AquaSplatting: A Hybrid 3D Representation for Robust Underwater Scene Reconstruction via Dual-Branch Rendering

Jiangbei Hu[1]*, Haobo Wang[2]*, Baixin Xu[3], Nan Ding[2], Zhimao Lu[2]†, Na Lei[1], Ying He[3]

[1]School of Software, Dalian University of Technology
[2] School of Computer Science, Dalian University of Technology
[3] College of Computing and Data Science, Nanyang Technological University

Abstract

3D Gaussian Splatting (3DGS) excels at real-time rendering of standard scenes, however, it struggles to reconstruct underwater environments due to severe challenges such as light scattering, color attenuation, and sparse coverage of Gaussian kernels in far-field aqueous regions. To address this, we introduce AquaSplatting, a hybrid framework that combines explicit and implicit modeling methods for robust underwater scene reconstruction. Our dual-branch architecture employs 3DGS in a geometry-guided branch to model solid surfaces like the seabed, while a medium-aware branch uses a compact, view-dependent MLP to represent volumetric water effects. Furthermore, a neural underwater hybrid rendering mechanism adaptively fuses these two representations based on accumulated opacity. Thanks to this dual-branch framework, our method can also synthesize restored images without water medium. To enhance efficiency, our proposed engagement-based pruning (EBP) strategy quantifies each Gaussian's contribution by accumulating its image-space gradients over multiple frames, enabling the principled removal of primitives with negligible impact. The entire framework is optimized using a comprehensive loss function that integrates photometric, exposure, semantic, and depth priors to maximize visual fidelity. Experiments on challenging underwater datasets demonstrate that AquaSplatting achieves the state-of-the-art in reconstruction quality surpassing prior methods while maintaining real-time performance.

Code — https://github.com/ChrisWanghb/AquaSplatting

Introduction

Reconstructing 3D underwater scenes has broad applications, from marine robotics to immersive virtual reality. However, optical effects such as absorption and scattering severely degrade underwater imaging quality, making accurate reconstruction a challenging task. As red light attenuates rapidly, color distortion and contrast reduction arise, while backscattering directs ambient light toward the camera. These adverse effects intensify with depth, linking object appearance to water properties and complicating the joint modeling of geometry and medium. An ideal representation of underwater environments must capture both solid structures (i.e. objects within water) and the surrounding water medium. Objects provide key cues about the appearance and geometry of the scene, while the water medium encodes critical environmental parameters. However, inherent challenges such as distance-dependent attenuation and wavelength-selective scattering make it difficult to accurately capture both components.

NeRF-based methods (Sethuraman, Ramanagopal, and Skinner 2023; Levy et al. 2023) have demonstrated the ability to jointly reconstruct objects and aquatic environments. SeaThru-NeRF (Levy et al. 2023) first incorporates the scattering-aware physical model into NeRF, allowing high-fidelity novel view synthesis and joint medium-free color-geometry recovery in underwater scenes. However, NeRF-based methods suffer from slow training speeds and cannot render in real time. 3DGS (Kerbl et al. 2023) enables real-time, photorealistic 3D reconstruction, but standard 3DGS implementations are designed for clear atmospheric conditions and ignore the effects of the surrounding medium, such as water. To tackle this, Watersplatting (Li et al. 2024) proposed to integrate rasterization with medium-aware modeling, which uses 3DGS for explicit geometry representation and an independent volumetric ray-marching branch to capture the scattering medium. However, this approach often produces noticeable "blobby" artifacts, especially in scenes with large water regions. This issue stems from the sparsity of Gaussians in distant background areas, leading to a degraded modeling capability for far-field water.

To overcome these challenges, we introduce *AquaSplatting*, a novel framework for reconstructing underwater scenes based on a hybrid 3D representation. Our method is built upon a dual-branch architecture: The geometry-guided branch employs 3DGS to explicitly model solid surfaces, such as the seabed and other submerged objects. The medium-aware branch implicitly represents the water medium, which employs a compact Multi-Layer Perceptron (MLP) to predict color and opacity for non-seafloor regions. This synergistic design allows our method to create a comprehensive scene representation, combining explicit foreground geometry with a continuous, implicit model of the surrounding water. Furthermore, we propose a neural underwater hybrid rendering mechanism that adaptively integrates

*These authors contributed equally.
†Corresponding author : lzm@dlut.edu.cn

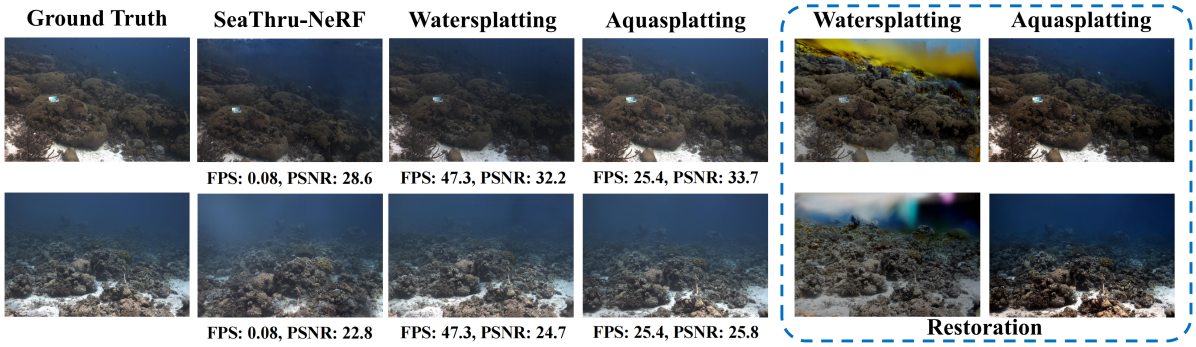


Figure 1: We introduce AquaSplatting, a novel framework for efficient underwater scene reconstruction that delivers state-of-the-art visual quality, while simultaneously enabling the high-fidelity restoration of the scene to its clear, water-free appearance.

the output from two branches. Based on the accumulated opacity, our renderer ensures seamless transitions between near-field objects and far-field water, significantly boosting visual fidelity. In addition, we introduce an engagement-based pruning (EBP) strategy that dynamically eliminates Gaussians with negligible contributions to the final render, enhancing both efficiency and training stability. Thanks to this dual-branch framework, our method can not only synthesize novel views of the underwater scene, but also generate restored, water-free images by disabling the effects of the medium branch. To validate our approach, we conducted extensive experiments on the publicly available underwater datasets (Levy et al. 2023; Tang et al. 2024). The results demonstrate that AquaSplatting outperforms existing baselines in both visual fidelity and real-time rendering speed. Our contributions are summarized as follows:

- We propose AquaSplatting, a novel hybrid framework that simultaneously models explicit scene geometry and the implicit water medium using a synergistic dual-branch architecture.
- We design an engagement-based pruning (EBP) strategy that dynamically removes ineffective Gaussians by quantifying their contribution to the final rendering, thus enhancing efficiency.
- Our method delivers a high visual fidelity reconstruction at real-time rendering speed, while also providing the capability to restore the scene to its clear, water-free appearance.

Related Work

Underwater scene reconstruction. Underwater scene representation requires the joint modeling of scene geometry and medium properties, with attenuation effects already extensively studied in 2D representations. The goal of underwater image restoration is to recover visual perception from degraded images (Yuan, Cai, and Cao 2021). Existing methods can be broadly categorized into two types: direct image enhancement approaches (Li et al. 2021; Rao et al. 2023), and model-based methods that use the Underwater Imaging Model to estimate attenuation coefficients for image restoration (Cong et al. 2023; Guo et al. 2025a). However, due to the ill-posed nature of single-image restoration, these methods often rely on strong prior assumptions for parameter estimation (Akkaynak and Treibitz 2019). In contrast, multi-view 3D reconstruction methods help alleviate this ill-posedness and improve the accuracy of parameter estimation (Boittiaux et al. 2024).

NeRF-based methods. NeRF-based methods have been widely adopted for underwater scene understanding because of their implicit volumetric modeling capabilities. ScatterNeRF (Ramazzina et al. 2023) and DehazeNeRF (Chen et al. 2024) incorporate atmospheric scattering into NeRF for foggy scene reconstruction and joint color–geometry correction. WaterNeRF (Sethuraman, Ramanagopal, and Skinner 2023) estimates medium parameters and applies histogram equalization to enhance underwater rendering. SeaThru-NeRF (Levy et al. 2023) integrates the underwater image formation model into NeRF's rendering, enabling domain adaptation. WaterHE-NeRF (Zhou et al. 2025) introduces a Retinex-based water-ray field and a joint loss to improve color fidelity. U2NeRF (Gupta, Mitra et al. 2024) employs an unsupervised Transformer to disentangle radiance, transmission, and scattering. NeuroPump (Guo et al. 2025b) jointly models refraction and scattering in a self-supervised framework and contributes the first real-world 360° underwater dataset with ground-truth. Despite their effectiveness, these approaches inherit NeRF's limitations in training cost and rendering latency, hindering real-time deployment.

3DGS-based methods. RestorGS (Qiao et al. 2025) pioneers a unified depth-aware Gaussian splatting framework that disentangles appearance and leverages depth-guided physical modeling for real-time restoration of diverse 3D degraded scenes, such as underwater, nighttime, and hazy conditions. While Gaussian Splashing (Mualem et al. 2024) effectively synthesizes dynamic water surfaces using 3D Gaussian primitives, it focuses on surface-level modeling and lacks support for complex underwater light transport and volumetric effects. UW-GS (Wang et al. 2025a) incorporates an appearance-sensitive model and physics-driven density control to handle underwater scattering, absorption, and motion interference. Seasplat (Yang, Leonard, and Girdhar 2024) enhances true color recovery and depth estimation by combining multiple physically grounded losses, including

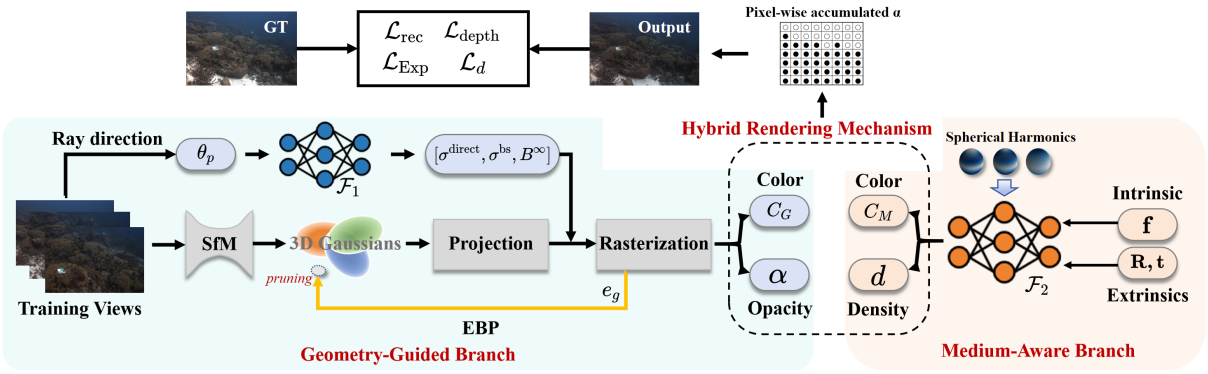


Figure 2: **Pipeline.** Our framework employs a dual-branch architecture for underwater scenes: a geometry-guided branch uses 3DGS to model explicit structures, while a media-aware MLP branch captures volumetric water effects. A neural rendering mechanism then adaptively fuses these branches based on accumulated transmittance. Finally, the model is optimized via a composite loss function integrating photometric, depth priors, exposure, and semantic for maximum visual fidelity.

backscatter, gray-world, and depth-weighted reconstruction, while preserving real-time rendering efficiency. Watersplatting (Li et al. 2024) introduces a dual-branch framework combining 3DGS with a separate volumetric field for the scattering medium, mitigating the inefficiencies of NeRF. However, it struggles to accurately reconstruct distant water backgrounds due to sparse point representations. In contrast, our method leverages a hybrid explicit-implicit representation, combining the geometric accuracy of Gaussian splats with the flexibility of neural volumetric modeling. This enables more balanced foreground-background reconstruction while maintaining high rendering efficiency.

Method

Problem Formulation

Underwater scenes are governed by complex optical phenomena, primarily light attenuation and scattering. Based on the revised underwater imaging model(Akkaynak and Treibitz 2018), the final image I is separated into a direct and backscatter component as

$$I \approx J \cdot \mathcal{A}^{\text{direct}}(\sigma^{\text{direct}}, z) + B^{\infty} \cdot (1 - \mathcal{A}^{\text{bs}}(\sigma^{\text{bs}}, z)),$$

where $\mathcal{A}(\sigma, z) = e^{-\sigma z}$, J denotes the water-free image of a scene at depth z, B^{∞} represents the background color of the water at an infinite distance. σ^{direct} and σ^{bs} are attenuation coefficients for the direct and backscatter components of the image that indicate the effect of water medium on the color.

A key limitation in prior methods is the assumption of a spectrally homogeneous water medium (Kim, Lee, and Lee 2024), where optical coefficients B^{∞}, σ^{direct}, and σ^{bs} are constant across color channels. This ignores the inherent wavelength dependence of light propagation, leading to inaccurate of color attenuation and scattering. Thus, we adopt a wavelength-aware model that independently defines the optical coefficients in each RGB channel. This results in nine learnable variables, $B^{\infty} = (B^{\infty}_r, B^{\infty}_g, B^{\infty}_b), \sigma^{\text{direct}} = (\sigma^{\text{direct}}_r, \sigma^{\text{direct}}_g, \sigma^{\text{direct}}_b), \sigma^{\text{bs}} = (\sigma^{\text{bs}}_r, \sigma^{\text{bs}}_g, \sigma^{\text{bs}}_b)$. By doing so, our model can represent the distinct behavior of each color channel, enabling a faithful reconstruction of underwater scenes.

Fig. 2 illustrates the pipeline of our proposed framework. Our method reconstructs underwater scenes based on a dual-branch architecture. A geometry-guided branch employs 3DGS to explicitly model scene geometry by predicting key parameters such as $B^{\infty}, \sigma^{\text{direct}}, \sigma^{\text{bs}}$, color C_G, and opacity α. Concurrently, a medium-aware branch represents the volumetric water medium using an implicit MLP that outputs auxiliary color C_M and density σ. A neural rendering mechanism then adaptively fuses these branches based on accumulated transmittance. The model is optimized with a composite loss function that combines photometric, depth priors, exposure, and semantics.

Dual-Branch Architecture

Geometry-Guided Branch. The geometry-guided branch of our framework is responsible for modeling solid surfaces within the scene, such as the seabed, rocks and other submerged objects. To this end, we employ 3DGS, which represents the scene's geometry as a collection of 3D anisotropic Gaussians. Each Gaussian is parameterized by its mean position and covariance matrix as the standard 3DGS method. The final color C_G is then synthesized via a differentiable tile-based rasterizer as

$$C_G = \sum_{i=1}^{N} \alpha_i c_i \prod_{j=1}^{i-1} (1 - \alpha_j),$$

where α_i and c_i is the opacity and color associated with the i-th Gaussian. The products are over the ordered list of Gaussians that overlap a given pixel. Please refer to 3DGS (Kerbl et al. 2023) for more details.

Owing to the unbounded rendering formulation of 3DGS, the expected pixel color along a camera ray $r(s) = \mathbf{o} + s\mathbf{d}$, which extends towards infinity, can be expressed as $C(r) = \int_0^{\infty} T(s)\sigma(s)c(s)ds$ (Kajiya and Von Herzen 1984). We further incorporate underwater light transport by modeling the transmittance from the first to the i-th Gaussian as:

$$T_i(s) = T_i^{\text{obj}} T^{w}(s) = \Big(\prod_{j=1}^{i-1} (1 - \alpha_j) \Big) \Big(e^{-\sigma^{\text{bs}} s} \Big),$$

T_i^{obj} denotes the occlusion term from the first $i-1$ Gaussians, while $T^w(s)$ models the light attenuation in water. The final color is then computed as:

$$C_G(r) \approx \sum_{i=1}^{N} C_i^{\text{obj}}(r) + \sum_{i=1}^{N} C_i^w(r),$$

where C^{obj} and C^w denote the color of objects and medium, respectively. We employ the direct extinction coefficient σ^{direct} and the medium backscattering coefficient σ^{bs} to model $C_i^{\text{obj}}(r)$ and $C_i^w(r)$, respectively.

$$C_i^{\text{obj}}(r) = T_i^{\text{obj}} T^w(s_i) \alpha_i c_i = T_i^{\text{obj}} \alpha_i c_i e^{-\sigma^{\text{direct}} s_i},$$

$$\begin{aligned} C_i^w(r) &= \int_{s_{i-1}}^{s_i} T^w(s)\, \sigma^w B^\infty \, ds \\ &= B^\infty \left(e^{-\sigma^{\text{bs}} s_{i-1}} - e^{-\sigma^{\text{bs}} s_i} \right). \end{aligned}$$

where σ^w is the water's attenuation coefficient, and B^∞ is the background color due to scattering. To handle view-dependent variation, we employ a Multi-Layer Perceptron (MLP) to predict σ^{direct}, σ^{bs} and B^∞ from the view direction θ_p

$$(\sigma^{\text{direct}}, \sigma^{\text{bs}}, B^\infty) = \mathcal{F}_1(\theta_p'),$$

where θ_p' is the positional encoding of θ_p, $\mathcal{F}_1$ is an MLP with the SoftPlus activation function.

Medium-Aware Branch. To complement the explicit geometry modeling, we introduce a *medium-aware branch* to represent non-seabed regions and distant volumetric content. This branch models scattering-aware color and density in a view-dependent manner via an MLP $\mathcal{F}_2$. Given the viewing pose $(\mathbf{R}, \mathbf{t})$ and the intrinsic parameter $\mathbf{f}$, it outputs the color C_M and volume density d for each sampled point:

$$C_M, d = \mathcal{F}_2(\mathbf{R}, \mathbf{t}, \mathbf{f}).$$

To better account for angular appearance variations induced by light scattering and absorption, we incorporate Spherical Harmonics (SH) to encode the viewing direction. This provides a compact and continuous representation of directional dependencies, enabling more accurate color modeling in distant or low-visibility regions (Rahaman et al. 2019). The SH-encoded direction is concatenated with the MLP output to yield the final radiance estimate.

Neural Underwater Hybrid Rendering Mechanism

To effectively model complex underwater scenes with varying depths and media properties, we propose a neural underwater hybrid rendering mechanism to combine the two branches: the geometry-guided branch reconstructs the detailed, local color C_G of near-field solid surfaces, while the medium-aware branch models the ambient, global color C_M of the far-field water medium. Directly blending these outputs can cause color artifacts and geometric inconsistencies. To mitigate this, we introduce an opacity-gated fusion strategy that adaptively incorporates medium-aware colors based on the accumulated opacity. During rasterization, color blending stops when accumulated opacity exceeds a threshold, indicating an opaque surface. For rays passing through distant regions, contributions from the medium-aware branch are retained. Specifically, the cumulative opacity α of the geometry-guided branch is computed during Gaussian splat filtering, sorting, and blending as $\alpha = 1 - \prod_{j=1}^{N}(1-\alpha_j)$. The final output color $\hat{C}$ is computed as

$$\hat{C} = \begin{cases} C_G, & \alpha > \gamma, \\ C_G + (1-\alpha) \cdot C_M \cdot d, & \alpha \le \gamma. \end{cases}$$

If the accumulated opacity α exceeds the threshold γ, the density of explicit Gaussian splats is sufficient to faithfully represent the local geometry and color, and the system relies solely on the geometry-guided branch. In contrast, when $\alpha \le \gamma$, the explicit representation is sparse and the medium-aware branch is introduced to supplement color estimation, particularly for distant water volumes and attenuation effects due to the medium.

This mechanism ensures that the implicit medium-aware branch complements the explicit branch only when necessary, improving the overall rendering quality while reducing artifacts. We provide more details about the fusion process in Algorithm 1 in the supplementary material. In particular, the per-pixel opacity α is computed on-the-fly during rasterization and can be directly extracted from the rendering pipeline without additional computational overhead.

Engagement-Based Pruning Strategy

While the standard 3DGS framework periodically resets the opacity of Gaussians to maintain model compactness, this abrupt intervention often destabilizes the training process. To address this issue, we propose a novel engagement-based pruning (EBP) strategy, which systematically removes Gaussians based on their quantified impact on the final rendered image. Specifically, we assign an engagement score e_g to each Gaussian as

$$e_g = \sum_{t \in \mathcal{N}_T} \frac{1}{h \cdot w} \left\| \frac{\partial I_{i,j}^t}{\partial x_g^t} \right\|_2^2,$$

where $I_{i,j}^t$ denotes the color rendered in pixel (i, j) in frame t, h and w is the height and width of the image, $x_g^{(t)}$ represents the position of the Gaussian point in frame t, and $\mathcal{N}_T$ is the set of multiple frames considered for the evaluation of engagement. In our implementation, we evaluate engagement scores every $T = 200$ training iterations and prune any Gaussians whose score e_g falls below a threshold k (we use $k = 10^{-6}$ in our experiments). This EBP strategy effectively suppresses the influence of negligible Gaussians while avoiding the training instability associated with the abrupt opacity resets used in the standard 3DGS framework.

Optimization

To guide our model towards a faithful reconstruction of the underwater scene, our framework is optimized using a comprehensive loss function that integrates photometric, semantic, exposure, and depth priors.

Reconstruction loss. Similarly to the standard 3DGS framework (Kerbl et al. 2023), the principal loss in image reconstruction is defined as a weighted sum of the L1 loss of color

at pixel level and the perceptual dissimilarity (D-SSIM):

$$\mathcal{L}_{\text{rec}} = (1 - \lambda_1)\mathcal{L}_1 + \lambda_1 \mathcal{L}_{\text{D-SSIM}}$$

Depth supervision via alpha-composited rendering. Underwater light attenuation severely degrades visibility, compromising the depth accuracy of distant structures. To counteract this, we directly supervise the depth map rendered from the 3D Gaussians via alpha compositing as

$$\hat{D} = \sum_{i \in \mathcal{N}} z_i \alpha_i' \prod_{j=1}^{i-1} (1 - \alpha_j')$$

where z_i is the depth of the i-th Gaussian, and α_i' is the opacity. To guide optimization, we use the pre-trained DepthAnything V2 model (Yang et al. 2024) to generate pseudo-ground-truth depth maps D_{pseudo}. We then formulate a depth consistency loss as:

$$\mathcal{L}_{\text{depth}} = \|D_{\text{pseudo}} - \hat{D}\|_1.$$

Exposure loss for illumination robustness. To handle various water conditions and ensure consistent brightness, we introduce an exposure regularization term that constrains the mean exposure of local image patches to a desired target value E as

$$\mathcal{L}_{\text{Exp}} = \frac{1}{N} \sum_{k=1}^{N} |L_k - E|,$$

where L_k is the mean luminance of the k-th non-overlapping region and N is the total number of such regions.

Semantic-guided opacity regularization. To enforce a robust decomposition of the underwater scene into seabed and non-seabed regions across our two branches, we generate supervisory masks using Grounded SAM (Ren et al. 2024). This process yields a binary mask, $\text{Mask}(i, j)$, for each image, where a value of 1 indicates a non-seabed region and 0 corresponds to the seabed. This mask guides the optimization of density d in the medium-aware branch, with the corresponding loss function defined as

$$\mathcal{L}_d = \sum_{i,j} \|\text{Mask}(i, j) - d(i, j)\|_2^2.$$

It is worth noting that while this semantic supervision from Grounded SAM enhances reconstruction quality, it is not a critical dependency. Our core framework still achieves strong performance without this component, a claim we validate in our ablation studies.

Experiment

Datasets. We validate our approach on two underwater datasets: (1) **SeaThru-NeRF** (Levy et al. 2023): A real-world dataset with four forward-facing scenes captured in diverse aquatic conditions. Its unbounded camera distances present a significant challenge for neural reconstruction. (2) **Underwater in the Wild (U-IW)** (Tang et al. 2024): A large-scale dataset build from internet videos, featuring extended and full 360°camera trajectories that provide rich spatial context for evaluation. Instead of relying on COLMAP (Schönberger and Frahm 2016), we adopt VGGT (Wang et al. 2025b), a fast feed-forward neural network that estimates camera poses, sparse point clouds, and 3D point tracks in just a few seconds. This approach provides our model with the accurate and efficient geometric initialization required for high-quality, real-time rendering.

Method	PSNR ↑	SSIM ↑	LPIPS ↓	FPS ↑	Time ↓
SeaThru-NeRF	26.45/19.11	0.815/0.639	0.247/0.402	0.55	2.7
HoGS	25.51/27.46	0.866/0.889	0.208/0.160	146.34	0.6
SeaSplat	27.39/27.21	0.865/0.887	0.209/0.161	42.54	1.6
UW-GS	28.81/27.00	0.918/0.891	0.144/0.153	38.27	0.8
Watersplatting	29.61/25.35	0.915/0.884	0.140/0.168	35.80	0.5
Ours	**30.32/28.25**	**0.933/0.906**	**0.136/0.152**	25.37	0.8

Table 1: Quantitative evaluation of existing methods. We evaluate visual performance by PSNR, SSIM, and LPIPS on two datasets (SeaThru-NeRF/U-IW) respectively. Efficiency is demonstrated by rendering speed (FPS) and training time (Time, in hours) across all test scenes.

Implementation Details. In our implementations, 3D Gaussian rendering is accelerated using a tile-based rasterizer. We employ Fourier positional encoding with $L = 10$ for camera viewpoints and $L = 4$ for ray directions. Following 3DGS, the ray termination threshold γ for opacity accumulation is set to 0.9999. The loss function combines multiple objectives, with weights set as $\lambda_1 = 0.2$ (reconstruction loss), $\lambda_d = 0.4$ (density regularization), $\lambda_{\text{Exp}} = 0.8$ (exposure loss) and $\lambda_{depth} = 0.1$ (depth supervision), including a term of water mask supervision. All experiments were conducted on a single NVIDIA RTX 4090 GPU to ensure a fair comparison of computational performance. For all methods, we report Frames Per Second (FPS) and total training time. To ensure robust and reproducible results, all reported metrics are the average of five independent runs.

Experimental Results

Reconstruction quality. As shown in Fig. 3, our method markedly enhances reconstruction fidelity by reducing visual artifacts and more accurately recovering distant geometric details. In contrast, existing approaches often exhibit pronounced "floating" artifacts and fail to reconstruct distant regions. Our framework maintains structural consistency across varying depths, rendering complex features with high fidelity in scenes like "J.G RedSea" and "Turtle". This consistent, high-quality performance across various datasets confirms the robustness of our method to challenging underwater conditions and scene complexities.

Quantitative evaluations. Our method consistently outperforms the state-of-the-art approaches in two datasets. As shown by the average metrics in Tab. 1, our approach achieves superior performance, securing the highest PSNR and SSIM scores while recording the lowest LPIPS values (Zhang et al. 2018). We also provide per-scene evaluations in Tab. S1 and Tab. S2 in the supplementary materials. The scene-wise evaluations further demonstrate the robustness of our approach across a wide range of underwater scenarios. Although our underwater medium model-

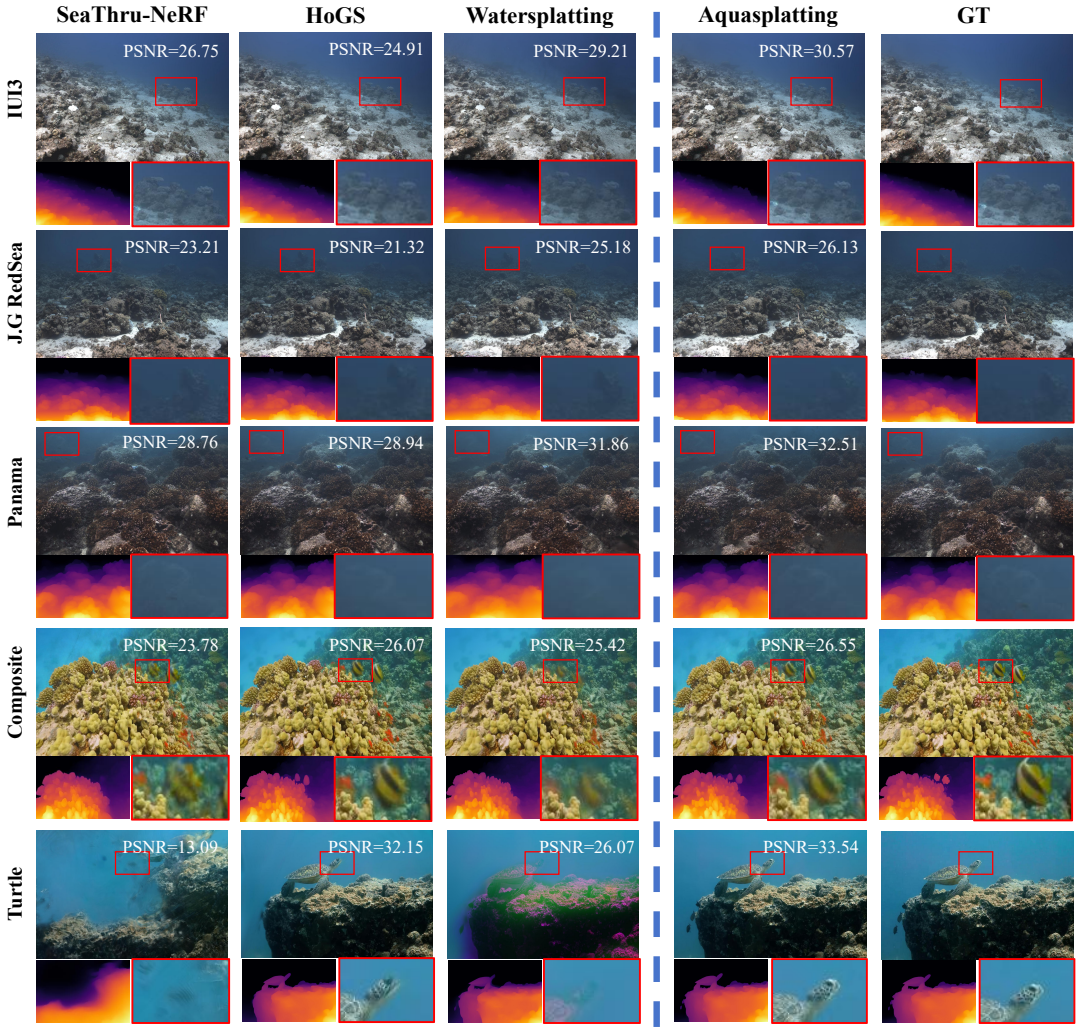


Figure 3: Qualitative comparison (Zoom-in for details). The depth maps are generated by DepthAnything v2.

ing and medium-aware branch introduce additional computational overhead compared to the original 3DGS framework, our method still achieves real-time rendering. Taken together, these quantitative results demonstrate that our method achieves an optimal balance between reconstruction accuracy, perceptual quality, and training efficiency. This underscores its effectiveness in mitigating underwater image degradation to reconstruct high-fidelity visual content.

Scene restoration. As shown in Fig. 4, our method generates high-fidelity scene restorations from novel viewpoints. Thanks to our hybrid representation, the restored images exhibit significantly enhanced clarity and superior structural integrity compared to the original underwater scenes. Compared against other novel view synthesis methods, AquaSplatting demonstrates outstanding performance in global scene recovery. Our approach more accurately preserves crucial scene depth cues and fine-grained structural details. Specifically, our exposure- and depth-guided optimization facilitates a more natural and coherent reconstruction of the scene's background, particularly in distant regions.

Ablation Studies

Effectiveness of hybrid rendering mechanism. To validate the effectiveness of our neural underwater hybrid rendering mechanism, we conducted two key ablation studies. First, we test a simplified model that naively combines the outputs of the geometry-guided C_G and medium-aware C_M branches via direct per-pixel summation as $\hat{C} = C_G + C_M$. This simplistic blending, which fails to distinguish between foreground objects and the background water medium, is analogous to injecting unstructured noise from the medium branch into the final reconstruction. Second, we evaluate a variant that relies solely on the geometry-guided branch or the medium-aware branch, completely disabling the hybrid rendering. As shown in Tab. 2, these models show a noticeable degradation in quantitative metrics. These results underscore the critical role of our complete hybrid rendering mechanism in achieving accurate and robust reconstruction.

Effectiveness of EBP. Fig. 5 demonstrates the effectiveness of our EBP strategy. The result after applying EBP (Top

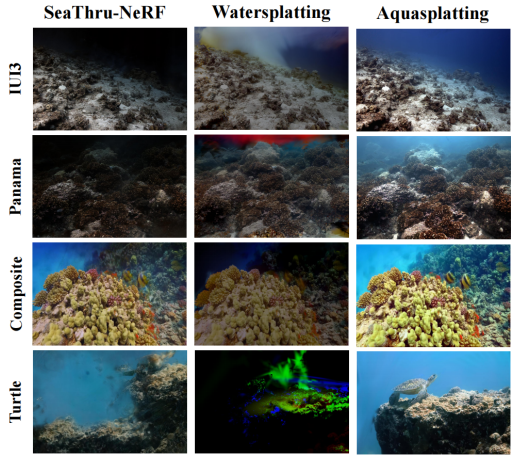


Figure 4: Comparison of image restoration. Our method achieves superior visual clarity and more consistent underwater scene restoration.

Geo-B	Med-B	Hybrid	PSNR ↑	SSIM ↑	LPIPS ↓
✓	×	×	27.98	0.889	0.161
×	✓	×	25.53	0.839	0.169
✓	✓	×	28.01	0.894	0.164
✓	✓	✓	**28.25**	**0.906**	**0.152**

Table 2: **Ablation study of the hybrid rendering mechanism.** We compare the quantitative performance of individual branches and their combinations on the U-IW Dataset.

right) preserves the full visual quality of the unpruned baseline (Top left), despite using remarkably fewer Gaussians. The total number of Gaussians can be reduced up to 3.2×. We also analyze the impact of different pruning threshold k settings on the final number of Gaussian primitives and the resulting rendering quality, as shown in Fig. 5(Bottom).

Effectiveness of loss functions. We ablate our loss components on the U-IW dataset, showing that each term progressively improves performance. As reported in the table of Fig. 6, adding the medium-aware density loss $\mathcal{L}_d$ to the principle reconstruction loss $\mathcal{L}_{\text{rec}}$ enhances foreground-background separation. The exposure regularization loss $\mathcal{L}_{\text{Exp}}$ further boosts structural similarity. Our full loss formulation achieves the top results across all metrics, demonstrating an effective balance of reconstruction quality and robustness. Fig. 6 visualizes the qualitative comparison results in different loss configurations.

Conclusion

We introduced *AquaSplatting*, a novel hybrid framework for high-fidelity, real-time reconstruction of underwater scenes. Our dual-branch architecture, which leverages 3DGS for objects and an MLP for the water, has proven to be highly effective. Key elements include a neural hybrid rendering to fuse branches and an EBP strategy for greater efficiency and artifact reduction. Extensive experiments indicate our approach achieves SOTA performance, delivering reconstructions with superior clarity and geometric accuracy.

Limitations & Future Work. Like most multi-view reconstruction techniques, our method requires accurate camera poses as input. Underwater conditions pose practical challenges, as factors such as light refraction and low illumination often impair the effectiveness of standard camera localization systems. We aim to expand our framework for larger-scale scenes with mixed media, such as water and fog, marking a crucial step toward solid 3D scene comprehension in complex environments.

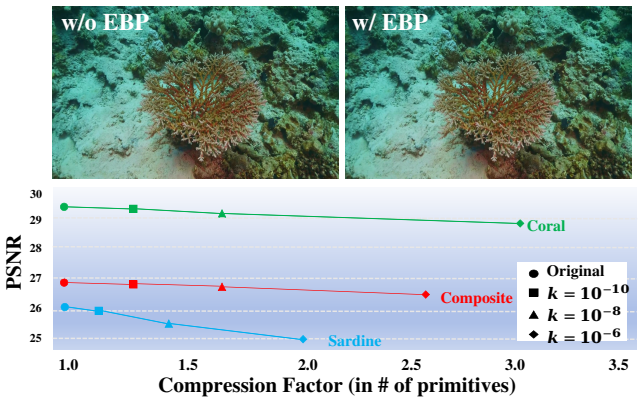
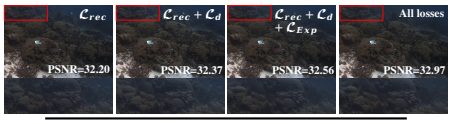


Figure 5: Our EBP strategy can reduce the total number of Gaussians by a factor of up to 3.2× with only a negligible impact on visual fidelity.



Configuration	PSNR↑	SSIM↑	LPIPS↓
$\mathcal{L}_{rec}$	29.82	0.901	0.162
$\mathcal{L}_{rec} + \mathcal{L}_d$	29.98	0.910	0.157
$\mathcal{L}_{rec} + \mathcal{L}_d + \mathcal{L}_{Exp}$	30.18	0.919	0.142
All losses	**30.32**	**0.933**	**0.136**

Figure 6: Qualitative and quantitative comparison of ablated loss configurations on the SeaThru-NeRF validation set.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (2021YFA1003003), the National Natural Science Foundation of China (62402083, T2225012, 62072071, 62262066, 62495080, 62495085), Fundamental Research Funds for the Central Universities (DUT25RC(3)031), and the Ministry of Education, Singapore, under its Academic Research Fund Grant (RT19/22).

References

Akkaynak, D.; and Treibitz, T. 2018. A revised underwater image formation model. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 6723–6732.

Akkaynak, D.; and Treibitz, T. 2019. Sea-Thru: A Method for Removing Water From Underwater Images. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1682–1691.

Boittiaux, C.; Marxer, R.; Dune, C.; Arnaubec, A.; Ferrera, M.; and Hugel, V. 2024. SUCRe: Leveraging scene structure for underwater color restoration. In *2024 International Conference on 3D Vision (3DV)*, 1488–1497. IEEE.

Chen, W.-T.; Yifan, W.; Kuo, S.-Y.; and Wetzstein, G. 2024. DehazeNeRF: Multi-image haze removal and 3D shape reconstruction using neural radiance fields. In *2024 International Conference on 3D Vision (3DV)*, 247–256. IEEE.

Cong, R.; Yang, W.; Zhang, W.; Li, C.; Guo, C.-L.; Huang, Q.; and Kwong, S. 2023. Pugan: Physical model-guided underwater image enhancement using gan with dual-discriminators. *IEEE Transactions on Image Processing*, 32: 4472–4485.

Guo, X.; Dong, Y.; Chen, X.; Chen, W.; Li, Z.; Zheng, F.; and Pun, C.-M. 2025a. Underwater image restoration via polymorphic large kernel cnns. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. IEEE.

Guo, Y.; Liao, H.; Ling, H.; and Huang, B. 2025b. NeuroPump: Simultaneous Geometric and Color Rectification for Underwater Images. arXiv:2412.15890.

Gupta, V.; Mitra, K.; et al. 2024. U2neRF: Unsupervised underwater image restoration and neural radiance fields. *arXiv preprint arXiv:2411.16172*.

Kajiya, J. T.; and Von Herzen, B. P. 1984. Ray tracing volume densities. In *Proceedings of the 11th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '84, 165–174. New York, NY, USA: Association for Computing Machinery. ISBN 0897911385.

Kerbl, B.; Kopanas, G.; Leimkühler, T.; and Drettakis, G. 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. arXiv:2308.04079.

Kim, S.; Lee, K.; and Lee, Y. 2024. Color-cued efficient densification method for 3d gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 775–783.

Levy, D.; Peleg, A.; Pearl, N.; Rosenbaum, D.; Akkaynak, D.; Korman, S.; and Treibitz, T. 2023. Seathru-nerf: Neural radiance fields in scattering media. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 56–65.

Li, C.; Anwar, S.; Hou, J.; Cong, R.; Guo, C.; and Ren, W. 2021. Underwater image enhancement via medium transmission-guided multi-color space embedding. *IEEE Transactions on Image Processing*, 30: 4985–5000.

Li, H.; Song, W.; Xu, T.; Elsig, A.; and Kulhanek, J. 2024. WaterSplatting: Fast Underwater 3D Scene Reconstruction Using Gaussian Splatting. arXiv:2408.08206.

Mualem, N.; Amoyal, R.; Freifeld, O.; and Akkaynak, D. 2024. Gaussian splashing: Direct volumetric rendering underwater. *arXiv preprint arXiv:2411.19588*.

Qiao, Y.; Shao, M.; Meng, L.; and Xu, K. 2025. RestorGS: Depth-aware Gaussian Splatting for Efficient 3D Scene Restoration. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 11177–11186.

Rahaman, N.; Baratin, A.; Arpit, D.; Draxler, F.; Lin, M.; Hamprecht, F.; Bengio, Y.; and Courville, A. 2019. On the spectral bias of neural networks. In *International conference on machine learning*, 5301–5310. PMLR.

Ramazzina, A.; Bijelic, M.; Walz, S.; Sanvito, A.; Scheuble, D.; and Heide, F. 2023. Scatternerf: Seeing through fog with physically-based inverse neural rendering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 17957–17968.

Rao, Y.; Liu, W.; Li, K.; Fan, H.; Wang, S.; and Dong, J. 2023. Deep color compensation for generalized underwater image enhancement. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(4): 2577–2590.

Ren, T.; Liu, S.; Zeng, A.; Lin, J.; Li, K.; Cao, H.; Chen, J.; Huang, X.; Chen, Y.; Yan, F.; et al. 2024. Grounded sam: Assembling open-world models for diverse visual tasks. *arXiv preprint arXiv:2401.14159*.

Schönberger, J. L.; and Frahm, J.-M. 2016. Structure-from-Motion Revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.

Sethuraman, A. V.; Ramanagopal, M. S.; and Skinner, K. A. 2023. WaterNeRF: Neural Radiance Fields for Underwater Scenes. arXiv:2209.13091.

Tang, Y.; Zhu, C.; Wan, R.; Xu, C.; and Shi, B. 2024. Neural underwater scene representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11780–11789.

Wang, H.; Anantrasirichai, N.; Zhang, F.; and Bull, D. 2025a. UW-GS: Distractor-aware 3d gaussian splatting for enhanced underwater scene reconstruction. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 3280–3289. IEEE.

Wang, J.; Chen, M.; Karaev, N.; Vedaldi, A.; Rupprecht, C.; and Novotny, D. 2025b. VGGT: Visual Geometry Grounded Transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Yang, D.; Leonard, J. J.; and Girdhar, Y. 2024. SeaSplat: Representing Underwater Scenes with 3D Gaussian Splatting and a Physically Grounded Image Formation Model. *arxiv*.

Yang, L.; Kang, B.; Huang, Z.; Zhao, Z.; Xu, X.; Feng, J.; and Zhao, H. 2024. Depth Anything V2. *arXiv:2406.09414*.

Yuan, J.; Cai, Z.; and Cao, W. 2021. TEBCF: Real-world underwater image texture enhancement model based on blurriness and color fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–15.

Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *CVPR*.

Zhou, J.; Liang, T.; Zhang, D.; Liu, S.; Wang, J.; and Wu, E. Q. 2025. WaterHE-NeRF: Water-ray matching neural radiance fields for underwater scene reconstruction. *Information Fusion*, 115: 102770.