterNeRF: Neural radiance fields for underwater scenes. In *OCEANS 2023 – MTS/IEEE U.S. Gulf Coast*, 2023.

[37] Zijun Tan, Chuan Fu, Tan Guo, Zhixiong Nan, Pengzhan Zhou, Xinggan Peng, and Fulin Luo. Pgmamba: A physical model-guided global mamba for underwater image enhancement. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 9359–9367, 2026.

[38] Yunkai Tang, Chengxuan Zhu, Renjie Wan, Chao Xu, and Boxin Shi. Neural underwater scene representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11780–11789, 2024.

[39] Jiandong Tian, Shijun Zhou, Baojie Fan, and Hui Zhang. A novel image formation model for descattering. *IEEE TPAMI*, 46(12):8173–8190, 2024.

[40] Dor Verbin, Peter Hedman, Ben Mildenhall, Todd Zickler, Jonathan T. Barron, and Pratul P. Srinivasan. Ref-nerf: Structured view-dependent appearance for neural radiance fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(11):9426–9437, 2025. doi: 10.1109/TPAMI.2024.3360018.

[41] Haoran Wang, Nantheera Anantrasirichai, Fan Zhang, and David Bull. Uw-gs: Distractor-aware 3d gaussian splatting for enhanced underwater scene reconstruction. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 3280–3289. IEEE, 2025.

[42] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4): 600–612, 2004.

[43] Junjie Wen, Jinqiang Cui, Zhenjun Zhao, Ruixin Yan, Zhi Gao, Lihua Dou, and Ben M. Chen. Syreanet: A physically guided underwater image enhancement framework integrating synthetic and real images. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5177–5183, 2023. doi: 10.1109/ICRA48891.2023.10161531.

[44] Changguang Wu, Jiangxin Dong, Chengjian Li, and Jinhui Tang. Plenodium: Underwater 3D scene reconstruction with plenoptic medium representation. In *Advances in Neural Information Processing Systems*, 2025.

[45] Tianyi Xie, Zeshun Zong, Yuxing Qiu, Xuan Li, Yutao Feng, Yin Yang, and Chenfanfu Jiang. Physgaussian: Physics-integrated 3d gaussians for generative dynamics. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4389–4398, 2024.

[46] Daniel Yang, John J Leonard, and Yogesh Girdhar. SeaSplat: Representing underwater scenes with 3D gaussian splatting and a physically grounded image formation model. In *ICRA*, pages 7632–7638, 2025.

[47] Zehao Yu, Anpei Chen, Binbin Huang, Torsten Sattler, and Andreas Geiger. Mip-splatting: Alias-free 3d gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 19447–19456, June 2024.

[48] Jieyu Yuan, Yujun Li, Yuanlin Zhang, Chunle Guo, Xiongxin Tang, Ruixing Wang, and Chongyi Li. 3D-UIR: 3D Gaussian for underwater 3D scene reconstruction via physics based appearance-medium decoupling. *IEEE Transactions on Image Processing*, 2026. arXiv:2505.21238.

[49] Yifan Zhan, Shohei Nobuhara, Ko Nishino, and Yinqiang Zheng. Nerfrac: Neural radiance fields through refractive surface. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 18402–18412, 2023.

[50] 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*, pages 586–595, 2018.

[51] Tianyi Zhang and Matthew Johnson-Roberson. Beyond nerf underwater: Learning neural reflectance fields for true color correction of marine imagery. *IEEE Robotics and Automation Letters*, 8(10):6467–6474, 2023.

[52] Zengxi Zhang, Zhiying Jiang, Jinyuan Liu, Xin Fan, and Risheng Liu. Waterflow: Heuristic normalizing flow for underwater image enhancement and beyond. In *Proceedings of the 31st ACM International conference on multimedia*, pages 7314–7323, 2023.

[53] Chen Zhao, Weiling Cai, Chenyu Dong, and Chengwei Hu. Wavelet-based fourier information interaction with frequency diffusion adjustment for underwater image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8281–8291, 2024.

[54] Jingchun Zhou, Jiaming Sun, Chongyi Li, Qiuping Jiang, Man Zhou, Kin-Man Lam, Weishi Zhang, and Xianping Fu. Hclr-net: Hybrid contrastive learning regularization with locally randomized perturbation for underwater image enhancement. *International Journal of Computer Vision*, 132(10):4132–4156, 2024.

[55] Jingchun Zhou, Tianyu Liang, Dehuan Zhang, Siyi Liu, Jingyi Wang, and Edmond Q. Wu. WaterHE-NeRF: Water-ray matching neural radiance fields for underwater scene reconstruction. *Information Fusion*, 115:102770, 2025.

[56] M. Zwicker, H. Pfister, J. van Baar, and M. Gross. Ewa volume splatting. In *Proceedings Visualization, 2001. VIS '01.*, pages 29–538, 2001. doi: 10.1109/VISUAL.2001.964490.

A Fournier-Forand Phase Function

The Fournier-Forand phase function [6] approximates the angular scattering probability for oceanic particles. Given scattering angle ψ, refractive index n, and Junge slope μ:

$$\tilde{P}(\psi) = \frac{1}{4\pi(1-\delta)^2\delta^\nu}\left[\nu(1-\delta) - (1-\delta^\nu) + \frac{\delta(1-\delta^\nu)}{\sin^2(\psi/2)}\cdot(\nu-(1-\delta))\right] + \frac{1-\delta_{180}^\nu}{16\pi(\delta_{180}-1)\delta_{180}^\nu}\left(3\cos^2\psi - 1\right), \tag{12}$$

where $\delta = \frac{4}{3(n-1)^2}\sin^2(\psi/2)$ and $\nu = \frac{3-\mu}{2}$.

B Scattering Angle Distribution

Camera rotation alone produces $14.7°$–$60.9°$ variation in ψ across views in our benchmark scenes, exercising a non-trivial portion of the FF curve between the forward-scattering lobe and the minimum near $90°$.

C Angle Scale Ablation

Table 4: Sensitivity to angle scale hyperparameters S_Z/S_A (four-scene average).

S_Z	S_A	SSIM↑	PSNR↑	LPIPS↓
$5°$	$20°$	0.920	26.99	0.158
$10°$	$45°$	0.921	28.11	0.160
$20°$	$90°$	0.924	29.45	0.155
$40°$	$180°$	0.923	27.92	0.159
$60°$	$270°$	0.922	28.40	0.158

SSIM varies by only 0.4% across a $12\times$ range of clamp magnitudes. This confirms the embedding refines the physical prior but cannot replace it: if the embedding were free to absorb the physical effect, larger clamps would yield substantial gains.

D Per-Scene PGM Analysis

Table 5 reports the per-scene breakdown. On its own, PGMNetwork helps on Curasao (+0.93 dB), whose mixed coral and sand make the water spatially heterogeneous, but lowers PSNR on the other three scenes, most strongly on uniformly turbid IUI3 (-2.65 dB). Combined with EWA, it improves all four scenes over the baseline: Curasao +1.15 dB, IUI3 +1.03 dB, JapaneseGardens +1.10 dB, and Panama +0.33 dB.

Table 5: PGMNetwork per-scene ablation. PGMNetwork alone helps only on Curasao; combined with EWA it improves all four scenes.

Configuration	Curasao	IUI3	Japanese	Panama
SSIM				
Baseline (no PGM)	0.9523	**0.9133**	0.8780	0.9409
+ PGMNetwork	0.9541	0.9046	0.8800	0.9443
+ PGMNetwork + EWA	**0.9542**	0.9132	**0.8834**	**0.9451**
PSNR				
Baseline (no PGM)	32.07	28.49	22.75	30.87
+ PGMNetwork	33.00	25.84	22.59	30.56
+ PGMNetwork + EWA	**33.22**	**29.52**	**23.85**	**31.20**

The EWA + PGM combination is therefore what yields consistent per-scene gains, in line with the synergy described in Section 3.5.

E Camera Distribution Sensitivity

We dissect query–training camera distribution effects directly from the llffhold=8 test split. Because test views are sampled every eighth frame along the capture trajectory, the first and last test view of each scene sit at the trajectory boundary with no training neighbour on one side, naturally producing the "training cameras only on one side of the query" condition. This setting directly tests two cases: all training cameras far from the query, and training cameras only on one side of it.

Re-evaluating the full model (30k iterations) on the four benchmark scenes per test view (Figure 7) reveals an asymmetric edge effect. In three of four scenes the *first* test view is the lowest-scoring view of that scene, by margins of 5.3 dB on Curasao (MTN_1288), 14.0 dB on IUI3 (MTN_5894), and 6.0 dB on JapaneseGardens (MTN_1090; see Figure 8 for the visible foreground hallucination in the region lacking neighbouring training cameras). Panama is an exception: all three of its test views fall within a 1.9 dB band (29.91–31.78 dB), reflecting a more closed capture trajectory in which both nominal "edges" still see bilateral coverage from neighbouring training cameras. This contrast is informative: where the trajectory genuinely terminates on one side of the query (Curasao/IUI3/JG first views), one-sided coverage costs 5–14 dB; where it does not (Panama), no such penalty arises. Pooling all 13 test views, edge views average PSNR/SSIM 26.83 / 0.903 versus interior views' 30.28 / 0.935 (-3.44 dB, $-3.6\,\%$ SSIM relative).

As a complementary and stricter ablation, we additionally trained Curasao on only one half of the camera rig (split by COLMAP-position median) and evaluated on the same held-out set: SSIM 0.947 (first half) vs. 0.949 (second half) vs. 0.954 (full train, Table 1), a $\leq 0.7\,\%$ relative drop comparable to SeaSplat under the same split. The leave-one-out interpolation thus generalises when one-sided gaps remain within trajectory extent, while the per-view edge-degradation above bounds the failure regime when test queries fall outside the convex hull of training cameras.

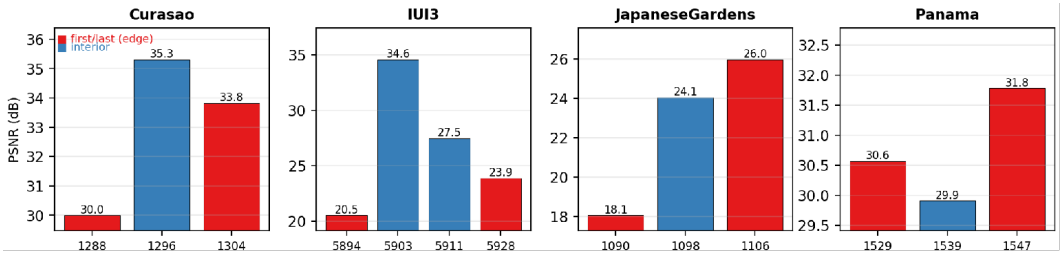

Figure 7: **Per-test-view PSNR across the four-scene benchmark** (full model). First/last (edge) views in red, interior views in blue. The first test view is the worst in three of four scenes (Curasao -5.3 dB, IUI3 -14 dB, JG -6 dB); Panama, with a more closed trajectory, shows no such penalty (all three views within 1.9 dB).

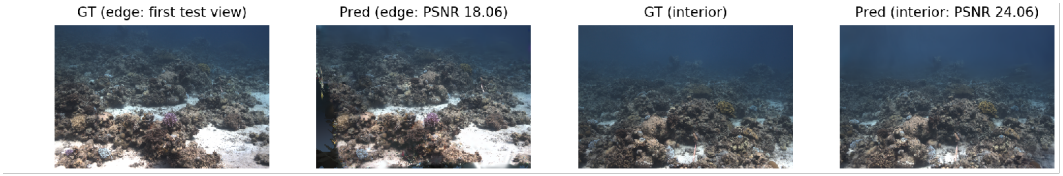

Figure 8: **Visual evidence of one-sided coverage on JapaneseGardens.** Edge view MTN_1090 (left two: GT, prediction at 18.06 dB) versus interior view MTN_1098 (right two: GT, prediction at 24.06 dB). The boundary view exhibits visibly degraded foreground reconstruction where training cameras are only on one side of the query.

F Physics Regularization Details

The physics regularization loss $\mathcal{L}_{\text{physics}}$ consists of five complementary priors applied to the recovered clean radiance $\hat{J}$.

Dark Channel Prior (DCP). Following He et al. [10], natural outdoor images contain pixels with near-zero intensity in at least one color channel. We enforce this prior on $\hat{J}$ via an asymmetric

penalty:

$$\mathcal{L}_{\text{DCP}} = \mathbb{E}\Big[\max(0,\,-\hat{J})^2 + \kappa \cdot \text{SmoothL1}\big(\max(0,\,-\hat{J}),\,0\big)\Big], \tag{13}$$

where κ=1000 penalizes negative radiance values (physically impossible) much more strongly than positive violations, preventing degenerate decompositions where the medium model over-corrects.

Gray World. Under the gray world hypothesis, the mean color of a properly white-balanced scene should be achromatic:

$$\mathcal{L}_{\text{GW}} = \sum_{c\in\{R,G,B\}} \big(\bar{J}_c - \tau\big)^2, \tag{14}$$

where $\bar{J}_c = \frac{1}{HW}\sum_{x,y}\hat{J}_c(x,y)$ is the spatial mean of channel c and τ=0.5 is the target neutral gray level.

White Balance. Bright scene points (e.g., white sand, coral highlights) should exhibit equal channel ratios after removing the water medium:

$$\mathcal{L}_{\text{WB}} = \left(\frac{\hat{J}_R(\mathbf{x}^*)}{\hat{J}_G(\mathbf{x}^*)} - 1\right)^2 + \left(\frac{\hat{J}_B(\mathbf{x}^*)}{\hat{J}_G(\mathbf{x}^*)} - 1\right)^2, \tag{15}$$

where $\mathbf{x}^* = \arg\max_{\mathbf{x}} \|\hat{J}(\mathbf{x})\|_2$ among pixels above the 97th intensity percentile.

Saturation. Clean radiance values should remain within the physically valid range $[0, s]$:

$$\mathcal{L}_{\text{Sat}} = \mathbb{E}\Big[\max(0,\,-\hat{J})^2 + \max(0,\,\hat{J}-s)^2\Big], \tag{16}$$

with s=1.0 enforcing the hard upper bound.

Edge-aware depth smoothness. Depth maps should be spatially smooth except at object boundaries indicated by image edges:

$$\mathcal{L}_{\text{depth}} = \mathbb{E}\Big[|\nabla_x z|\, e^{-|\nabla_x I|} + |\nabla_y z|\, e^{-|\nabla_y I|}\Big], \tag{17}$$

where the exponential weighting allows depth discontinuities at image edges while encouraging smoothness in textureless regions.

Loss weights. λ_{DCP}=0.5, λ_{GW}=0.05, λ_{WB}=0.05, λ_{sat}=4.0, λ_{depth}=1.0.

G Loss Ablation

D-SSIM weight λ_{dssim}=0.4 and perceptual weight λ_{percep}=0.02 were selected via grid search on the four-scene benchmark. We found λ_{dssim} is the most impactful hyperparameter; higher values (0.6–0.7) benefit the larger SaltPond scene.

D-SSIM weight. The structural similarity term provides complementary gradient signals that help preserve edges and textures. Table 6 shows the impact of varying λ_{dssim} on reconstruction quality.

Table 6: D-SSIM weight ablation on curasao. Higher λ_{dssim} prioritizes structural similarity over pixel-wise accuracy.

λ_{dssim}	SSIM↑	PSNR↑	LPIPS↓
0.15	0.9461	31.09	0.1426
0.2 (3DGS default)	0.9474	31.02	0.1401
0.3	0.9462	30.74	**0.1367**
0.4 (ours)	**0.9540**	**33.22**	0.1390

Perceptual weight. We use λ_{percep}=0.02 for the VGG-based perceptual loss, selected via grid search on the four-scene benchmark; this primarily improves LPIPS with minimal impact on SSIM.

H Hyperparameter Sensitivity

We analyse sensitivity to remaining key hyperparameters. Unless noted, results are 4-scene averages.

Appearance embedding dimension. We use d=32. Validated on SaltPond (d=64 vs d=32: SSIM 0.7398 vs. 0.7410 at 60k iterations), suggesting 32 dimensions provide sufficient capacity without overfitting.

SH medium correction parameters. Table 7 shows the impact of SH degree on the medium correction MLP. Degree-3 with lr=10^{-3} provides optimal performance.

Table 7: SH medium correction hyperparameters (4-scene average). Degree-3 significantly outperforms lower-order approximations.

SH degree	Learning rate	SSIM↑	PSNR↑	LPIPS↓
1	10^{-3}	0.9152	27.85	0.1709
2	10^{-3}	0.9132	27.60	0.1738
3	10^{-3} (ours)	**0.9240**	**29.45**	**0.1550**

Backscatter / attenuation learning rate. We use 5×10^{-3} for both the backscatter source B_s^c and attenuation β_B^c, reduced from the initial 10^{-2} via grid search; the lower rate consistently improved PSNR across scenes, particularly on the IUI3-RedSea.

Table 8: Backscatter/attenuation learning rate ablation. Lower rate (5×10^{-3}) provides better PSNR with minimal SSIM trade-off.

Configuration	SSIM↑	PSNR↑	LPIPS↓
2-scene validation (Curasao + IUI3)			
10^{-2}	0.9284	29.23	0.1745
5×10^{-3}	**0.9337**	**31.37**	**0.1670**

I Camera Rotation and Learned Interpolation Ablation

We ablate the two key components of our physics-geometry synergy. Table 9 shows the impact of removing each.

Table 9: Camera rotation and learned interpolation ablation (4-scene average). Camera rotation encoding is the most impactful component.

Configuration	SSIM↑	PSNR↑	LPIPS↓
Full model	**0.9240**	**29.45**	**0.1550**
w/o camera rotation	0.9200	28.19	0.1570
w/o learned interpolation	0.9220	28.27	0.1570

Removing camera-rotation encoding causes a significant SSIM drop (-0.40%), confirming that transforming the global light direction into per-view coordinates via $\mathbf{R}_i$ is essential. Removing the learned interpolation causes a smaller drop (-0.20% SSIM), suggesting the attention-based interpolation primarily benefits test-time generalisation rather than training-set reconstruction.

J Training Schedule Ablation

Physics activation iteration. Our two-stage training activates the physics model at T_s=5,000. Earlier activation leads to unstable training due to noisy depth estimates; later activation under-utilises capacity (Table 10).

Total training iterations. 30,000 iterations are sufficient for convergence on the four benchmark scenes and D3 (18–68 images). Longer training does not improve metrics and can lead to overfitting.

Table 10: Physics activation timing (2-scene average: Curasao + IUI3).

Activation iter T_s	SSIM↑	PSNR↑	LPIPS↓
5,000 (ours)	**0.9307**	**30.02**	0.1787
10,000	0.9272	28.85	**0.1763**
15,000	0.9234	28.13	0.1792

Table 11: Total training iterations ablation. 30k is used for the four benchmark scenes and D3 (18–68 images); 60k for SaltPond (260 images).

Total iterations	SSIM↑	PSNR↑	LPIPS↓
30,000 (ours)	**0.9240**	**29.45**	**0.1550**
40,000	0.9151	27.11	0.1687

K Per-Scene Detailed Results

Table 12 expands on Table 1 with complete per-scene results.

Table 12: Complete per-scene results on the four SeaThru-NeRF scenes. Best results in **bold**.

Method	Curasao			IUI3-RedSea			Japanese-RedSea			Panama			Average		
	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS
3DGS	0.873	28.31	0.221	0.843	22.98	0.246	0.854	21.49	0.216	0.893	29.20	0.152	0.866	25.50	0.209
SeaSplat	0.898	29.70	0.189	0.869	26.85	0.213	0.872	22.57	0.180	0.895	28.16	0.148	0.884	26.82	0.183
SeaThru-NeRF	0.824	23.54	0.352	0.784	22.65	0.341	0.734	18.82	0.359	0.851	22.78	0.290	0.798	21.95	0.336
WaterSplatting	0.922	29.49	0.148	0.889	29.38	0.203	0.876	23.32	**0.152**	0.943	31.15	**0.104**	0.908	28.34	**0.152**
Ours	**0.954**	**33.22**	**0.139**	**0.913**	**29.52**	**0.195**	**0.883**	**23.85**	0.169	**0.945**	**31.20**	0.116	**0.924**	**29.45**	0.155

L PGMNetwork and EWA Configuration

PGMNetwork architecture. A two-layer MLP PGM : $\mathbb{R}^3 \to \mathbb{R}^3_+$ with hidden width 16 and activations ReLU (hidden) / Softplus (output, ensuring $\beta_D > 0$). Input is the 3D world-space position $\mathbf{x}_j$ of each Gaussian; output is the per-Gaussian attenuation $[\beta_{D,R}, \beta_{D,G}, \beta_{D,B}]$. The output layer is initialized to produce $\beta_D \approx 1.0$ everywhere, enabling smooth transition from the global baseline. Learning rate 10^{-3} is optimal; higher rates (5×10^{-3}) cause instability.

EWA blur strength. Table 13 shows the trade-off between aliasing reduction and sharpness for the EWA splatting blur strength s_{blur}. s_{blur}=0.3 provides the optimal balance.

Table 13: EWA blur strength ablation (4-scene average, config with SH medium + PGMNetwork).

Blur strength s_{blur}	SSIM↑	PSNR↑	LPIPS↓
0.0 (no EWA)	0.9210	28.00	0.1586
0.1	0.9215	28.25	**0.1570**
0.3 (ours)	**0.9240**	**29.45**	0.1550
0.5	0.9218	28.15	0.1590
1.0	0.9200	27.80	0.1620

M SaltPond Dataset Results

The SaltPond scene from the SeaSplat dataset [46] contains $200+$ training images, requiring longer training than the SeaThru-NeRF scenes (18–29 images each). Table 14 shows progressive optimization on this challenging scene; our final configuration surpasses SeaSplat across all three metrics.

Table 14: SaltPond optimization progression. Our method surpasses SeaSplat [46] on SSIM (0.747 vs. 0.730), PSNR (27.48 vs. 27.42 dB) and LPIPS (0.247 vs. 0.271).

Configuration	SSIM↑	PSNR↑	LPIPS↓
SeaSplat [46]	0.730	27.42	0.271
Ours (30k, baseline)	0.7399	26.79	0.2648
Ours (60k, baseline)	0.7410	27.26	0.2584
Ours (60k + λ_{dssim}=0.5)	0.7416	26.63	0.2582
Ours (60k + λ_{dssim}=0.68)	0.7467	26.62	0.2472
Ours (60k + embed=64)	0.7398	27.18	0.2566
Ours (60k + PGMNetwork)	0.7371	26.00	0.2689
Ours (final)	**0.747**	**27.48**	**0.247**

Key findings: (i) 60k iterations are essential for $200+$ image datasets ($+0.16\%$ SSIM over 30k); (ii) SaltPond requires higher λ_{dssim} than the 4-scene optimum (0.4), indicating dataset-dependent tuning; (iii) PGMNetwork *degrades* performance on SaltPond (uniform water), unlike the 4-scene benchmark where heterogeneous scenes benefit.

N Implementation Details

Appearance Embedding MLP.

- Input: $[\mathbf{e}_i;\ \bar{z};\ \tilde{\theta}_Z^{(i)}/90;\ \tilde{\theta}_A^{(i)}/360] \in \mathbb{R}^{35}$
- Architecture: $35 \to 64 \to 64 \to 2$, ReLU hidden, unbounded output (sigmoid clamping after)
- Initialization: last layer $\mathcal{N}(0, 0.01^2)$ weights, zero bias
- Learning rate: follows the Gaussian optimizer schedule

Embedding Interpolation Network.

- Encoder: $\mathbb{R}^{36} \to \mathbb{R}^{64}$ (2-layer MLP, ReLU)
- Multi-head attention: 4 heads, d_{head}=16
- Decoder: $\mathbb{R}^{64} \to \mathbb{R}^{32}$ (2-layer MLP, ReLU)
- Learning rate: 10^{-4}; leave-one-out loss weight λ_{interp}=0.5

Higher-Order Scattering Modulation (SH Medium).

- Input: SH basis $\text{SH}_3(\mathbf{v}) \in \mathbb{R}^{16}$ (degree-3)
- Architecture: $16 \to 64 \to 3$ (per-channel multiplicative modulation ΔB^c)
- Activation: ReLU hidden, output constrained around 1.0; learning rate 10^{-3}
- Modulation: $B^c \leftarrow B^c \cdot \Delta B^c$, preserving the physical single-scattering prior

Training configuration. $30{,}000$ iterations ($60{,}000$ for SaltPond); physics activation at iter $5{,}000$; full-image batch; Adam (Gaussians use the custom 3DGS schedule); single NVIDIA RTX 4090; 30–40 minutes per scene at 30k iterations. Learning rates: backscatter/attenuation 5×10^{-3}; SH medium MLP 10^{-3}; PGMNetwork 10^{-3}; interpolation network 10^{-4}.

Table 15: Final loss weights, determined through ablation.

Loss component	Weight
D-SSIM	$\lambda_{\text{dssim}} = 0.4$
ℓ_1 reconstruction	$1 - \lambda_{\text{dssim}} = 0.6$
Perceptual (VGG)	$\lambda_{\text{percep}} = 0.02$
Dark Channel Prior	$\lambda_{\text{DCP}} = 0.5$
Gray World	$\lambda_{\text{GW}} = 0.05$
White Balance	$\lambda_{\text{WB}} = 0.05$
Saturation	$\lambda_{\text{sat}} = 4.0$
Depth smoothness	$\lambda_{\text{depth}} = 1.0$
Interpolation	$\lambda_{\text{interp}} = 0.5$

O Global Sun Direction Parameterization

During a typical underwater dive session lasting 30–60 minutes, the sun's position changes negligibly, justifying a single pair of learnable angles (θ_Z^w, θ_A^w) shared across all training images. The per-image appearance embedding residuals $(\delta\theta_Z, \delta\theta_A)$ absorb any minor temporal variations or unmodeled photometric effects.

P Failure Cases and Limitations

Extremely turbid water. On scenes with high turbidity (e.g., JapaneseGardens-RedSea), our method achieves the best SSIM (0.883 vs. 0.876) and PSNR (23.85 vs. 23.32 dB), but the LPIPS gap to WaterSplatting (0.169 vs. 0.152) is larger than on other scenes. Extreme scattering pushes our single-scattering physical model toward its limits, where unconstrained learning-based methods may better capture perceptual quality.

Spatially uniform vs. heterogeneous water. The PGMNetwork is designed to capture spatial variations in medium attenuation. On its own, it improves the heterogeneous Curasao scene but can introduce optimization noise elsewhere, most strongly in uniform-turbidity scenes (IUI3, SaltPond); paired with EWA, it improves all four benchmark scenes (Table 5). An automated scene-heterogeneity detection metric to adaptively gate PGMNetwork activation is left for future work.

Dependence on SfM quality. Our physics-geometry synergy assumes accurate camera poses from SfM. Poor SfM results lead to incorrect base illumination angles and degraded reconstruction.

Limited phase function. We use the Fournier-Forand phase function which assumes spherical particles. Real underwater environments may contain non-spherical particles (plankton, sediment) with different scattering properties.

Q Training Convergence

We trained three variants on Curasao to isolate the optimization effect of physics integration: *w/o physics* (vanilla 3DGS reconstructing the underwater image directly), *Physics-only* (image formation $I = J \odot T + B$ without appearance embedding or learned interpolation), and *Full model*. The physical decomposition is activated at iteration $5{,}000$ for the latter two. Figure 9 reports the smoothed (EMA α=0.995) reconstruction ℓ_1 loss across $30{,}000$ iterations. All three variants reach a comparable plateau on the underwater fitting objective: this is by design, since every variant minimises the same ℓ_1 against the captured underwater views and the formation model is deliberately invertible. The physics-equipped runs exhibit a brief loss bump near iteration $5{,}000$ when the medium parameters become trainable, then re-converge below the vanilla baseline. The benefit of physics is therefore not a smaller fitting loss on the underwater image, but a structured decomposition that yields a clean radiance J usable for restoration and downstream geometry—neither of which is captured by the underwater ℓ_1 in this figure.

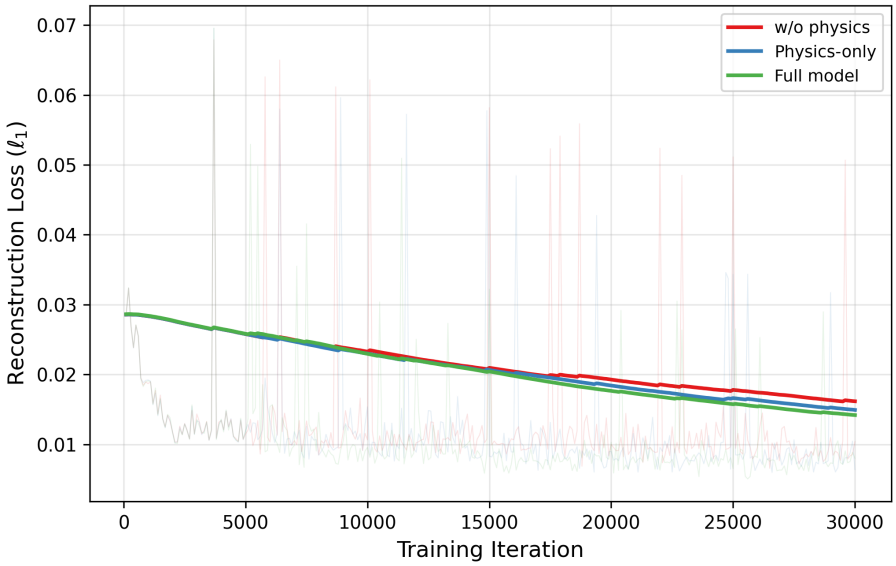

Figure 9: Reconstruction ℓ_1 loss on Curasao (EMA-smoothed). Faint lines show the unsmoothed signal. The dashed boundary at $5{,}000$ iterations marks the activation of the physics decomposition for the latter two variants.

R Synthetic Benchmark with Ground Truth

Real underwater scenes have no clean reference, so restoration, geometry, and physical parameters cannot be scored directly. We therefore build a synthetic benchmark from two scenes captured in air. Dr. Johnson from Deep Blending [11] has 263 views (230/33 train/test). Truck from Tanks and Temples [17] has 251 views (219/32 train/test).

Protocol. A clean 3DGS reconstruction of each scene renders the clean radiance J and the depth z for every view. Each clean view is then degraded with fixed medium parameters. Because the water is applied to one 3D reconstruction, the same 3D point receives the same water thickness in every view. Every method trains only on the degraded images and is scored against the held-out clean images. Unsupervised decomposition cannot fix the per-channel scale of J, since only the product $\beta^c z$ is observable for channel c. Restored images are therefore aligned to the ground truth by one per-channel scale before computing metrics.

Two degradation settings. *Setting A (depth-only)* follows the image formation model of prior work [2]: $I = J\, e^{-\beta_D z} + B_\infty\,(1 - e^{-\beta_B z})$. It excludes our view- and position-dependent term, so it does not favor our method. *Setting B (view-dependent)* applies our forward model with a per-pixel scattering angle computed from each camera rotation. It matches our model by construction, so we report it as a mechanism study rather than a neutral benchmark.

Restoration. Table 16 reports restoration under Setting A. Both physics-based methods beat vanilla 3DGS by 3.2–4.4 dB, so an explicit medium model, not reconstruction capacity, drives restoration. Our method improves over SeaSplat by 0.55 dB on Dr. Johnson and 0.78 dB on Truck.

Table 16: Restoration on the synthetic benchmark, Setting A (PSNR / SSIM, per-channel aligned).

Method	Dr. Johnson	Truck
3DGS	24.09 / 0.849	21.50 / 0.882
SeaSplat [46]	27.90 / 0.910	24.72 / 0.927
Ours	**28.45 / 0.920**	**25.50 / 0.935**

S Depth Accuracy on the Synthetic Benchmark

Each synthetic view carries a ground-truth depth map from the same generator. We use standard scale-invariant metrics. Each prediction is aligned to the ground truth by one median scale factor per view, then scored with AbsRel, RMSE, and RMSE-log. Pixels on the far NaN-fill plateau are excluded. All methods use the same evaluation code and one configuration each.

Table 17: Depth accuracy on the synthetic benchmark. Setting A is neutral; Setting B is aligned with our model by construction. Bold marks the best per scene and metric.

		Truck			**Dr. Johnson**		
	Method	AbsRel↓	RMSE↓	RMSE-log↓	AbsRel↓	RMSE↓	RMSE-log↓
A	3DGS	0.0719	0.075	0.1179	0.0775	0.088	0.1133
	SeaSplat	0.0628	0.070	0.1277	0.0384	0.039	0.0513
	Ours	**0.0511**	**0.058**	**0.1033**	**0.0311**	**0.034**	**0.0471**
B	3DGS	0.1099	0.100	0.1540	0.2237	0.209	0.3089
	SeaSplat	0.0918	0.110	0.2035	0.1898	0.174	0.2510
	Ours	**0.0655**	**0.080**	**0.1319**	**0.0945**	**0.095**	**0.1454**

Under the neutral Setting A, our method has the lowest error on all three metrics in both scenes. Under Setting B, angle-agnostic baselines degrade much more: averaged over the two scenes, AbsRel is 0.0800 for our method, 0.1408 for SeaSplat, and 0.1668 for 3DGS.

Caveat. The reference depth is rendered from a clean 3DGS reconstruction, so it inherits the alpha-composited depth convention of 3DGS. Vanilla 3DGS shares this inductive bias, whereas a medium model changes the opacity distribution. The comparison is therefore biased toward the 3DGS baseline.

T Recovery of Physical Parameters

We degrade the clean Truck scene with our forward model (Setting B) and known parameters: sun direction θ_Z^w=50°, θ_A^w=180°; attenuation β_D=$[0.35, 0.25, 0.20]$ m^{-1}; Fournier–Forand parameters μ=3.7, n=1.34; backscatter amplitude B_s=$[3, 6, 12]$. Training uses only the degraded images.

Attenuation and backscatter. The recovered coefficients keep the correct spectral structure: attenuation is strongest in red and backscatter is strongest in blue. Starting from B_s=$[5, 10, 20]$, optimization reaches about $[3.3, 5.4, 11.0]$, close to the ground truth. Absolute attenuation is identifiable only up to a global per-channel scale, because the observations constrain the product $\beta^c z$.

Sun direction. In the primary configuration, the recovered zenith and azimuth differ from the ground truth by $6.0°$ and $4.3°$. Additional initializations show that absolute sun-direction recovery becomes less stable when the initialization is far from the ground truth. We therefore treat the absolute sun direction as an approximate estimate rather than an exactly identifiable quantity. The view dependence of ψ comes from the camera rotations and does not rely on recovering the absolute direction exactly.

Operating range. Beyond an optical thickness of about 2.8, the inverse problem degrades measurably. This bounds the operating range of the method; the benchmark scenes, captured in shallow sunlit water, lie within it.

U Robustness to Pose Errors

We perturb all Curasao training poses with random-axis rotation noise of standard deviation $\sigma \in \{0.5°, 1.0°, 2.0°\}$, plus translation noise of 0.5% of the camera distance per degree. Test poses stay clean so that evaluation targets remain aligned. Table 18 shows a clear sensitivity: $0.5°$ of noise already lowers PSNR by 7.8 dB. Accurate poses are a prerequisite, as for the 3DGS family in general: projection error grows directly with pose error, and our medium model also reads the camera rotation. The perturbation is deliberately harsh, since it hits every training camera at once, whereas real SfM errors are correlated and partly absorbed by bundle adjustment.

Table 18: Training-pose noise on Curasao.

Rotation noise	SSIM↑	PSNR↑	LPIPS↓
Clean	0.954	33.22	0.139
σ=0.5°	0.714	25.47	0.368
σ=1.0°	0.666	24.91	0.394
σ=2.0°	0.616	23.38	0.424

V Comparison with Plenodium

Plenodium [44] released code. We evaluate it on the same datasets with identical train/test splits and the same evaluation code.

Table 19: Same-split comparison with Plenodium on the four benchmark scenes (SSIM / PSNR).

Scene	Plenodium [44]	Ours
Curasao	**0.9559 / 34.35**	0.954 / 33.22
IUI3-RedSea	0.9065 / 29.09	**0.913 / 29.52**
JapaneseGardens	**0.8914 / 24.75**	0.883 / 23.85
Panama	0.9388 / **31.53**	**0.945** / 31.20
Average	0.9232 / **29.93**	**0.924** / 29.45

Table 20: Comparison with Plenodium on SaltPond.

Method	SSIM↑	PSNR↑	LPIPS↓
Plenodium [44]	0.648	26.88	0.392
Ours	**0.747**	**27.48**	**0.247**

On the four benchmark scenes the two methods are close: average SSIM is 0.924 vs. 0.9232, and Plenodium leads average PSNR by 0.48 dB. On SaltPond, which has a different water condition and capture distribution, our method reaches 0.747 SSIM and Plenodium 0.648. Plenodium's plenoptic medium is a flexible learned representation well suited to view synthesis, while our formation model additionally outputs the medium coefficients and the illumination direction. We regard the two approaches as complementary.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: All four claimed contributions (physics-geometry synergy, physically grounded medium modeling, cross-view attention, strong empirical performance) are supported by experiments in Section 4 and ablations in Section 4.3 and the appendix.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Section 5 discusses the single-sun assumption, the multi-scattering regime, unprobed phase-function peaks, the approximate identifiability of the sun direction and attenuation scale, the operating range in optical thickness, and the dependence on accurate SfM poses (quantified in the appendix).

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [N/A]

Justification: The paper does not include formal theoretical results or proofs.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: Section 4.1 details all hyperparameters, training schedule, and architecture choices. The appendix provides loss formulations, network architectures, and regularization weights. All datasets used are publicly available.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No]

   Justification: Code is not included with the submission. The training code, per-scene configurations, and the synthetic benchmark (generation scripts and evaluation protocol) will be publicly released. All datasets used (SeaThru-NeRF, SaltPond, D3, Deep Blending, Tanks and Temples) are publicly available and properly cited.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Section 4.1 and the appendix provide complete training details including optimizer, learning rates, loss weights, data splits (llffhold=8), and hardware.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [N/A]

   Justification: Each scene requires 30–40 minutes of training; running multi-seed evaluation across 6 scenes, 4 baselines, and all ablations is computationally infeasible at this scale. We follow the standard 3DGS/NeRF underwater benchmark protocol on fixed train/test splits, consistent with prior work (SeaSplat, WaterSplatting, SeaThru-NeRF).

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.

- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Section 4.1 specifies the GPU (NVIDIA RTX 4090) and training iterations (30k). The appendix reports per-scene training time (30–40 minutes) and VRAM usage.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research uses publicly available datasets, involves no human subjects, and the underwater reconstruction application has no foreseeable negative societal impact.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

Justification: Underwater 3D reconstruction is a foundational technique for marine science and does not raise significant negative societal concerns.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: The method reconstructs underwater scenes from existing images and poses no risk of misuse.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: All datasets and baselines are properly cited. The codebase builds on 3DGS which is released under an open-source license.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.

- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: Code and trained models will be released with documentation upon acceptance.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: No human subjects or crowdsourcing involved.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: No human subjects research.

    Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: LLMs are not used as a component of the proposed method.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.