Spatiotemporal Degradation-Aware 3D Gaussian Splatting for Realistic Underwater Scene Reconstruction

Anonymous Author(s)

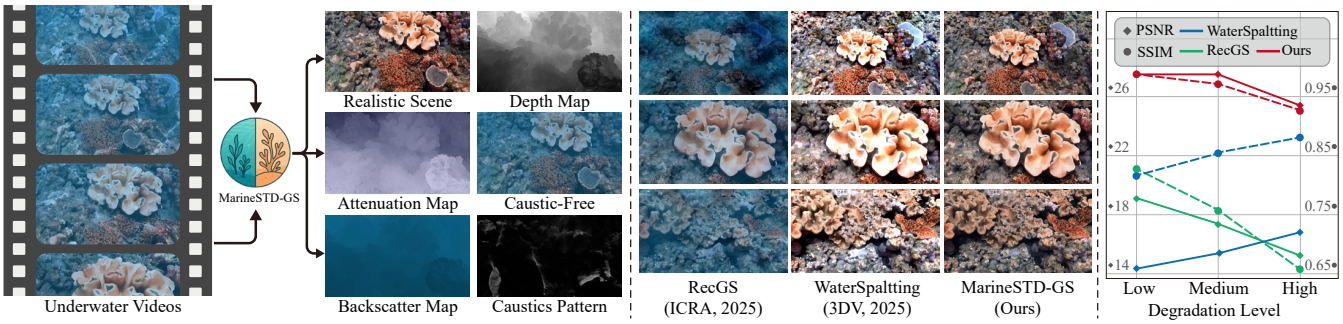

Figure 1: (a) Given underwater video sequences with complex spatiotemporal degradations, MarineSTD-GS performs holistic scene understanding by disentangling realistic scene representations—recovering true color and geometry (depth), and estimating spatial degradation parameters (e.g., attenuation and backscatter) as well as transient caustic patterns. (b) Compared to state-of-the-art underwater 3D reconstruction methods, our approach produces more faithful and consistent scene appearances, effectively reducing color distortion and caustic-induced flickering. (c) Quantitative results show that MarineSTD-GS consistently outperforms all baselines across varying degradation levels, demonstrating superior robustness. For more visual Comparison and Results, Please refer to our anonymous project page via https://marinestd-gs.github.io/.

Abstract

Reconstructing realistic underwater scenes from underwater video remains a meaningful yet challenging task in the multimedia domain. The inherent spatiotemporal degradations in underwater imaging, including caustics, flickering, attenuation, and backscattering, often lead to inaccurate geometry and appearance in existing 3D reconstruction methods. While a few recent works have explored underwater degradation-aware reconstruction, they often address either temporal or spatial degradation alone, falling short in more real-world underwater scenarios where both types of degradation occur. We propose MartineSTD-GS, a novel 3D Gaussian Splatting-based framework that explicitly models both temporal and spatial degradations for realistic underwater scene reconstruction. Specifically, we introduce two paired Gaussian primitives: Intrinsic Gaussians represent the true scene, while Degraded Gaussians render the observations. The degraded colors are physically derived from the intrinsic ones via a Spatiotemporal Degradation Modeling (SDM) module, enabling self-supervised disentanglement of realistic appearance from degraded images. To ensure stable training and accurate geometry, we further propose a Multi-Stage Optimization strategy and a Depth-Guided Geometry Loss. We also construct a simulated benchmark with diverse degradations and ground-truth appearances for comprehensive evaluation. Experiments on both simulated and real-world datasets show that MarineSTD-GS robustly handles spatiotemporal degradations and outperforms existing methods in novel view synthesis with realistic, water-free scene appearances.

CCS Concepts

• **Computing methodologies** → **Reconstruction**; **Rendering**; **Scene understanding**; **Image processing**.

Keywords

Underwater 3D Reconstruction, Underwater Image Restoration, 3D Gaussian Splatting, Novel View Synthesis

ACM Reference Format:

Anonymous Author(s). 2025. Spatiotemporal Degradation-Aware 3D Gaussian Splatting for Realistic Underwater Scene Reconstruction. In *Proceedings of ACM Multimedia 2025 (MM '25).* ACM, Dublin, Ireland, 9 pages. https://doi.org/XXXXXXX.XXXXXXX

MM '25, Dublin, Ireland

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Reconstructing realistic scenes from underwater video is critical for a wide range of marine multimedia applications, including underwater archaeology [23, 47], ecological monitoring [37, 41], and immersive virtual reality [16, 24]. However, underwater imaging is inherently affected by a range of degradation factors, which can be broadly divided into spatial and temporal categories. Spatial degradations such as distance-dependent attenuation and backscattering often cause color distortions and haze effects [7]. On the

other hand, temporal degradations including caustics and flickering lead to inconsistent lighting and severe local brightness fluctuations [1, 11, 29]. These spatiotemporal degradations violate the multi-view consistency assumption required by most 3D reconstruction methods such as Neural Radiance Field (NeRF)[25] and 3D Gaussian Splatting (3DGS) [17], resulting in inaccurate geometry and appearance, and ultimately hindering downstream applications [10].

Recent efforts [18, 20, 38, 45] in underwater-specific 3D reconstruction have attempted to address certain types of degradation. SeaThruNeRF [18], WaterSplatting [20], and SeaSplat [38] introduce attenuation and scattering models to decouple realistic scene appearance from spatially degraded underwater observations. RecGS [45], on the other hand, incorporates low-pass filtering and recurrent updates into the 3DGS pipeline to mitigate flickering artifacts. However, since these methods typically address only one aspect of degradation, they struggle to recover realistic geometry and appearance when both spatial and temporal degradations occur, as commonly seen in real underwater environments.

To address this challenge, we propose MarineSTD-GS, a novel 3D Gaussian-based framework that explicitly models both spatial and temporal degradation in underwater imaging and disentangles realistic scene representations from degraded video through self-supervised learning. It employs two paired types of Gaussian primitives: Intrinsic Gaussians represent the underlying scene, while Degraded Gaussians are used to render degraded observations. These primitives share geometric attributes, and the color of each Degraded Gaussian is derived from its intrinsic counterpart via a physically grounded Spatiotemporal Degradation Modeling (SDM) module. To improve geometry learning under weak visual cues, we introduce a Depth-Guided Geometry Loss that leverages monocular depth priors. Furthermore, a Multi-Stage Optimization strategy stabilizes training under coupled degradations and facilitates appearance refinement. To enable comprehensive evaluation, we also construct a simulated underwater dataset covering diverse scenes, caustic patterns, and degradation levels.

Our contributions can be summarized as follows:

- We present MarineSTD-GS, the first 3DGS-based framework that explicitly models both spatial and temporal degradations in underwater imaging, enabling the disentangled reconstruction of realistic scene content and water-related degradation parameters.
- We develop a Multi-Stage Optimization strategy for stable training and a Depth-Guided Geometry Loss to enhance both global structure and local geometric fidelity.
- We construct and release a comprehensive simulated underwater dataset with diverse scenes, caustic patterns, and degradation levels to support standardized evaluation.
- Extensive experiments on both simulated and real-world data show that MarineSTD-GS achieves state-of-the-art performance in color accuracy, flickering suppression, and visibility restoration.

2 Related Work

2.1 3D Gaussian Splatting

3D Gaussian Splatting [17] represents scenes with a set of learnable 3D Gaussian primitives $\{\boldsymbol{\mu}, \Sigma, o, \boldsymbol{c}\}$, where $\boldsymbol{c}$ is modeled via view-dependent spherical harmonics (SH). It employs a differentiable rasterization pipeline to render images and depths, supervised by photometric losses such as L1 and D-SSIM. Owing to its high fidelity and efficiency, 3DGS has been adapted to various settings, including sparse views [21, 35, 48], large-scale scenes [22], and wild image collections [36, 42]. Additional improvements involve geometric priors like monocular depth [21, 35], or regularization strategies targeting edge [13], frequency [43], and rank constraints [15]. Over-splatting control is addressed in [9, 46]. GS-W [42] and Splatfacto-W [36] are particularly relevant for modeling dynamic appearances, a situation analogous to underwater scenes with dynamic appearance. Recent underwater reconstruction methods [20, 38, 45] incorporate degradation modeling via imaging equations or volumetric rendering. In contrast, MarineSTD-GS employs a dual-Gaussian design and SDM module to explicitly model both spatial and temporal degradations within a unified 3DGS framework.

2.2 Underwater Restoration and Reconstruction

Underwater imaging suffers from spatial degradations (e.g., attenuation, backscattering) and temporal artifacts (e.g., caustics, flickering). Traditional restoration methods handle color degradation via image priors [3, 6, 8] or learning-based enhancement [5, 19, 34], while recent works [7] leverage 3D structure for parameter estimation. Temporal degradations are often modeled as additive [12] or multiplicative [14, 32] components of illumination; RecGS [45] mitigates flickering via recurrent modeling but ignores spatial effects. In 3D reconstruction, NeRF-based methods [18, 27, 33, 44] simulate underwater light propagation volumetrically but suffer from inefficiency and blurred geometry. 3DGS-based approaches [20, 38] improve efficiency by incorporating physical models but do not account for transient lighting. Unlike prior work that models only one type of degradation, our MarineSTD-GS jointly addresses both spatial and temporal effects, enabling more robust and faithful scene reconstruction.

3 Method

3.1 Overview

The pipeline of MarineSTD-GS is illustrated in Fig. 2. To enable self-supervised learning of the 3D representation of the realistic scene from degraded underwater images, we design two types of 3D Gaussian primitives. The Intrinsic Gaussians represent the underlying scene content, while the Degraded Gaussians are used to render the degraded underwater images corresponding to specific training views at given times. Each Intrinsic Gaussian is paired with a corresponding Degraded Gaussian, and the two share the same mean position, covariance, and opacity. The color attributes of Intrinsic Gaussians capture the realistic appearance of the scene, while those of the corresponding Degraded Gaussians are derived from the Intrinsic ones through the proposed Spatiotemporal Degradation Modeling (SDM), which physically models the spatiotemporal color degradation process.

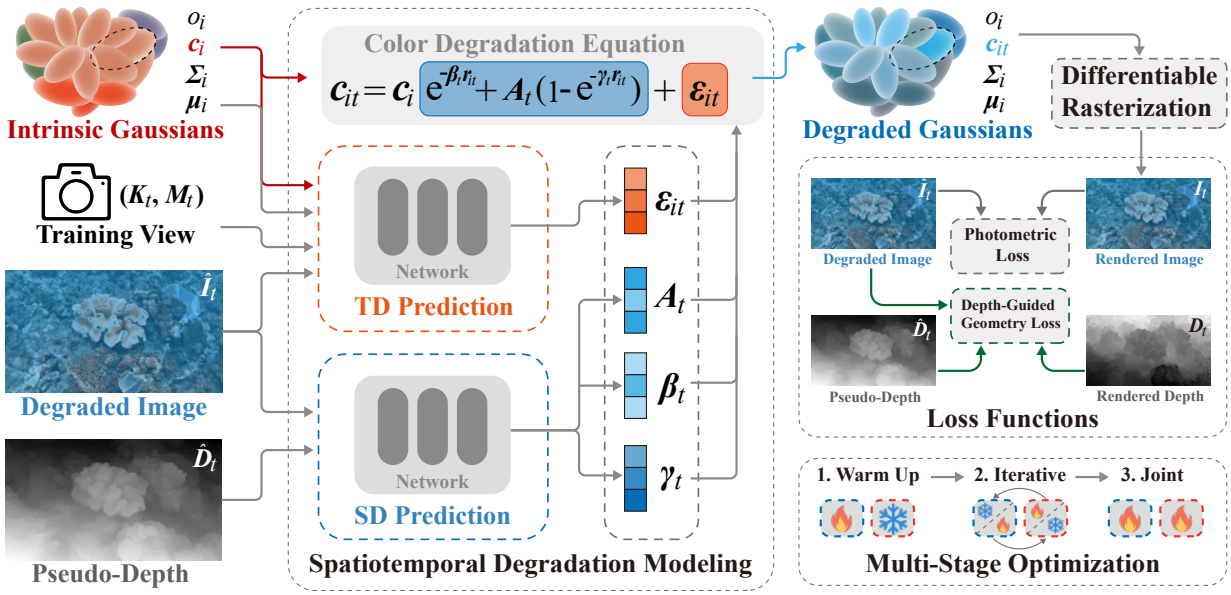

Figure 2: Pipeline of MarineSTD-GS. Given a training view at a specific time, the Spatiotemporal Degradation Modeling (SDM) module predicts the color degradation parameters of each Gaussian under the current spatiotemporal condition, and derives the degraded colors of Degraded Gaussians from their corresponding Intrinsic counterparts for rendering the underwater image and depth map. During optimization, in addition to photometric losses for reconstructing the degraded input, a Depth-Guided Geometry loss is employed to enhance geometric consistency. A Multi-Stage Optimization strategy is also adopted to ensure the stable training of the SDM components.

During training, in addition to a photometric reconstruction loss that encourages the rendered image to match the degraded input, we introduce a Depth-Guided Geometry Loss, which leverages geometric priors from state-of-the-art monocular depth estimators, such as Depth-Anything-V2 [39] and texture information from the input image to improve geometry optimization. Furthermore, a Multi-Stage Optimization strategy is proposed to stabilize the training of SDM components, enabling Intrinsic Gaussians to capture more accurate scene representations.

We next present the details of the SDM module in Section 3.2, the loss functions in Section 3.3, and the Multi-Stage Optimization strategy in Section 3.4.

3.2 Spatiotemporal Degradation Modeling

To more accurately model underwater degradation, the SDM module, as illustrated in Fig. 2, consists of a physically grounded Color Degradation Equation, which describes how scene colors are degraded under varying spatiotemporal conditions, along with two dedicated branches for predicting different types of degradation parameters.

3.2.1 Color Degradation Equation. Inspired by the revised underwater imaging model proposed by Akkaynak *et al.* [2], we model the relationship between the color of the i-th Intrinsic Gaussian and its corresponding Degraded Gaussian at time t as follows:

$$\boldsymbol{c}_{it} = \boldsymbol{c}_i \underbrace{e^{-\boldsymbol{\beta}_t \cdot r_{it}} + \boldsymbol{A}_t (1 - e^{-\boldsymbol{\gamma}_t \cdot r_{it}})}_{\text{SD term}} + \underbrace{\epsilon_{it}}_{\text{TD term}}. \quad (1)$$

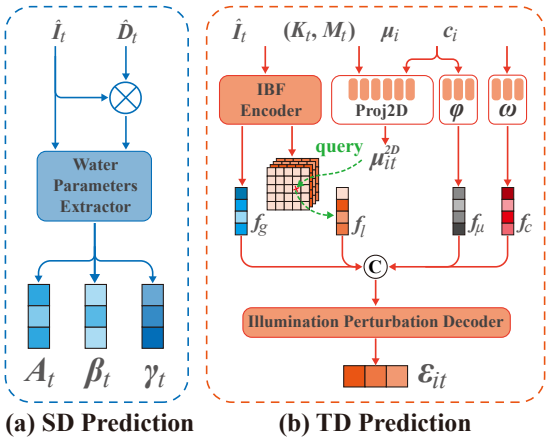

Figure 3: Pipelines of the SD-Prediction and TD-Prediction branches.

Equation 1 consists of two components: the SD term and the TD term. The SD term is inspired by the underwater image formation model proposed by Akkaynak *et al.*, and models the effects of spatial degradation (SD) such as attenuation and backscattering. Here, r_{it} denotes the distance between the t-th viewpoint and the i-th Gaussian. The water parameters $\boldsymbol{\beta}_t$, $\boldsymbol{\gamma}_t$, and $\boldsymbol{A}_t$ represent the attenuation coefficient, backscatter coefficient, and underwater ambient light, respectively, all of which are shared by all Gaussians at time t

and are predicted by the SD Prediction branch. In contrast, the TD term captures the influence of temporal degradation (TD) caused by dynamic lighting phenomena such as caustics and flickering. Specifically, ϵ_{it} denotes the transient illumination perturbation for the i-th Gaussian at time t, which is Gaussian-specific and predicted by the TD Prediction branch.

3.2.2 Spatial Degradation Prediction. To effectively predict the water parameters, the SD Prediction branch directly extracts them from the degraded image $\hat{\boldsymbol{I}}_t$. To increase sensitivity to distant regions where water effects are more pronounced, we first apply a depth-aware enhancement to $\hat{\boldsymbol{I}}_t$ using the pseudo-depth map $\hat{\boldsymbol{D}}_t$. The resulting enhanced image, denoted as $\hat{\boldsymbol{I}}'_t = \hat{\boldsymbol{D}}_t \otimes \hat{\boldsymbol{I}}_t$, suppresses foreground areas with minimal degradation, thereby increasing the relative emphasis on distant regions where underwater effects are stronger. We then concatenate $\hat{\boldsymbol{I}}'_t$ and $\hat{\boldsymbol{I}}_t$ along the channel dimension and feed the result into a Water Parameters Extractor (WPE) to obtain the final water parameters: $\boldsymbol{A}_t$, $\boldsymbol{\beta}_t$, and $\boldsymbol{\gamma}_t$. The detailed structure of the WPE can be found in the supplementary material.

3.2.3 Temporal Degradation Prediction. To better predict the local instantaneous brightness variations of the i-th Gaussian at time t, we design an Instantaneous Brightness Feature (IBF) Encoder. This module extracts a low-resolution feature map $\boldsymbol{F}_l$ and a global feature vector $\boldsymbol{f}_g$ from the degraded image $\hat{\boldsymbol{I}}_t$. The feature map $\boldsymbol{F}_l$ retains spatially uneven brightness patterns caused by flickering or caustics, while the global vector $\boldsymbol{f}_g$ provides an overall brightness representation. Next, we compute the 2D projection coordinate μ_{it}^{2D} of the i-th Gaussian under the t-th view using the camera intrinsics $\boldsymbol{K}_t$ and extrinsics $\boldsymbol{M}_t$, via a function $\mu_{it}^{2D} = \text{Proj2D}(\boldsymbol{K}_t, \boldsymbol{M}_t, \mu_i)$. Based on this coordinate, the corresponding local brightness feature vector $\boldsymbol{f}_l$ is queried from $\boldsymbol{F}_l$ using bilinear interpolation. Meanwhile, the intrinsic color $\boldsymbol{c}_i$ and 3D position $\boldsymbol{\mu}_i$ of the i-th Gaussian are encoded into feature vectors via two learnable encoders, $\phi(\cdot)$ and $\omega(\cdot)$, respectively. Finally, all feature vectors are concatenated and fed into an Illumination Perturbation Decoder to produce the transient illumination perturbation ϵ_{it} for the i-th Gaussian at time t, the structure of which can be found in the supplementary material.

3.3 Loss Function

3.3.1 Photometric Loss. In our self-supervised setup, we adopt the standard 3DGS photometric loss to supervise both the Intrinsic Gaussians and the SDM module. Specifically, we compute the L1 and D-SSIM losses between the rendered image and the degraded input, combined as:

$$L_{\text{photo}} = \lambda_1 \cdot \|\hat{\boldsymbol{I}}_t - \boldsymbol{I}_t\|_1 + \lambda_2 \cdot \text{D-SSIM}(\hat{\boldsymbol{I}}_t, \boldsymbol{I}_t), \tag{2}$$

where $\hat{\boldsymbol{I}}_t$ is the rendered image at time t, $\boldsymbol{I}_t$ is the corresponding degraded input image, and λ_1, λ_2 are weighting coefficients.

3.3.2 Depth-Guided Geometry Loss. Due to spatiotemporal degradation, the multi-view consistency assumed by 3DGS-based methods often breaks down, making photometric loss alone insufficient for accurate geometry learning. We observe that recent monocular depth estimators [39] remain robust even under severe texture degradations caused by caustics. Motivated by this, we propose a Depth-Guided Geometry Loss (L_{dgg}) that leverages pseudo-depth priors to enhance geometric reconstruction. It comprises a coarse depth supervision term and an adaptive edge-aware smoothness term.

Coarse Depth Supervision Term. Since the estimated pseudo-depth maps typically reflect normalized relative disparities rather than absolute depth values, direct supervision of the rendered depth maps using L1 or L2 loss is not suitable. Instead, we employ the Pearson correlation coefficient [35], which is invariant to translation and scale, to measure the similarity between the pseudo-depth map $\boldsymbol{D}_t$ and the rendered depth map $\hat{\boldsymbol{D}}_t$. The Coarse Depth Supervision Term is defined as:

$$L_{\text{cds}} = 1 - \text{Pearson}(\boldsymbol{D}_t, \hat{\boldsymbol{D}}_t). \tag{3}$$

Adaptive Edge-aware Depth Smoothness Term. While L_{cds} provides coarse global supervision, we further introduce an adaptive edge-aware smoothness term to improve local geometric consistency. It computes spatial gradients of the rendered depth map $\boldsymbol{D}_t$ and applies adaptive weights derived from the gradients of the pseudo-depth map $\hat{\boldsymbol{D}}_t$ and input image $\boldsymbol{I}_t$, modulated by $\hat{\boldsymbol{D}}_t$.

$$\begin{aligned} L_{\text{ads}} = \frac{1}{N} \sum_{x,y} \big(& |\nabla_x \boldsymbol{D}_t(x,y) \cdot \boldsymbol{w}_x(x,y)| \\ & + \left|\nabla_y \boldsymbol{D}_t(x,y) \cdot \boldsymbol{w}_y(x,y)\right| \big), \end{aligned} \tag{4}$$

where the adaptive weights $\boldsymbol{w}_x$ and $\boldsymbol{w}_y$ are defined as:

$$\boldsymbol{w}_x = (1 - \hat{\boldsymbol{D}}_t(x,y)) \cdot e^{-|\nabla_x \hat{\boldsymbol{D}}_t(x,y)|} + \hat{\boldsymbol{D}}_t(x,y) \cdot e^{-|\nabla_x \hat{\boldsymbol{I}}_t(x,y)|}, \tag{5}$$

$$\boldsymbol{w}_y = (1 - \hat{\boldsymbol{D}}_t(x,y)) \cdot e^{-|\nabla_y \hat{\boldsymbol{D}}_t(x,y)|} + \hat{\boldsymbol{D}}_t(x,y) \cdot e^{-|\nabla_y \hat{\boldsymbol{I}}_t(x,y)|}, \tag{6}$$

with $\hat{\boldsymbol{D}}_t(x,y) \in [0,1]$ assigning higher values to distant regions. This design encourages more reliance on RGB-based edge guidance in areas where the depth signal is less reliable.

The final Depth-Guided Geometry Loss is a weighted combination of the two terms described above:

$$L_{\text{dgg}} = \lambda_{\text{cds}} L_{\text{cds}} + \lambda_{\text{ads}} L_{\text{ads}}, \tag{7}$$

where λ_{cds} and λ_{ads} are weighting coefficients.

3.3.3 Total Loss. The overall training objective combines the photometric loss and the Depth-Guided Geometry Loss with an additional regularization term applied to the transient illumination perturbations ϵ_{it}. Specifically, we impose an ℓ_2 regularization on all predicted ϵ_{it} values to prevent the model from overfitting to color reconstruction errors by relying excessively on the transient term. This regularization helps avoid trivial solutions and promotes stable disentanglement between intrinsic scene color and transient lighting effects. The total loss is defined as:

$$L_{\text{total}} = L_{\text{photo}} + L_{\text{dgg}} + \lambda_\epsilon \cdot L_{\epsilon\text{-reg}}, \tag{8}$$

where the regularization term is defined as $L_{\epsilon\text{-reg}} = \sum_{i,t} ||\epsilon_{it}||_2^2$, and λ_ϵ controls its strength.

3.4 Multi-Stage Optimization

In practice, we observe that the transient illumination perturbations predicted by the TD Prediction branch may interfere with the backscatter components estimated by the SD Prediction branch, resulting in optimization conflicts within the SDM module. To ensure stable training of the two core branches, we adopt a multi-stage optimization strategy that divides the training process into three phases:

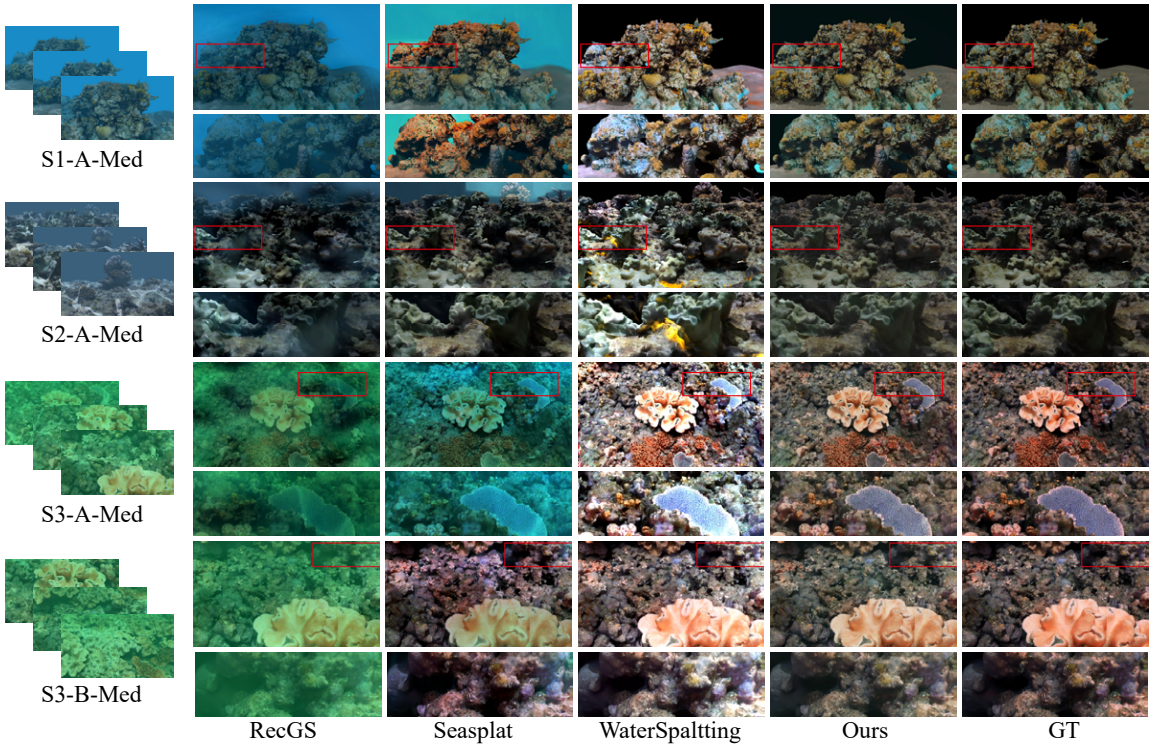

Figure 4: Qualitative comparisons of novel view synthesis on simulated scenes. Our method removes global color distortions from attenuation and backscattering, and suppresses local overexposure under varying caustic patterns, resulting in appearances closest to the ground truth.

Table 1: Quantitative evaluation of novel view synthesis on simulated scenes, with results reported separately for each scene category. The best results are highlighted in bold, and the second-best are underlined. Efficiency metrics are also reported.

Methods	Paper/Year	S1 Scene			S2 Scene			S3 Scene			FPS	Avg. Time
		PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS		
3DGS	SIGGRAPH/2023	17.656	0.792	0.167	15.218	0.695	0.163	13.804	0.692	0.578	125.43	0.40h
gsplat	Arxiv/2024	17.871	0.782	0.179	15.257	0.685	0.183	13.741	0.673	0.599	107.24	0.16h
GS-W	ECCV/2024	18.010	0.778	0.198	17.001	0.687	0.220	14.340	0.666	0.570	67.68	1.18h
Splatfact-W	Arxiv/2024	16.933	0.772	0.193	14.176	0.663	0.195	14.132	0.695	0.581	91.42	0.16h
SeaThru-NeRF	CVPR/2023	23.995	0.885	0.105	15.348	0.529	0.420	15.934	0.711	0.358	0.23	13.5h
Seasplat	Arxiv/2024	24.064	0.908	0.097	18.292	0.813	0.124	17.379	0.830	0.294	37.71	0.52h
RecGS	ICRA/2025	18.576	0.777	0.159	18.551	0.728	0.164	14.230	0.685	0.549	392.16	0.33h
WaterSpaltting	3DV/2025	18.739	0.882	0.075	14.279	0.767	0.137	14.275	0.809	0.152	84.53	0.12h
Ours	——/2025	**27.719**	**0.934**	**0.059**	**25.853**	**0.863**	**0.085**	**26.355**	**0.918**	**0.097**	104.91	0.38h

Stage I: Warm-up. In this phase, the TD Prediction branch participates in degraded color computation but remains frozen, i.e., no gradients are propagated through its parameters. This allows the SD Prediction branch to independently learn reasonable water parameters without interference.

Stage II: Iterative training. We alternately stop the gradient flow of the TD and SD branches in an interleaved manner. At each iteration, only one branch is updated while the other provides a fixed degradation signal to support supervision. This mitigates mutual interference and facilitates decoupled learning of the two degradation sources.

Stage III: Joint training. In the final stage, both the SD and TD Prediction branches are jointly optimized. This enables the model to refine degradation parameter learning and improve disentanglement, leading to more accurate reconstruction and realistic scene representation.

After optimization, the Intrinsic Gaussians are able to effectively represent the underlying realistic scene. They can be directly used

for novel view synthesis as if observed without underwater degradations such as attenuation, backscatter, caustics, or flickering.

4 Experiments

4.1 Experimental Setup

Simulated Dataset. To enable comprehensive evaluation, we construct three major categories of synthetic underwater scenes using Blender, each incorporating both spatial and temporal degradation factors, since obtaining real underwater scenes without degradations is nearly impossible. The three scene categories are designed as follows: *S1* features detailed textures and varied colors; *S2* simulates a large-scale open environment with significant contrast loss in distant regions; *S3* contains scenes with dominant greenish color distortions. For each category, we generate five distinct synthetic settings: three with different caustic patterns under medium spatial degradation, and two additional variants using a fixed caustic pattern (Pattern A) but with varying levels of spatial degradation (low and high). This design enables systematic evaluation of model robustness under diverse combinations of degradation intensity and structure. Each scene contains 120 images at a resolution of 540×960. Further construction details are provided in the supplementary material.

Real-world Dataset. We evaluate our method on a variety of real underwater scenes including four representative scenes from each of the BVICoral [4] and Flseas_VI [28] datasets, which exhibit strong caustic and flickering effects along with color degradation. Additionally, we use four scenes from the SeaThru-NeRF dataset [18] and two scenes (D3 and D5) from the SeaThru dataset [3].

Comparison Methods. We compare our method against eight state-of-the-art baselines, including four general 3D scene reconstruction methods (3DGS [17], gsplat [40], GS-W [42], and Splatfacto-W [36]) and four recent underwater-specific approaches (SeaThru-NeRF [18], WaterSplatting [20], SeaSplat [38], and RecGS [45]).

Evaluation Metrics. We primarily evaluate novel view synthesis performance on realistic scenes that have been restored from underwater degradation. On the simulated dataset, we compute PSNR, SSIM, and LPIPS, using the rendered images and corresponding ground truth. For real-world scenes containing color charts, we further assess color fidelity using CIEDE2000 (ΔE_{00}) [31] and average angular error $\bar{\psi}$ (in degrees) [7]. We also report the average training time (Avg. Time) and rendering speed (FPS) at a resolution of 540×960 to evaluate the overall efficiency of each method.

Implementation Details. Our method is implemented based on gsplat [40]. We initialize the 3D Gaussians using a sparse point cloud reconstructed by COLMAP [30], with the intrinsic color attributes initialized from the corresponding point colors. To model the Lambertian-like appearance of underwater scenes [26], we set the spherical harmonics (SH) degree of intrinsic Gaussians to zero. We adopt a multi-stage training strategy as described in Sec. 3.4, with each stage lasting for 10,000 steps, resulting in a total of 30,000 training iterations. All experiments are conducted on a single NVIDIA RTX 3090 GPU. Additional implementation details are provided in the supplementary material.

Table 2: Quantitative color correction results on real-world scenes.

Methods	Curasao		D3		D5		Avg.	
	ΔE_{00}	$\bar{\psi}$	ΔE_{00}	$\bar{\psi}$	ΔE_{00}	$\bar{\psi}$	ΔE_{00}	$\bar{\psi}$
3DGS	25.63	26.35	20.18	22.65	32.38	26.62	26.06	25.21
gsplat	23.24	26.20	20.34	22.45	30.93	26.52	24.84	25.06
GS-W	25.76	26.45	27.82	26.66	35.17	27.64	29.58	26.92
Splatfact-W	28.31	26.92	22.12	23.29	31.66	27.45	27.36	25.89
SeaThru-NeRF	20.97	**23.29**	21.13	24.92	34.44	26.92	25.51	25.04
Seasplat	24.33	23.59	19.86	22.80	37.04	27.86	27.08	24.75
RecGS	39.53	29.00	20.15	22.55	24.09	23.66	27.92	25.07
WaterSpaltting	37.16	26.53	24.30	23.05	**20.45**	21.83	27.30	23.80
Ours	**19.98**	23.40	**19.64**	**22.27**	22.23	**20.95**	**20.62**	**22.21**

Table 3: Quantitative evaluation of robustness under varying spatial degradation levels.

Methods	Low		Medium		High	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
3DGS	17.417	0.808	15.623	0.729	13.988	0.638
gsplat	17.470	0.793	15.811	0.721	14.073	0.629
GS-W	17.862	0.789	16.131	0.716	14.383	0.634
Splatfact-W	16.065	0.771	15.214	0.715	13.970	0.637
SeaThru-NeRF	18.479	0.705	18.152	0.687	18.289	0.700
Seasplat	21.535	0.895	19.430	0.839	17.242	0.780
RecGS	19.119	0.789	17.391	0.731	15.245	0.645
WaterSpaltting	14.367	0.781	15.379	0.812	16.800	0.835
Ours	**27.520**	**0.927**	**27.502**	**0.913**	**25.396**	**0.875**

Table 4: Quantitative evaluation of ablation experiments on simulated scenes under Caustic Pattern A.

Methods	PSNR	SSIM	LPIPS
w/o SD Prediction	16.804	0.730	0.310
w/o TD Prediction	24.440	0.874	0.111
w/o WPE	20.711	0.809	0.227
w/o L_{dgg}	25.791	0.893	0.086
w/o $L_{\epsilon-\text{reg}}$	19.388	0.774	0.194
w/o MS-Opt.	25.705	0.887	0.094
Full Model	**26.806**	**0.905**	**0.080**

4.2 Experimental Results

Quantitative Results on Simulated Scenes. Table 1 reports the quantitative results of novel view synthesis on the simulated dataset. Our method consistently achieves the best performance across all metrics and scene categories. Notably, on the S1 scenes, it outperforms the second-best method (Seasplat) by 3.655 dB in PSNR. For S3, our advantage increases to 8.976 dB, indicating strong robustness to severe color distortions.

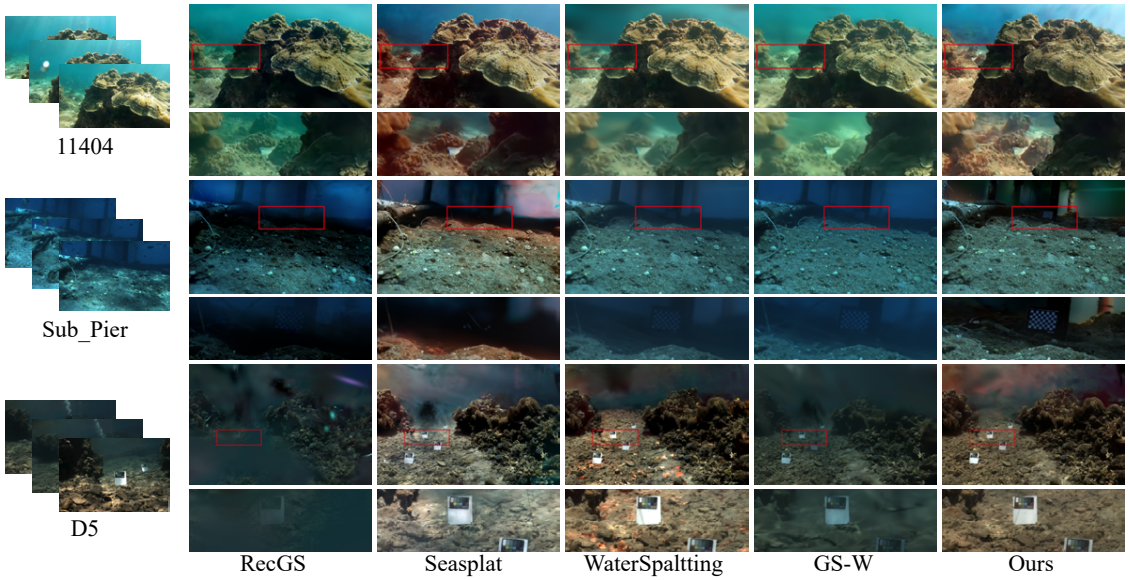

Figure 5: Qualitative comparisons of novel view synthesis on real-world scenes. Our method corrects attenuation-induced color casts, removes backscattering haze, and suppresses caustic-induced illumination artifacts for consistent scene appearance.

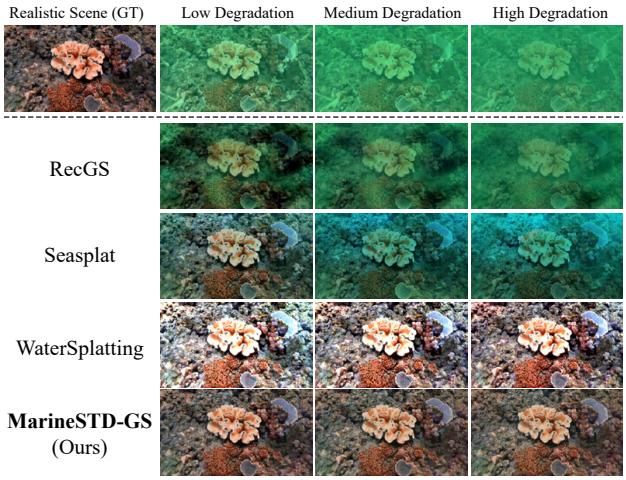

Figure 6: Qualitative comparisons under different spatial degradation levels on simulated scenes. Our method consistently preserves scene appearance and color fidelity, even under severe degradation, closely matching the ground truth.

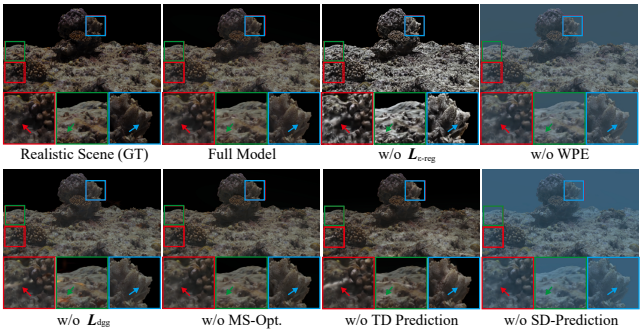

Figure 7: Visual comparisons of novel view synthesis results under different ablation settings.

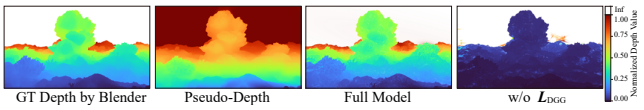

Figure 8: Impact of depth-guided geometry supervision using pseudo-depth. Although pseudo-depth deviates from the ground truth, it helps improve geometric consistency.

Quantitative Results on Real-world Scenes. Table 2 presents quantitative color correction results on real-world scenes. Our method leads on both evaluation metrics, reducing ΔE_{00} by 4.22 compared to gsplat, and lowering the angular error $\bar{\psi}$ by 1.59 compared to WaterSplatting, demonstrating its effectiveness in real-world color fidelity recovery.

Qualitative Comparisons. Fig. 4 and Fig. 5 show qualitative comparisons. RecGS, which ignores spatial degradations, fails to remove the color cast and haze caused by attenuation and backscattering. SeaSplat introduces strong color artifacts, while WaterSplatting struggles to suppress local overexposure under caustics. GS-W demonstrates partial effectiveness in mitigating temporal degradations such as flickering, but is unable to handle spatial effects like backscattering. In contrast, our method corrects color distortions, enhances visibility, and suppresses caustic-induced illumination artifacts across both simulated and real-world scenes. For example, in

the Sub_Pier and D5 scenes, our method restores distant structures and produces consistent color and lighting.

Robustness Under Different Degradation Levels. Table 3 compares the robustness of different methods under varying levels of spatial degradation. Our method consistently achieves the best performance, maintaining high fidelity even under severe degradation. Notably, it retains a minimum lead of 5.985 dB in PSNR over the second-best method across all degradation levels. Fig. 6 visually confirms this robustness, showing that our reconstructed realistic scenes preserve color fidelity and appearance closest to the ground truth across different degradation intensities.

4.3 Ablation Study

All ablation experiments are conducted on all three simulated scenes under Caustic Pattern A, covering all spatial degradation levels.

Effectiveness of SDM Components. We first evaluate the contributions of key components in the SDM module, including the TD Prediction branch, the SD Prediction branch, and the Water Parameter Extractor (WPE) within the SD branch. As shown in Table 4, removing any of these components leads to a noticeable drop in performance, with at least 2.366 dB reduction in PSNR. Fig. 7 further demonstrates that models without these components either fail to suppress flickering-induced local highlights or are unable to properly correct water degradation effects.

Impact of Optimization Strategies. We then assess the impact of our proposed optimization strategies. As shown in Table 4, each component contributes positively to the overall performance. As shown in Fig. 7, without the regularization term $L_{\epsilon\text{-reg}}$, the model tends to rely excessively on the transient term, causing the Intrinsic Gaussians to degenerate into overly monotonic colors. Without the multi-stage optimization strategy (MS-Opt.), the learned scene appearance shows reduced detail and stability. Additionally, Fig. 8 highlights the role of L_{dgg}: although the pseudo-depth map may deviate from the ground truth, it still guides the reconstructed geometry to better match the true structure.

These results collectively validate the necessity of each component in our framework. The SDM branches, optimization strategies, and geometric guidance all contribute to improved robustness, color fidelity, and structural accuracy, confirming the effectiveness of our design for modeling complex underwater degradations.

4.4 More Applications

Our method explicitly models spatiotemporal degradation in underwater video sequences, enabling the disentanglement of two components from a single sequence: a realistic 3D scene representation and a set of per-frame water parameters. As a result, in addition to enabling novel view synthesis (NVS) of clean, realistic scenes—our primary objective—we can also support NVS with realistic water effects such as attenuation and backscattering, but without transient degradations like caustics.

Figure 9 illustrates two extended applications enabled by this disentangled representation:

(1) Intrinsic Water Effects Rendering. Given an underwater video sequence from Scene 1, we extract both the intrinsic 3D scene representation and a collection of water parameters for each frame, denoted as $\{\boldsymbol{\beta}_t^1, \boldsymbol{\gamma}_t^1, \boldsymbol{A}_t^1\}_{t=1}^{N^1}$. We then compute the average water parameters (e.g., $\bar{\boldsymbol{\beta}}^1$, $\bar{\boldsymbol{\gamma}}^1$, $\bar{\boldsymbol{A}}^1$) and reuse them to render novel views of Scene 1 with realistic underwater effects. This provides plausible scene rendering with consistent, interpretable water degradation, free from caustic flickering.

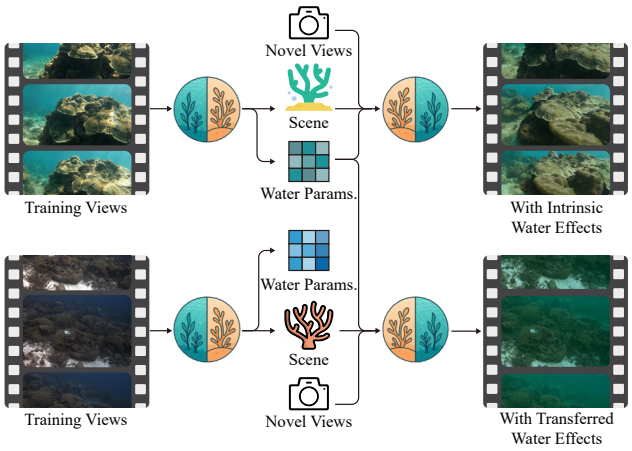

Figure 9: Applications enabled by disentangled scene and water representations. Top: Rendering novel views with intrinsic water effects from the original scene. Bottom: Transferring water effects from one scene to another to synthesize novel underwater views under different media conditions.

(2) Cross-scene Water Effect Transfer. Given two scenes (Scene 1 and Scene 2), we disentangle the water parameters and scene content for both. By applying Scene 1's water parameters to the intrinsic representation of Scene 2, we can render novel underwater views of Scene 2 under the visual conditions of Scene 1. This enables the simulation of diverse underwater environments, supports domain-specific data generation, and provides a new strategy for constructing synthetic underwater datasets with known ground-truth scene geometry and content.

5 Conclusion

We present MarineSTD-GS, a 3D Gaussian-based framework that explicitly models spatiotemporal degradation in underwater videos and reconstructs realistic scene representations via self-supervised learning. By combining a dual-Gaussian design with a physically grounded Spatiotemporal Degradation Modeling (SDM) module, our method jointly learns both the intrinsic scene and associated degradation factors. The proposed Depth-Guided Geometry Loss improves geometric reconstruction, while the Multi-Stage Optimization strategy stabilizes training and enhances texture fidelity. Experiments on our simulated dataset—covering diverse scene types, caustic patterns, and degradation levels—as well as real-world scenes demonstrate that MarineSTD-GS consistently outperforms existing methods, achieving state-of-the-art performance in color correction, visibility enhancement, and suppression of transient lighting artifacts. Finally, our joint modeling of scene and degradation parameters enables extended applications, such as novel view synthesis with controllable water effects and cross-scene water parameter transfer, highlighting the its versatility for si mulation, virtual reality, and data generation in underwater environments.

References

[1] Panagiotis Agrafiotis, Dimitrios Skarlatos, Timothy Forbes, Charalambos Poullis, Margarita Skamantzari, and Andreas Georgopoulos. 2018. Underwater photogrammetry in very shallow waters: main challenges and caustics effect removal. (2018).

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

[3] Derya Akkaynak and Tali Treibitz. 2019. Sea-thru: A method for removing water from underwater images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 1682–1691.

[4] Nantheera Anantrasirichai. 2024. BVI-Coral: Underwater scenes for 3D reconstruction. doi:10.5281/zenodo.11093417

[5] Mohamed Badran and Marwan Torki. 2023. DAUT: Underwater Image Enhancement Using Depth Aware U-shape Transformer. In *2023 IEEE International Conference on Image Processing (ICIP)*. IEEE, 1830–1834.

[6] Dana Berman, Deborah Levy, Shai Avidan, and Tali Treibitz. 2020. Underwater single image color restoration using haze-lines and a new quantitative dataset. *IEEE transactions on pattern analysis and machine intelligence* 43, 8 (2020), 2822–2837.

[7] Clémentin Boittiaux, Ricard Marxer, Claire Dune, Aurélien Arnaubec, Maxime Ferrera, and Vincent Hugel. 2024. SUCRe: Leveraging scene structure for underwater color restoration. In *2024 International Conference on 3D Vision (3DV)*. IEEE, 1488–1497.

[8] Paulo LJ Drews, Erickson R Nascimento, Silvia SC Botelho, and Mario Fernando Montenegro Campos. 2016. Underwater depth estimation and image restoration based on single images. *IEEE computer graphics and applications* 36, 2 (2016), 24–35.

[9] Guangchi Fang and Bing Wang. 2024. Mini-Splatting: Representing Scenes with a Constrained Number of Gaussians. *arXiv preprint arXiv:2403.14166* (2024).

[10] Ben Fei, Jingyi Xu, Rui Zhang, Qingyuan Zhou, Weidong Yang, and Ying He. 2024. 3d gaussian splatting as new era: A survey. *IEEE Transactions on Visualization and Computer Graphics* (2024).

[11] Timothy Forbes, Mark Goldsmith, Sudhir Mudur, and Charalambos Poullis. 2018. DeepCaustics: Classification and removal of caustics from underwater imagery. *IEEE Journal of Oceanic Engineering* 44, 3 (2018), 728–738.

[12] Rafael Garcia, Tudor Nicosevici, and Xevi Cufí. 2002. On the way to solve lighting problems in underwater imaging. In *OCEANS'02 MTS/IEEE*, Vol. 2. IEEE, 1018–1024.

[13] Yuanhao Gong. 2024. Eggs: Edge guided gaussian splatting for radiance fields. In *Proceedings of the 29th International ACM Conference on 3D Web Technology*. 1–5.

[14] Nuno Gracias, Shahriar Negahdaripour, Laszlo Neumann, Ricard Prados, and Rafael Garcia. 2008. A motion compensated filtering approach to remove sunlight flicker in shallow water images. In *OCEANS 2008*. IEEE, 1–7.

[15] Junha Hyung, Susung Hong, Sungwon Hwang, Jaeseong Lee, Jaegul Choo, and Jin-Hwa Kim. 2024. Effective Rank Analysis and Regularization for Enhanced 3D Gaussian Splatting. *arXiv preprint arXiv:2406.11672* (2024).

[16] Matthew Johnson-Roberson, Mitch Bryson, Ariell Friedman, Oscar Pizarro, Giancarlo Troni, Paul Ozog, and Jon C Henderson. 2017. High-resolution underwater robotic vision-based mapping and three-dimensional reconstruction for archaeology. *Journal of Field Robotics* 34, 4 (2017), 625–643.

[17] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Trans. Graph.* 42, 4 (2023), 139–1.

[18] Deborah Levy, Amit Peleg, Naama Pearl, Dan Rosenbaum, Derya Akkaynak, Simon Korman, and Tali Treibitz. 2023. SeaThru-NeRF: Neural radiance fields in scattering media. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 56–65.

[19] Chongyi Li, Chunle Guo, Wenqi Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao. 2019. An underwater image enhancement benchmark dataset and beyond. *IEEE transactions on image processing* 29 (2019), 4376–4389.

[20] Huapeng Li, Wenxuan Song, Tianao Xu, Alexandre Elsig, and Jonas Kulhanek. 2024. WaterSplatting: Fast Underwater 3D Scene Reconstruction Using Gaussian Splatting. *arXiv preprint arXiv:2408.08206* (2024).

[21] Jiahe Li, Jiawei Zhang, Xiao Bai, Jin Zheng, Xin Ning, Jun Zhou, and Lin Gu. 2024. Dngaussian: Optimizing sparse-view 3d gaussian radiance fields with global-local depth normalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 20775–20785.

[22] Jiaqi Lin, Zhihao Li, Xiao Tang, Jianzhuang Liu, Shiyong Liu, Jiayue Liu, Yangdi Lu, Xiaofei Wu, Songcen Xu, Youliang Yan, et al. 2024. Vastgaussian: Vast 3d gaussians for large scene reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5166–5175.

[23] Risheng Liu, Zhiying Jiang, Shuzhou Yang, and Xin Fan. 2022. Twin adversarial contrastive learning for underwater image enhancement and beyond. *IEEE Transactions on Image Processing* 31 (2022), 4922–4936.

[24] Gerard Llorach-Tó, Enoc Martínez, Joaquín Del Río Fernández, and Emilio García-Ladona. 2023. Experience OBSEA: a web-based 3D virtual environment of a seafloor observatory. In *OCEANS 2023-Limerick*. IEEE, 1–6.

[25] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Commun. ACM* 65, 1 (2021), 99–106.

[26] Zak Murez, Tali Treibitz, Ravi Ramamoorthi, and David Kriegman. 2015. Photometric stereo in a scattering medium. In *Proceedings of the IEEE international conference on computer vision*. 3415–3423.

[27] Andrea Ramazzina, Mario Bijelic, Stefanie Walz, Alessandro Sanvito, Dominik Scheuble, and Felix Heide. 2023. Scatternerf: Seeing through fog with physically-based inverse neural rendering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 17957–17968.

[28] Yelena Randall. 2023. *Flsea: Underwater visual-inertial and stereo-vision forward-looking datasets*. Master's thesis. University of Haifa (Israel).

[29] Jonathan Sauder and Devis Tuia. 2024. Self-Supervised Underwater Caustics Removal and Descattering via Deep Monocular SLAM. In *European Conference on Computer Vision*. Springer, 214–232.

[30] Johannes L Schonberger and Jan-Michael Frahm. 2016. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4104–4113.

[31] Gaurav Sharma, Wencheng Wu, and Edul N Dalal. 2005. The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *COLOR research & application* 30, 1 (2005), 21–30.

[32] ASM Shihavuddin, Nuno Gracias, and Rafael Garcia. 2012. Online Sunflicker Removal using Dynamic Texture Prediction.. In *VISAPP (1)*. 161–167.

[33] Yunkai Tang, Chengxuan Zhu, Renjie Wan, Chao Xu, and Boxin Shi. 2024. Neural Underwater Scene Representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 11780–11789.

[34] Zhengyong Wang, Liquan Shen, Mai Xu, Mei Yu, Kun Wang, and Yufei Lin. 2023. Domain adaptation for underwater image enhancement. *IEEE Transactions on Image Processing* 32 (2023), 1442–1457.

[35] Haolin Xiong, Sairisheek Muttukuru, Rishi Upadhyay, Pradyumna Chari, and Achuta Kadambi. 2023. Sparsegs: Real-time 360 {\deg} sparse view synthesis using gaussian splatting. *arXiv preprint arXiv:2312.00206* (2023).

[36] Congrong Xu, Justin Kerr, and Angjoo Kanazawa. 2024. Splatfacto-w: A nerfstudio implementation of gaussian splatting for unconstrained photo collections. *arXiv preprint arXiv:2407.12306* (2024).

[37] Xinwei Xue, Tianjiao Ma, Yidong Han, Long Ma, and Risheng Liu. 2023. Learning Deep Scene Curve for Fast and Robust Underwater Image Enhancement. *IEEE Signal Processing Letters* (2023).

[38] Daniel Yang, John J Leonard, and Yogesh Girdhar. 2024. Seasplat: Representing underwater scenes with 3d gaussian splatting and a physically grounded image formation model. *arXiv preprint arXiv:2409.17345* (2024).

[39] Lihe Yang, Bingyi Kang, Zilong Huang, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. 2024. Depth anything: Unleashing the power of large-scale unlabeled data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10371–10381.

[40] Vickie Ye, Ruilong Li, Justin Kerr, Matias Turkulainen, Brent Yi, Zhuoyang Pan, Otto Seiskari, Jianbo Ye, Jeffrey Hu, Matthew Tancik, et al. 2024. gsplat: An open-source library for Gaussian splatting. *arXiv preprint arXiv:2409.06765* (2024).

[41] Matan Yuval and Tali Treibitz. 2024. Releasing a dataset of 3D models of artificial reefs from the northern red-sea for 3D printing and virtual reality applications. *Remote Sensing Applications: Society and Environment* 36 (2024), 101305.

[42] Dongbin Zhang, Chuming Wang, Weitao Wang, Peihao Li, Minghan Qin, and Haoqian Wang. 2025. Gaussian in the wild: 3d gaussian splatting for unconstrained image collections. In *European Conference on Computer Vision*. Springer, 341–359.

[43] Jiahui Zhang, Fangneng Zhan, Muyu Xu, Shijian Lu, and Eric Xing. 2024. Fregs: 3d gaussian splatting with progressive frequency regularization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 21424–21433.

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

[45] Tianyi Zhang, Weiming Zhi, Kaining Huang, Joshua Mangelson, Corina Barbalata, and Matthew Johnson-Roberson. 2024. RecGS: Removing Water Caustic with Recurrent Gaussian Splatting. *arXiv preprint arXiv:2407.10318* (2024).

[46] Zheng Zhang, Wenbo Hu, Yixing Lao, Tong He, and Hengshuang Zhao. 2024. Pixel-gs: Density control with pixel-aware gradient for 3d gaussian splatting. *arXiv preprint arXiv:2403.15530* (2024).

[47] Jiajia Zhou, Junbin Zhuang, Yan Zheng, Yasheng Chang, and Suleman Mazhar. 2024. HIFI-Net: A Novel Network for Enhancement to Underwater Optical Images. *IEEE Signal Processing Letters* 31 (2024), 885–889.

[48] Zehao Zhu, Zhiwen Fan, Yifan Jiang, and Zhangyang Wang. 2025. Fsgs: Real-time few-shot view synthesis using gaussian splatting. In *European Conference on Computer Vision*. Springer, 145–163.