SMP-UWGS: Coupled Physics-Geometry Optimization for Scalable Multi-Partition Underwater 3D Reconstruction

Yuxuan Liu[1] and Jinhui Zhang[1]⋆

Beijing Institute of Technology, Beijing, China
3120250892@bit.edu.cn, zhangjinh@bit.edu.cn

https://github.com/vicliuuuu/SMP-UWGS

Abstract. We propose SMP-UWGS, an underwater reconstruction framework based on 3D Gaussian Splatting (3DGS) that couples scalable multi-partition geometry with physical light transport modeling. Our method introduces two key components: SMP-GAUSSIAN, a multi-partition architecture enabling distributed optimization for large-scale scenes, and DPR-Net (Differentiable Physical Rendering Network), which combines WaterParamPredict module for estimating water optical parameters with a Dual-Branch Differential Refinement (DBDR) module to model attenuation and backscattering. With tailored loss functions, the framework jointly estimates water optical parameters and improves color fidelity and geometric accuracy. Experiments on public datasets show that SMP-UWGS achieves state-of-the-art efficiency while enabling scalable, high-fidelity underwater reconstruction, benefiting applications such as ecological monitoring, habitat mapping, and autonomous underwater robotics.

Keywords: 3D Gaussian Splatting · Underwater 3D Reconstruction · Physics-Aware Rendering · Multi-Partition Gaussian

1 Introduction

The ocean covers the vast majority of Earth's surface, yet its underwater realm remains one of the planet's most immense visual frontiers. Thousands of underwater sites and incidents require detailed mapping annually, ranging from the UNESCO-estimated millions of shipwrecks and archaeological sites to the thousands of marine casualties reported by the IMO. These urgent needs in archaeology, safety, and ecological monitoring drive the critical demand for large-scale, high-fidelity underwater 3D reconstruction.

Neural rendering has revolutionized scene reconstruction, with NeRF [27] enabling detailed geometry recovery via continuous volumetric fields, and 3D Gaussian Splatting (3DGS) [19] offering real-time optimization and rendering through

⋆ Corresponding author.

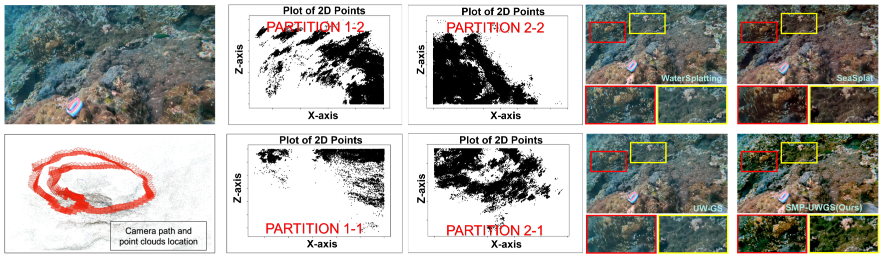

Fig. 1: Overview of the SMP-UWGS framework for large-scale underwater reconstruction. The estimated camera trajectory and sparse geometry are projected onto the XZ plane for spatial partitioning. An optical compensation module based on DPR-Net is applied to each partition, and the optimized partitions are integrated into a globally consistent 3D representation.

explicit anisotropic Gaussian primitives. However, extending these paradigms to large-scale underwater environments remains challenging.

On one side, physics-aware approaches such as SeaThru-NeRF [22] explicitly model wavelength-dependent attenuation and scattering to recover physically consistent colors. However, their reliance on dense volumetric sampling and global optimization of optical parameters leads to prohibitive GPU memory consumption and training time when scaling to large scenes. On the other side, recent large-scale explicit representations achieve scalability through spatial partitioning and point-based pruning. However, by treating the medium as uniform air, they neglect the depth-dependent nature of underwater light transport, which results in significant color distortions and degraded reconstruction quality in deeper regions.

This impasse arises from an overlooked coupling: underwater radiative transport is inherently depth-conditioned and thus inseparable from scene geometry. Since attenuation and backscatter depend explicitly on camera-to-surface distance, spatial partitioning and physical modeling cannot be treated as independent stages. A partitioning strategy that ignores depth continuity fragments the attenuation field, while a global physical model that ignores spatial locality becomes computationally prohibitive at scale.

To bridge this gap, we argue that scalability and physical consistency must be jointly optimized rather than sequentially combined. Fig. 1 provides an overview of the proposed SMP-UWGS framework. We therefore propose SMP-UWGS, a unified framework that reformulates large-scale underwater reconstruction as a depth-aware radiative Gaussian learning problem. This formulation necessitates coordinated design across representation, physical modeling, and optimization. At the representation level, we introduce a partition-driven Gaussian optimization strategy that enables region-wise training while preserving cross-region depth and visibility coherence, ensuring scalability without disrupting geometric continuity. At the physical modeling level, we embed a depth-conditioned differentiable light transport formulation directly into the Gaussian primitives. A two-stage estimation mechanism provides physically consistent initialization of attenuation and backscatter, followed by joint refinement to maintain radiative balance across depth. At the optimization level, we adopt a hybrid physical–statistical objective that alternates radiative reconstruction with structural and optical regularization, stabilizing the simultaneous learning of geometry and medium properties. Through this unified design, SMP-UWGS treats spatial scalability and underwater light transport as a single constrained learning problem, rather than independent modules stitched together post hoc.

We conduct comparisons and ablation studies on the SeaThru-NeRF [22], BVI-Coral [36], and UVEB [39] datasets, as well as our OTNN dataset. Across all large-scale scenes, SMP-UWGS consistently improves reconstruction fidelity while reducing memory consumption and training time, achieving state-of-the-art performance in PSNR, SSIM, and LPIPS.

Our contributions are summarized as follows:

1) We reformulate large-scale underwater reconstruction as a depth-aware radiative Gaussian learning problem and propose a partition-driven Gaussian representation with cross-region depth and visibility coherence, enabling scalable optimization while preserving global geometric consistency.
2) We introduce a depth-conditioned differentiable light transport formulation embedded within Gaussian primitives, jointly modeling attenuation and backscatter to achieve physically grounded radiance recovery.
3) We design a hybrid physical–statistical optimization objective that stabilizes the coupled estimation of geometry and medium properties, ensuring robust convergence and high-fidelity reconstruction under complex underwater degradation.

2 Related Works

2.1 Underwater Imaging

Underwater environments often suffer from color cast, image distortion, and contrast loss due to wavelength-dependent light absorption and scattering [14]. Owing to the similarity between haze and underwater imaging, classical enhancement methods such as the dark-channel prior [15], gray-world assumption [12], and histogram equalization have been adapted. To better capture underwater light behavior, several works employ physical modeling, including light propagation [9], radiative transfer theory [11], and light reflectance models [9]. The Revised Underwater Image Formation Model (UIFM) [1] further separates direct and backscattered components, and has been widely adopted in recent restoration studies [17, 18, 42]. However, single-image enhancement often produces inconsistent color and geometry, limiting its applicability to 3D reconstruction [33].

Neural approaches such as SeaThru-NeRF [22] and UW-GS [36] address scattering effects via physically grounded models, but real-time reconstruction remains a challenge.

2.2 Large Scene Reconstruction

Large-scale scene reconstruction has evolved from traditional structure-from-motion (SfM) [30] and multi-view stereo (MVS) [13] to NeRF–based methods [27]. To improve scalability, Block-NeRF [32] and Mega-NeRF [35] divide scenes into spatial partitions, while Switch-NeRF [43] adopts a mixture-of-experts design and Grid-NeRF [40] combines neural and grid-based representations. Despite these advances, rendering remains slow and fine details are often lost. The emergence of 3DGS [19] enabled real-time rendering but faced memory limitations for large scenes. VastGaussian [24] alleviates this through progressive spatial partitioning and parallel optimization, achieving high-fidelity reconstruction at urban scale [25]. Appearance variations are further managed by decoupled modeling—NeRF-W [26] introduces appearance embeddings, while VastGaussian [24] applies a CNN-based photometric correction during training. Overall, these studies highlight a trend toward integrating physical modeling with scalable optimization for large, dynamically lit environments [31, 37].

2.3 3D Gaussian Splatting

3DGS has emerged as an explicit and efficient representation for 3D scenes, enabling high-fidelity real-time rendering [19]. Unlike NeRF [27], which rely on implicit volumetric sampling and suffer from slow optimization, 3DGS models scenes with trainable Gaussian primitives for fast rasterization [7]. Advances in this field include multiscale anti-aliasing [3, 4], grid-based acceleration [5, 6], deblurring techniques [8, 21], and compact or dynamic extensions [23, 38]. However, 3DGS remains challenged in scattering media. To address this, methods such as scatterNeRF [28] and SeaThru-NeRF [22] integrate physical light models for foggy and underwater scenes, while UW-GS [36] adapt 3DGS with color appearance modeling and volumetric light transport. These efforts demonstrate the versatility of 3DGS, yet underline its need for domain-specific adaptation in complex optical environments.

3 Method

This section presents an extended 3DGS framework that integrates a physical light transport model for scalable underwater reconstruction. Specifically, Subsection 3.1 introduces a scalable multi-partition Gaussian architecture for memory-efficient large-scene modeling. Subsection 3.2 presents DPR-Net, a differentiable physical rendering module consisting of WaterParamPredict and Dual-Branch Differential Refinement (DBDR) for attenuation and backscatter compensation. Finally, Subsection 3.3 describes the loss functions used for stable underwater optimization.

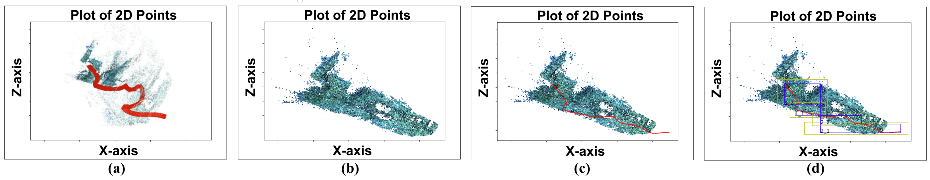

Fig. 2: Visualization of the proposed spatial partitioning strategy on the MingWreck2 scene (OTNN dataset): (a) reconstructed 3D point cloud and cameras' location, (b) projection onto the XZ plane, (c) camera trajectory with initial spatial partitions, and (d) boundary extension guided by camera poses.

3.1 Scalable Multi-Partition Gaussian Architecture for Regional-Scale Modeling

As shown in Figure 2, we propose a scalable multi-partition Gaussian architecture for large-scale underwater scene reconstruction, which adapts to complex camera distributions.

Dynamic Spatial Partitioning Strategy. The reconstruction process begins by projecting camera coordinates onto the x-z plane, effectively simplifying vertical variations while focusing on horizontal distribution patterns. Given a set of training cameras $\mathcal{C} = \{c_1, c_2, \ldots, c_N\}$ with positions $p_i = (x_i, y_i, z_i)$, we first sort the cameras by their x-coordinate.The sorted camera set is then divided into m spatial partitions, each containing approximately N/m cameras. To ensure seamless transitions between adjacent partitions, we designate the rightmost camera in each partition as a boundary camera B_x^i. For each x-axis partition $\mathcal{C}_i^m$, we further refine the spatial division along the z-axis into n sub-regions.The camera distribution across sub-regions is defined as:

$$\mathcal{C}_{ij}^{mn} = \mathcal{C}_i^z \left[\lfloor N_i/n \rfloor \cdot (j-1) : \lfloor N_i/n \rfloor \cdot j\right] \text{for } j = 1, \ldots, n-1 \tag{1}$$

The final sub-region accommodating remaining cameras. This hierarchical partitioning scheme generates $m \times n$ overlapping sub-regions with balanced camera distributions, where shared boundary cameras maintain geometric consistency throughout the reconstruction volume.

Boundary-Aware Region Enhancement. To address potential discontinuities at partition boundaries, we implement a comprehensive region enhancement strategy based on expanded spatial domains. Starting from original partition boundaries, we generate extended bounding boxes with adaptive expansion rates:

$$\begin{aligned} \text{extend\_camera\_bbox} = [\, &\min_x - e \cdot w,\ \max_x + e \cdot w, \\ &\min_z - e \cdot h,\ \max_z + e \cdot h] \end{aligned} \tag{2}$$

Edge partitions use an extended rate of 2.0 to enhance boundary coverage, while internal regions use 1.5, relative to the base extension unit.The architecture implements a three-tier Gaussian fusion strategy to ensure comprehensive scene coverage:

- **Global Gaussian Layer**: Maintains complete scene representation for global consistency and coherence.
- **Regional Gaussian Layer**: Refines local features within extended boundaries through Gaussian stitching, defined as $\mathcal{G}_{\text{enhanced}} = \mathcal{G}_{\text{global}} \cup \mathcal{G}_{\text{focus}}$.
- **Neighborhood Gaussian Layer**: Ensures explicit boundary continuity across edge partitions by incorporating adjacent region information.

The effectiveness of local feature enhancement is quantified by the Gaussian focusing ratio $r_{\text{focus}} = \frac{|\mathcal{G}_{\text{focus}}|}{|\mathcal{G}_{\text{enhanced}}|}$, which measures the proportion of focused Gaussian elements within the enhanced region.

Cross-Partition Coherence Mechanism. To ensure geometric consistency across partition boundaries, we introduce a visibility-based camera selection mechanism. Specifically, we project the 3D bounding volume of each partition onto the image planes of all candidate cameras and compute the visibility coverage rate, denoted as ρ. We then select cameras that satisfy a global coverage threshold τ (i.e., $\rho \geq \tau$) to participate in the joint optimization of adjacent partitions. To enhance boundary coverage, edge partitions employ a more relaxed threshold, defined as $\tau_e = 0.3\tau$. Consequently, cross-region Gaussian fusion follows this visibility criterion:

$$\mathcal{G}_{ij} = \{\mathbf{g} \in G_j \mid \pi_{C_j}(\mathbf{g}) \in I\} \tag{3}$$

Edge partitions employ relaxed projection $\pi_{C_j}^{\text{relaxed}}$ for extended boundary handling. The final Gaussian representation for each partition is obtained through comprehensive deduplication:

$$\mathcal{G}_i^{\text{final}} = \text{Unique}\left(\bigcup_{j=1}^{N} \mathcal{G}_{ij} \cup \mathcal{G}_i\right) \tag{4}$$

Depth-Aware Region Weighting Strategy. To compensate for depth-dependent attenuation in underwater scenes, we introduce a region-wise weighting scheme that integrates visibility and depth statistics into the optimization process.

For each partition i, the regional weight is defined as

$$w_i = (\bar{r}_i + \beta)\left(1 + \tau_{init}\, \log(1 + d_i')\right) \gamma_{geo}, \tag{5}$$

where $\bar{r}_i$ denotes the average visibility ratio estimated via bounding-box projection and intersection analysis, $\beta = 0.01$ prevents degenerate weights, and d_i' represents the normalized regional depth after scale normalization. τ_{init} controls the depth amplification strength and provides a stable initialization prior; it is later refined by the WaterParamPredict network to estimate spatially-varying attenuation parameters. The geometric factor γ_{geo} enhances sparse boundary regions (2.0 for corner partitions, 1.5 for edge partitions, and 1.0 otherwise).

The regional depth is computed in camera coordinates as

$$\bar{d}_i = \text{median}\{z_c \mid \mathbf{p} \in \mathcal{P}_i,\ z_c > \epsilon\}, \tag{6}$$

where z_c denotes the depth value in the camera coordinate system. All final weights are clamped to $[0, 1]$ to ensure numerical stability.

Depth-Adaptive Gradient Modulation. The total objective is defined as $\mathcal{L}_{total} = \sum_i w_i \mathcal{L}_i$, Thus, w_i directly modulates the effective gradient magnitude of each region during backpropagation.

Due to underwater attenuation, deeper regions inherently suffer from reduced signal strength and weaker reconstruction gradients. By assigning larger weights to deeper partitions, the proposed scheme compensates for attenuation-induced gradient decay and encourages balanced optimization across depth layers. Shallow regions, which are typically well-observed, receive relatively smaller amplification, preventing overfitting to high-SNR areas. Importantly, this mechanism achieves depth-adaptive optimization without modifying the internal state of the optimizer. The underwater image $I(x, y)$ is formed from the radiance field

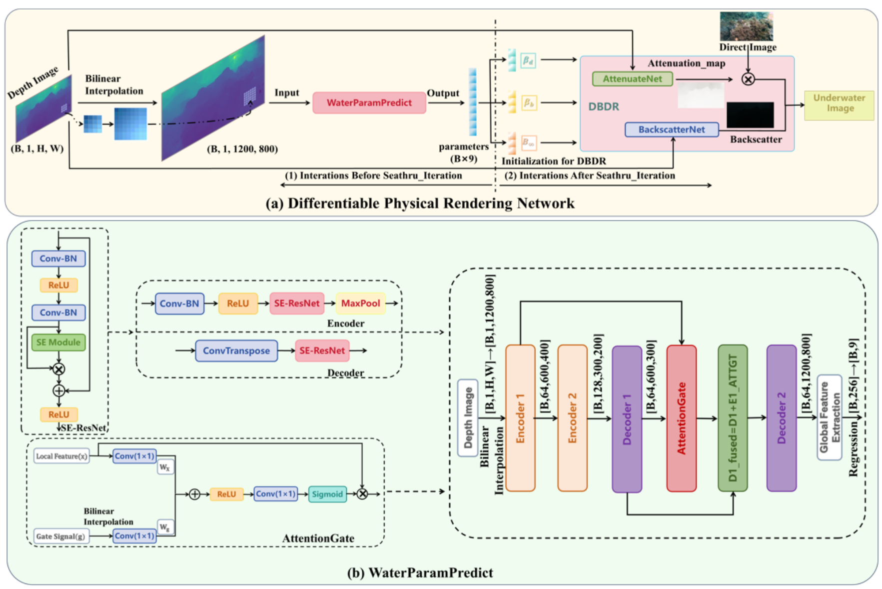

Fig. 3: Overview of the DPR-Net pipeline. (**a**) *System Framework:* DPR-Net embeds into 3DGS to recover underwater radiance. WaterParamPredict initializes optical parameters from depth, which are then jointly refined by the parallel DBDR module AttenuateNet + BackscatterNet) to ensure color fidelity. (**b**) *WaterParamPredict Architecture:* A U-Net backbone enhanced with SE-ResNet encoders and attention-gated decoders. It regresses nine global optical parameters ($\beta_d, \beta_b, B_\infty$) from depth maps to provide physically consistent initialization.

$J(x, y)$ rendered by 3DGS, with depth $z(x, y)$ affecting light attenuation and backscattering. The image formation equation is [2, 22]:

$$I(x, y) = J(x, y) \cdot e^{-\beta_d(x,y)\cdot z(x,y)} + B_\infty(x, y) \left(1 - e^{-\beta_b(x,y)\cdot z(x,y)}\right) \quad (7)$$

This differentiable formulation enables gradient propagation from image observations to both optical parameters and Gaussian primitives within the 3DGS framework. While the partitioning strategy improves efficiency and preserves local feature structures, underwater scenes still require dedicated modeling of depth-dependent attenuation and backscatter. To address this issue, we introduce a physics-guided optical estimation module inspired by SeaThru, termed DPR-Net (Figure 3(a)). The training process is organized into three progressive stages: (1) global parameter prediction via WaterParamPredict, (2) local light transport refinement through DBDR, and (3) partition-accelerated 3D reconstruction.

WaterParamPredict. WaterParamPredict follows a U-Net architecture [29] augmented with SE-ResNet units and attention gating to estimate global underwater optical parameters, including the attenuation coefficient β_d, backscatter coefficient β_b, and background light B_∞ from the input depth map. The input depth map is not obtained from an external sensor. Instead, it is rendered from the current Gaussian representation by projecting Gaussian centers into the camera coordinate system and using their depth values as rendering colors. This produces a differentiable depth map that is consistent with the evolving 3DGS geometry.

As illustrated in Figure 3(b), the encoder employs SE-ResNet modules [16,20,34] to progressively extract depth-aware features while adaptively recalibrating channel responses. The decoder restores spatial resolution through upsampling and integrates attention gating [10] to selectively fuse encoder features based on high-level semantic guidance. This mechanism enables the network to emphasize regions that are most informative for optical parameter estimation. Finally, convolution layers followed by a Sigmoid activation generate the parameter maps, allowing the model to focus on image regions most relevant to underwater optical effects.

3.2 Physically-Guided Rendering Pipeline with Differentiable Optical Parameter Estimation

After two decoding stages, the network maps the fused high-level features to a 9-dimensional vector $\mathbf{p} = (p_0, \dots, p_8)$ through a linear layer, which represents the RGB attenuation coefficients, backscatter coefficients, and background light parameters. The physical parameters are then obtained using activation constraints:

$$\beta_d^c = \mathrm{softplus}(p_c), \beta_b^c = \mathrm{softplus}(p_{c+3}), B_\infty^c = \sigma(p_{c+6}), \quad c \in \{R, G, B\}. \tag{8}$$

where softplus$(\cdot)$ guarantees positive attenuation and scattering coefficients, and $\sigma(\cdot)$ bounds the background light within $[0, 1]$.

DBDR. The pretraining stage stabilizes joint optimization by reducing the sensitivity of exponential terms $e^{-\beta z}$ to optical parameter variations. To address spatial variations in turbidity and illumination, we introduce a DBDR mechanism, where attenuation and backscatter fields are jointly refined from the global parameters predicted by WaterParamPredict.

Specifically, we estimate spatially varying attenuation and backscatter coefficients using a shared convolutional predictor:

$$[\beta_d(x,y), \beta_b(x,y)] = \mathrm{ReLU}(\mathrm{Scale} \cdot \sigma(W * z(x,y))) \tag{9}$$

where $z(x,y)$ denotes the depth map and W represents shared convolutional weights producing two-channel outputs corresponding to attenuation and backscatter coefficients.

3.3 Hybrid Physical-Statistical Loss for Robust Underwater Optimization

Underwater scene reconstruction is challenged by degradation from light attenuation, backscatter, and turbidity, which introduce ambiguity in recovering geometry and appearance. To resolve this, we design a hybrid physical-statistical loss that combines reconstruction fidelity with statistical priors to disambiguate underwater light transport effects.

We address view-dependent degradation using feature-wise linear modulation (FiLM). Given an underwater image I_u and a view embedding $\mathbf{e}_i$ associated with the i-th camera view, the network produces a decoupled image:

$$I_d = \mathcal{F}(I_u, \mathbf{e}_i) \odot I_u \tag{10}$$

The physical loss operates on the decoupled output:

$$\mathcal{L} = (1-\lambda)\|I_d - I_{gt}\|_1 + \lambda(1 - \mathrm{SSIM}(I_d, I_{gt})) \tag{11}$$

where I_{gt} denotes the ground-truth image. This formulation separates water-induced effects from scene content while preserving reconstruction fidelity.

To restore color fidelity and mitigate visibility degradation caused by underwater light scattering and absorption, we introduce a Physics-Aware Dark Channel Prior (PA-DCP). Standard DCP often fails in underwater environments by misinterpreting the naturally attenuated red channel as dense turbidity, leading to severe color casts (e.g., excessive blue-green shifts). To alleviate this issue, we compute the backscatter-compensated signal $\hat{J} = \hat{I} - B$, where $\hat{I}$ denotes the estimated radiance and B represents the backscatter component. Channel contributions are then reweighted using normalized adaptive weights $w_c \propto (\beta_d^c + \epsilon)^{-1}$, where β_d^c denotes the attenuation coefficient of channel c. This reduces the influence of heavily attenuated channels (typically Red) during dark channel estimation.

Formally, the physics-aware dark channel map D_{PA} is computed as

$$D_{PA}(\mathbf{x}) = \min_c \left(\min_{\mathbf{y} \in \Omega(\mathbf{x})} \left(w_c \cdot \hat{J}^c(\mathbf{y}) \right) \right), \tag{12}$$

where $\Omega(\mathbf{x})$ denotes a local patch. To stabilize the estimation, we impose a color consistency loss:

$$\mathcal{L}_{\mathrm{color}} = \lambda_{\mathrm{dcp}} \left(\|\mathrm{ReLU}(D_{PA})\|_1 + \lambda_{\mathrm{neg}} \|\mathrm{ReLU}(-D_{PA})\|_{\mathrm{smooth}} \right), \tag{13}$$

where the L_1 term suppresses residual backscatter in the dark channel, while the Smooth-L_1 term ($\lambda_{\text{neg}} = 1000$) penalizes large negative responses caused by over-compensation.

To improve geometric coherence, we introduce an edge-aware depth smoothness loss:

$$\mathcal{L}_{\text{depth}} = \lambda_d \sum_{x,y} w_{x,y} \left\| \nabla D_{x,y} \right\|_1 , \tag{14}$$

where $w_{x,y} = \exp(-\|\nabla I_{x,y}\|)$ reduces smoothing near strong image gradients. This encourages locally smooth depth while preserving discontinuities aligned with image edges.

To handle transparency and background consistency, we incorporate an alpha-background regularization term that stabilizes foreground–background blending:

$$\mathcal{L}_{bg} = \lambda_b \mathcal{L}_\alpha(\hat{I}, \text{learned\_bg}, \alpha), \tag{15}$$

where α denotes the opacity map used in alpha compositing.

To maintain physically plausible underwater optical parameters, we apply mild L_2 regularization:

$$\mathcal{L}_{\text{param}} = \lambda_p \left(\|\beta_d\|_2^2 + \|\beta_b\|_2^2 + \|B_\infty\|_2^2 \right), \tag{16}$$

which prevents degenerate solutions during optimization.

The total loss is defined as

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{phys}} + \mathcal{L}_{\text{color}} + \mathcal{L}_{\text{depth}} + \mathcal{L}_{bg} + \mathcal{L}_{\text{param}}. \tag{17}$$

Training proceeds in stages: we first optimize the physical rendering loss $\mathcal{L}_{\text{phys}}$, and gradually introduce the remaining terms after the rendering process stabilizes. This hybrid objective combines physical consistency with statistical regularization to improve color stability, geometric smoothness, and foreground–background interaction in underwater reconstruction.

4 Experimental Configuration

Datasets. Evaluation is conducted on four underwater datasets covering diverse marine environments. SeaThru-NeRF and BVI-Coral [22, 36] provide representative scenes under different water conditions. UVEB [39] contains terrain exploration and marine biology sequences for evaluating robustness in complex environments. OTNN (our dataset) includes real-world underwater shipwreck and archaeological scenes, such as the Blue Thistle WWII wreck in the Red Sea and several Ming Dynasty shipwrecks in the South China Sea.

Radiometric Linearization. Since our physical image formation model relies on the linearity of light transport, we standardize all input imagery to linear radiance space prior to training. Consistent with standard practices in neural rendering (e.g., NeRF, 3DGS), images from consumer cameras and public benchmarks (assumed to be encoded in sRGB) are converted using the inverse IEC

Table 1: Quantitative comparison with state-of-the-art methods on underwater novel view synthesis. Performance is evaluated using PSNR, SSIM, and LPIPS across four benchmark datasets. Top two results are highlighted in red and yellow.

SeaThru-NeRF												
Scene	**IUI3 Red Sea**			**Curaçao**			**J.G. Red Sea**			**Panama**		
Method	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
SeaThru-NeRF [22]	25.908	0.785	0.304	30.193	0.873	0.210	21.841	0.767	0.249	27.846	0.834	0.224
3DGS [19]	22.980	0.843	0.246	28.313	0.873	0.221	21.493	0.854	0.216	29.200	0.893	0.152
UW-GS [36]	27.652	0.863	0.185	30.171	0.851	0.172	22.947	0.830	0.190	30.190	0.911	0.160
SeaSplat [41]	26.670	0.870	0.210	30.300	0.900	0.190	22.700	0.870	0.180	28.760	0.900	0.150
SMP-UWGS(Ours)	27.829	0.907	0.180	31.551	0.937	0.154	23.289	0.895	0.174	31.642	0.925	0.137

BVI-Coral												
Scene	**Lagoon**			**Seabed**			**Reef**			**Marine**		
Method	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
SeaThru-NeRF	23.542	0.759	0.286	22.705	0.641	0.346	18.551	0.408	0.552	18.103	0.571	0.458
3D Gaussian	22.351	0.739	0.285	21.919	0.743	0.283	20.381	0.714	0.286	16.120	0.670	0.365
UW-GS	23.210	0.754	0.188	24.809	0.800	0.263	22.008	0.712	0.284	19.406	0.703	0.308
SeaSplat	26.250	0.779	0.289	25.346	0.774	0.252	24.316	0.754	0.258	19.986	0.710	0.336
SMP-UWGS(Ours)	25.953	0.816	0.221	24.461	0.769	0.299	23.022	0.754	0.291	20.159	0.718	0.313

	OTNN						UVEB					
Scene	**MingWreck2**			**BlueThistle**			**BioConservation**			**GeoExploration**		
Method	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
SeaThru-NeRF	22.186	0.791	0.254	23.934	0.796	0.265	19.137	0.707	0.270	17.421	0.675	0.260
3D Gaussian	20.089	0.756	0.287	21.964	0.772	0.299	18.549	0.742	0.321	15.418	0.639	0.288
UW-GS	23.719	0.797	0.229	24.008	0.802	0.352	21.980	0.721	0.272	24.442	0.775	0.372
SeaSplat	23.828	0.732	0.267	22.012	0.785	0.263	23.245	0.767	0.232	27.303	0.765	0.311
SMP-UWGS(Ours)	23.706	0.801	0.237	24.530	0.813	0.293	22.487	0.772	0.263	26.707	0.788	0.303

61966-2-1 transfer function. This ensures that optimization of physical parameters (β, B_∞) occurs in a radiometrically correct space, avoiding biases introduced by non-linear gamma encoding.

Implementation Details. The framework is implemented in PyTorch and trained on NVIDIA RTX 3090 GPUs. Initial point clouds and camera poses are obtained using COLMAP.

For small-scale scenes, DPR-Net is applied for the first 30,000 iterations. For large-scale scenes, we employ a partition-based optimization strategy with 12,000 total iterations. The training schedule includes two stages: WaterParamPredict is used before iteration 3000 to estimate global optical parameters, and the DBDR module is activated between iterations 3000 and 8000 to jointly refine attenuation and backscatter. After iteration 8000, the standard optimization process resumes.

Evaluation Metrics and Compared Methods. Comprehensive assessment employs PSNR for pixel-level accuracy, SSIM for structural similarity, and LPIPS for perceptual quality evaluation. Training time and GPU memory consumption are additionally reported to quantify computational efficiency.

We compare our method with several representative approaches, including 3DGS, underwater-adapted Gaussian splatting methods (UW-GS and SeaSplat), and the neural rendering method SeaThru-NeRF.

5 Results Analysis

Efficiency and memory. As shown in Fig. 4, the radar chart compares different methods across six normalized metrics, where outer rings indicate better performance. Our method (SMP-UWGS) achieves the largest coverage area, demonstrating superior balance between reconstruction quality and computational efficiency. We compare training time, memory consumption, and rendering speed on the MingWreck1 dataset with NVIDIA RTX 3090 GPU. SMP-UWGS achieves the best balance between efficiency and memory usage, completing training in only 13 minutes—much faster than 3DGS (42 min), SeaSplat (81 min), and SeaThru-NeRF (10 h). It also requires just 10.9 GB of VRAM, significantly less than SeaSplat (16 GB) and SeaThru-NeRF (14.6 GB), while delivering the highest rendering speed of 434 FPS.

These gains arise from the proposed multi-partition optimization and adaptive parallelization. SMP-UWGS establishes a new efficiency–performance tradeoff for underwater 3D reconstruction, achieving SOTA scalability and speed.

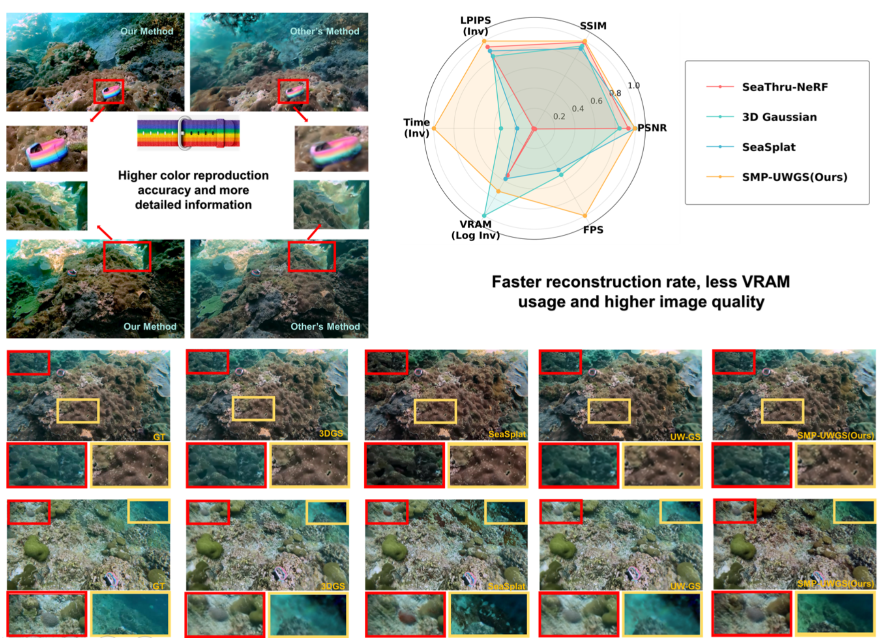

Fig. 4: Quantitative and qualitative comparison of underwater reconstruction methods. The top-left plot compares color reproduction accuracy and detail preservation among different methods. The top-right radar chart summarizes performance across multiple quantitative metrics. The bottom row presents visual reconstruction results, where the original image serves as the reference for comparison with SMP-UWGS, 3DGS, SeaSplat, and UW-GS.

Convergence and training-budget fairness. To verify that the efficiency gain is not caused by an insufficient training budget for competing methods, we further analyze PSNR convergence on the SeaBed scene using identical training and validation splits. As shown in Fig. 5, SMP-UWGS reaches stable reconstruction quality within 12k iterations and avoids large PSNR oscillations. In contrast, SeaSplat requires substantially longer optimization and still exhibits unstable intermediate reconstructions. These results indicate that the proposed physics-aware initialization and multi-partition optimization improve both convergence speed and reconstruction stability. **Quantitative results.** As shown in

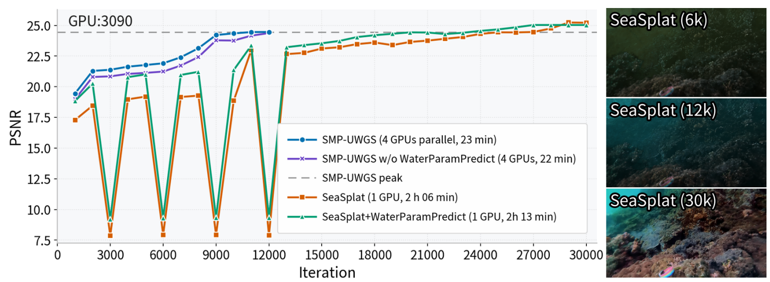

Fig. 5: Convergence and training-budget comparison on the SeaBed scene. Left: PSNR curves under different training budgets. Right: SeaSplat renderings at 6k, 12k, and 30k iterations.

Table 1, SMP-UWGS consistently outperforms prior methods across all underwater benchmarks, achieving leading performance in both perceptual quality and geometric fidelity. Experiments on the SeaThru-NeRF dataset fully demonstrate the importance of DPR-NET, validating its effectiveness in handling challenging underwater conditions. In the BVI-Coral dataset, which features highly complex terrain and intricate morphological structures that are challenging to capture, SMP-UWGS attains the highest PSNR and SSIM in most scenes, surpassing recent physically-inspired and Gaussian-based approaches. On the OTNN and UVEB datasets, SMP-UWGS maintains competitive or best performance under varying attenuation and scattering conditions, validating its scalability and adaptability to large-scale underwater environments.

Although the theoretical visibility constraints and boundary conditions within partitions may lead to the omission of some points, our method still achieves excellent performance thanks to the effective compensation mechanism and robust underwater modeling module. While the multi-partition approach may exhibit slight blurring in some areas, particularly at junctions and edges, our overall level of detail processing remains state-of-the-art.

Furthermore, as shown in Figure 4, when compared with other methods, our method performs exceptionally well in terms of color reproduction and authenticity. These results highlight that the integration of multi-partition optimization and physically-guided light transport modeling enables SMP-UWGS to produce artifact-free, high-fidelity reconstructions across diverse marine settings, setting a new standard for underwater novel view synthesis and large-scale 3D reconstruction.

Ablation study. Table 2 reports the ablation results on the Lagoon dataset, evaluating the contribution of each component in SMP-UWGS to reconstruction quality and computational efficiency.

Removing DPR-Net leads to a noticeable degradation in PSNR, SSIM, and LPIPS, highlighting the importance of physically grounded light transport modeling for accurate underwater appearance reconstruction. Further decomposing DPR-Net shows that both WaterParamPredict and DBDR contribute complementary improvements: omitting either module results in consistent performance drops, confirming their roles in stabilizing water parameter estimation and refining optical attenuation and backscattering effects. WaterParamPredict provides a physically plausible initialization for underwater optical parameters, which stabilizes subsequent joint optimization and reduces the iterations needed to reach high reconstruction quality.

For the SMP framework, removing the entire partition mechanism alters the efficiency–fidelity trade-off, particularly increasing the training time.

While auxiliary components such as DARWS and BARE provide relatively moderate gains in quantitative metrics, they contribute to improved optimization stability and rendering consistency, supporting a more compact and efficient training pipeline overall.

Table 2: Ablation study on the Lagoon dataset. Each row reports the performance after removing a specific component from the full model. DPR-Net consists of WaterParamPredict and DBDR, while the partition strategies includes DARWS and BARE along with the partition architecture.

Model Setting	**PSNR**	**SSIM**	**LPIPS**	**Time**
Baseline (3DGS)	22.351	0.739	0.285	43min
Full Model w/o SMP	26.275	0.793	0.271	1h13min
DPR-Net (Physical Modeling)				
Full Model w/o DPR-Net	23.002	0.726	0.297	10min
Full Model w/o WaterParamPredict	24.803	0.770	0.240	17min
Full Model w/o DBDR	24.970	0.783	0.235	12min
Partition Strategies				
Full Model w/o DARWS	25.230	0.796	0.230	14min
Full Model w/o BARE	23.019	0.729	0.301	14min
Full Model	**25.953**	**0.816**	**0.221**	14min

6 Conclusion and Limitations

Conclusion In this work, we propose SMP-UWGS, a unified and scalable framework for large-scale underwater scene reconstruction that integrates 3DGS with physically informed light transport modeling. By introducing a scalable multi-partition architecture together with DPR-Net, WaterParamPredict, and DBDR, our method explicitly models light attenuation and backscattering effects while preserving computational efficiency. Extensive evaluations on public datasets demonstrate that SMP-UWGS achieves favorable trade-offs between reconstruction fidelity and efficiency, making it suitable for practical underwater applications such as ecological monitoring, underwater archaeology, and rescue operations. More broadly, this study highlights the importance of embedding physics-aware modeling into explicit scene representations for robust large-scale underwater reconstruction.

Limitations Despite these contributions, several limitations remain. First, the proposed physical rendering model assumes homogeneous ambient illumination, which is common in fog-based or haze-based underwater formulations. This assumption limits direct applicability to scenarios dominated by active, co-moving light sources, where the illumination field becomes spatially non-uniform and direction-dependent. Extending the model to accommodate dynamic lighting conditions remains an important direction for future research. Second, similar to standard 3DGS-based approaches, our framework assumes a static scene during optimization. As a result, dynamic objects may introduce temporal artifacts or ghosting effects, since motion cannot be explicitly disentangled from geometry and appearance variations. Future work will explore incorporating dynamic masking strategies or 4D Gaussian representations to better handle moving elements.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant U22A2048, in part by the Hebei Natural Science Foundation under Grant 264Z1819G, and in part by the State Key Laboratory of Autonomous Intelligent Unmanned Systems under Grant ZZKF2025ZD-2-2.

References

1. Akkaynak, D., Treibitz, T.: A revised underwater image formation model. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6723–6732 (2018)
2. Akkaynak, D., Treibitz, T.: Sea-thru: A method for removing water from underwater images. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1682–1691 (2019)

3. Barron, J.T., Mildenhall, B., Tancik, M., et al.: Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 5855–5864 (2021)
4. Barron, J.T., Mildenhall, B., Verbin, D., et al.: Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5470–5479 (2022)
5. Barron, J.T., Mildenhall, B., Verbin, D., et al.: Zip-nerf: Anti-aliased grid-based neural radiance fields. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 19697–19705 (2023)
6. Chen, A., Xu, Z., Geiger, A., Yu, J., Su, H.: Tensorf: Tensorial radiance fields. In: European conference on computer vision. pp. 333–350. Springer (2022)
7. Chen, G., Wang, W.: A survey on 3d gaussian splatting. arXiv preprint arXiv:2401.03890 (2024)
8. Chen, W., Liu, L.: Deblur-gs: 3d gaussian splatting from camera motion blurred images. Proceedings of the ACM on Computer Graphics and Interactive Techniques **7**(1), 1–15 (2024)
9. Dekker, A.G., Brando, V.E., Anstee, J.M., Pinnel, N., Kutser, T., Hoogenboom, E.J., Peters, S., Pasterkamp, R., Vos, R., Olbert, C., et al.: Imaging spectrometry of water. In: Imaging spectrometry: Basic principles and prospective applications, pp. 307–359. Springer (2002)
10. Deng, W., Shi, Q., Li, J.: Attention-gate-based encoder–decoder network for automatical building extraction. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **14**, 2611–2620 (2021)
11. Dutré, P., Bekaert, P., Bala, K.: Advanced global illumination. AK Peters/CRC Press (2018)
12. Ebner, M.: Color constancy based on local space average color. Machine Vision and Applications **20**(5), 283–301 (2009)
13. Furukawa, Y., Curless, B., Seitz, S.M., Szeliski, R.: Towards internet-scale multi-view stereo. In: 2010 IEEE computer society conference on computer vision and pattern recognition. pp. 1434–1441. IEEE (2010)
14. Gonzalez-Sabbagh, S.P., Robles-Kelly, A.: A survey on underwater computer vision. ACM Computing Surveys **55**(13s), 1–39 (2023)
15. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. IEEE Transactions on Pattern Analysis and Machine Intelligence **33**(12), 2341–2353 (2010)
16. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7132–7141 (2018)
17. Huo, F., Li, B., Zhu, X.: Efficient wavelet boost learning-based multi-stage progressive refinement network for underwater image enhancement. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1944–1952 (2021)
18. Kar, A., Dhara, S.K., Sen, D., Biswas, P.K.: Zero-shot single image restoration through controlled perturbation of koschmieder's model. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 16205–16215 (2021)
19. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G., et al.: 3d gaussian splatting for real-time radiance field rendering. ACM Trans. Graph. **42**(4), 139–1 (2023)
20. Koonce, B.: Resnet 50. In: Convolutional neural networks with swift for tensorflow: image recognition and dataset categorization, pp. 63–72. Springer (2021)
21. Lee, B., Lee, H., Sun, X., Ali, U., Park, E.: Deblurring 3d gaussian splatting. In: European Conference on Computer Vision. pp. 127–143. Springer (2024)

22. Levy, D., Peleg, A., Pearl, N., et al.: Seathru-nerf: Neural radiance fields in scattering media. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 56–65 (2023)
23. Li, Z., Niklaus, S., Snavely, N., Wang, O.: Neural scene flow fields for space-time view synthesis of dynamic scenes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6498–6508 (2021)
24. Lin, J., Li, Z., Tang, X., Liu, J., Liu, S., Liu, J., Lu, Y., Wu, X., Xu, S., Yan, Y., et al.: Vastgaussian: Vast 3d gaussians for large scene reconstruction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5166–5175 (2024)
25. Lin, L., Liu, Y., Hu, Y., Yan, X., Xie, K., Huang, H.: Capturing, reconstructing, and simulating: the urbanscene3d dataset. In: European Conference on Computer Vision. pp. 93–109. Springer (2022)
26. Martin-Brualla, R., Radwan, N., Sajjadi, M.S., Barron, J.T., Dosovitskiy, A., Duckworth, D.: Nerf in the wild: Neural radiance fields for unconstrained photo collections. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 7210–7219 (2021)
27. Mildenhall, B., Srinivasan, P.P., Tancik, M., et al.: Nerf: Representing scenes as neural radiance fields for view synthesis. In: Proceedings of the European Conference on Computer Vision. pp. 405–421 (2020)
28. Ramazzina, A., Bijelic, M., Walz, S., et al.: Scatternerf: Seeing through fog with physically-based inverse neural rendering. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 17957–17968 (2023)
29. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
30. Schönberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4104–4113 (2016)
31. Sun, C., Sun, M., Chen, H.T.: Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5459–5469 (2022)
32. Tancik, M., Casser, V., Yan, X., Pradhan, S., Mildenhall, B., Srinivasan, P.P., Barron, J.T., Kretzschmar, H.: Block-nerf: Scalable large scene neural view synthesis. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8248–8258 (2022)
33. Tang, Y., Zhu, C., Wan, R., Xu, C., Shi, B.: Neural underwater scene representation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 11780–11789 (2024)
34. Targ, S., Almeida, D., Lyman, K.: Resnet in resnet: Generalizing residual architectures. arXiv preprint arXiv:1603.08029 (2016)
35. Turki, H., Ramanan, D., Satyanarayanan, M.: Mega-nerf: Scalable construction of large-scale nerfs for virtual fly-throughs. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12922–12931 (2022)
36. Wang, H., Anantrasirichai, N., Zhang, F., Bull, D.: Uw-gs: Distractor-aware 3d gaussian splatting for enhanced underwater scene reconstruction. In: 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 3280–3289. IEEE (2025)
37. Wang, P., Liu, Y., Chen, Z., Liu, L., Liu, Z., Komura, T., Theobalt, C., Wang, W.: F2-nerf: Fast neural radiance field training with free camera trajectories. In:

Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4150–4159 (2023)
38. Xie, X., Gherardi, R., Pan, Z., Huang, S.: Hollownerf: Pruning hashgrid-based nerfs with trainable collision mitigation. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 3480–3490 (2023)
39. Xie, Y., Kong, L., Chen, K., Zheng, Z., Yu, X., Yu, Z., Zheng, B.: Uveb: A large-scale benchmark and baseline towards real-world underwater video enhancement. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 22358–22367 (2024)
40. Xu, L., Xiangli, Y., Peng, S., Pan, X., Zhao, N., Theobalt, C., Dai, B., Lin, D.: Grid-guided neural radiance fields for large urban scenes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8296–8306 (2023)
41. Yang, D., Leonard, J.J., Girdhar, Y.: Seasplat: Representing underwater scenes with 3d gaussian splatting and a physically grounded image formation model. In: 2025 IEEE International Conference on Robotics and Automation (ICRA). pp. 7632–7638. IEEE (2025)
42. Zhang, F., You, S., Li, Y., Fu, Y.: Atlantis: Enabling underwater depth estimation with stable diffusion. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 11852–11861 (2024)
43. Zhenxing, M., Xu, D.: Switch-nerf: Learning scene decomposition with mixture of experts for large-scale neural radiance fields. In: The Eleventh International Conference on Learning Representations (2022)