RUSplatting: Robust 3D Gaussian Splatting for Sparse-View Underwater Scene Reconstruction

Zhuodong Jiang[1]
davidjiang326@gmail.com
Haoran Wang[1]
yp22378@bristol.ac.uk
Guoxi Huang[1]
guoxi.huang@bristol.ac.uk
Brett Seymour[2]
Brett_Seymour@nps.gov
Nantheera Anantrasirichai[1]
n.anantrasirichai@bristol.ac.uk

[1] Visual Information Laboratory, University of Bristol, Bristol, UK
[2] Submerged Resources Center, National Park Service, Denver, CO, USA

Abstract

Reconstructing high-fidelity underwater scenes remains a challenging task due to light absorption, scattering, and limited visibility inherent in aquatic environments. This paper presents an enhanced Gaussian Splatting-based framework that improves both the visual quality and geometric accuracy of deep underwater rendering. We propose decoupled learning for RGB channels, guided by the physics of underwater attenuation, to enable more accurate colour restoration. To address sparse-view limitations and improve view consistency, we introduce a frame interpolation strategy with a novel adaptive weighting scheme. Additionally, we introduce a new loss function aimed at reducing noise while preserving edges, which is essential for deep-sea content. We also release a newly collected dataset, Submerged3D, captured specifically in deep-sea environments. Experimental results demonstrate that our framework consistently outperforms state-of-the-art methods with PSNR gains up to 1.90dB, delivering superior perceptual quality and robustness, and offering promising directions for marine robotics and underwater visual analytics. The code of RUSplatting is available at https://github.com/theflash987/RUSplatting and the dataset Submerged3D can be downloaded at https://zenodo.org/records/15482420.

1 Introduction

Underwater exploration has a wide range of applications, including marine ecosystem research, underwater archaeology, and geology. These studies heavily rely on effective data acquisition and visualisation. However, underwater exploration is significantly constrained by the challenging environment, including the risks associated with diving, a shortage of professional divers, and the high costs of both equipment and operations. Additionally, certain

viewpoints or data may lack sufficient overlap, making it difficult to produce high-quality 3D reconstructions. Consequently, there is a pressing need for technologies capable of effectively reconstructing complex marine environments using limited data, enabling onshore visualisation and subsequent analysis.

Advancements in Novel View Synthesis (NVS) have introduced new possibilities for reconstructing and visualising environments in 3D space from multi-view image data. These methods allow the generation of unseen viewpoints using limited training images. A notable breakthrough in NVS is Neural Radiance Fields (NeRF) [21], which implicitly models 3D scenes. Despite their impressive performance, NeRFs require substantial training time and are impractical for real-time or interactive applications [2, 28]. Moreover, their representations are inherently difficult to interpret or modify intuitively [28], limiting their flexibility in dynamic environments, such as underwater. In contrast, 3D Gaussian Splatting (3DGS) [15] reconstructs scenes explicitly using Gaussian point sets, offering faster training, higher rendering speeds and refresh rates, as well as more flexible scene manipulation.

Nevertheless, most existing NeRF-based and 3DGS-based methods focus on reconstructing scenes in clear media and therefore struggle to produce high-quality results in hazy and distorted environments like underwater. Due to the physical properties of water, underwater imaging is affected by attenuation and scattering, where the former causes colour casting and the latter results in low contrast [20]. It is therefore essential to account for these properties when modelling the entire scene.

In the face of these challenges, we propose a new framework: **R**obust **U**nderwater scene representation using 3D Gaussian **Splatting** (RUSplatting), which demonstrates greater robustness than previous works, such as UW-GS [27] and WaterSplatting [18]. Previous approaches integrate medium effects via affine transformations of Gaussian attributes but assume dense, temporally coherent input frames, limiting effectiveness under sparse-view conditions. They also do not address distortions common in deep-sea imaging, such as strong noise and severe colour imbalance.

Our proposed RUSplatting framework, shown in Fig. 1, further advances underwater 3D reconstruction by explicitly decoupling the estimation of attenuation and scattering parameters for individual RGB channels instead of treating them uniformly, enabling more accurate and realistic colour restoration. As discussed in [23], the light attenuation underwater has different degrees of changes due to the wavelength. Red light, which has a longer wavelength, is absorbed more rapidly than green and blue light. Furthermore, to mitigate image noise prevalent in deep-sea conditions, we introduce an Edge-Aware Smoothness Loss (ESL) to directly optimise the reconstruction pipeline, dynamically reducing noise while preserving critical geometric edges. Additionally, to address the inherent challenges of sparse-view datasets, we incorporate an Intermediate Frame Interpolation (IFI) mechanism, enriching viewpoint coverage and increasing initialised Gaussian points. However, since interpolated frames may contain artefacts that could deteriorate the learning process, we propose an Adaptive Frame Weighting (AFW) strategy that adjusts the loss based on the predictive uncertainty [13, 14]. Finally, we introduce a new dataset, Submerged3D, capturing detailed objects in deep-sea environments, supporting future research in underwater robotics, archaeology, and marine biology.

In summary, our key contributions are as follows:

- We propose a novel framework, RUSplatting, for 3D representation of deep underwater scenes under sparse-view conditions.
- RUSplatting enables more accurate colour estimations through colour-channel decoupling, which particularly improves clarity in high-turbidity conditions.

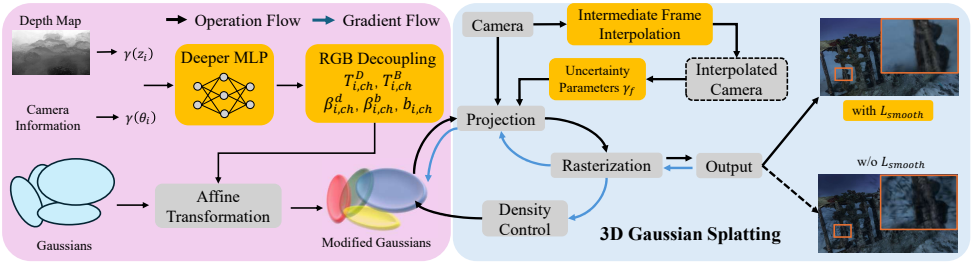

Figure 1: Pipeline of RUSplatting. Yellow highlights indicate the proposed contributions: (1) colour-channel decoupling with a deeper MLP for improved estimation of underwater medium parameters; (2) IFI to enrich representations and enhance frame overlap; (3) an AFW strategy to suppress contributions from poorly interpolated frames; and (4) an ESL to reduce noise while preserving fine structural details.

- We introduce a novel Edge-Aware Smoothness Loss (ESL) to suppress noise commonly encountered in deep-sea, low-light conditions while preserving structural details.
- We introduce Intermediate Frame Interpolation (IFI) and Adaptive Frame Weighting (AFW) to address viewpoint sparsity.
- We present Submerged3D, the first dataset captured in deep underwater environments, highlighting real-world challenges such as limited lighting and severe turbidity.

Extensive experiments on both existing and new datasets demonstrate that RUSplatting consistently outperforms state-of-the-art methods across various evaluation metrics, with PSNR improvements of up to 21.37%, 4.99% and 5.09% compared to SeeThru-NeRF [17], WaterSplatting [18] and UW-GS [27], respectively.

2 Related Work

Underwater images. Processing underwater images remains a major challenge in underwater computer vision due to colour distortion and noise arising from water's physical properties and its impact on light propagation [7]. The Underwater Image Formation Model (UIFM) [3, 4] was introduced to describe this process, accounting for light absorption and backscattering. While the original UIFM assumes a homogeneous water column, the Revised UIFM (R-UIFM) [1] improves realism by allowing spatial variability in water properties, making it widely adopted in underwater image processing. For example, PRWNet [12] uses a deep learning-based multi-stage refinement approach that enhances image quality by addressing spatial and frequency domain degradations through Wavelet Boost Learning. However, the survey in [10] shows that applying enhancement prior to 3D reconstruction results in poorer performance compared to embedding the physical properties of water directly into the reconstruction process.

NVS methods for underwater. While the NeRF-based and GS-based technologies have rapidly advanced, most struggle in scattering media such as fog or water. To tackle this limitation, ScatterNeRF [24] was introduced to effectively separate objects from the scattering medium in foggy environments. SeaThru-NeRF [17] extends this line of research by estimating water medium coefficients using a novel architecture grounded in the Revised Un-

derwater Image Formation Model (R-UIFM) [1]. AquaNeRF [8] improves SeaThru-NeRF with single surface rendering to avoid distractors. Tang *et al.* [26] propose a neural representation that incorporates dynamic rendering and tone mapping to enable more accurate reconstruction of underwater scenes under varying illumination conditions.

For 3DGS-based methods, UW-GS [27] and WaterSplatting [18] employ MLP-based colour appearance modules to address colour casting. In contrast, SeaSplat [29] optimises backscatter parameters using a dark channel prior loss. UW-GS [27], UDR-GS [5], and Aquatic-GS [20] leverage DepthAnything [30] to estimate depth from 2D inputs and incorporate it into their loss functions. UW-GS [27] further improves reconstruction of distant objects through a physics-driven density control module.

3 Preliminaries

The 3DGS [15] employs a collection of spatial Gaussian cloud points to represent a 3D scene. To synthesise a novel view, these Gaussians are projected onto a 2D image plane. The projected Gaussian is defined as

$$G_i(\boldsymbol{x}) = \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^\top \Sigma_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i)\right), \tag{1}$$

where the center of Gaussian G_i is located at $\boldsymbol{\mu}_i$, with covariance matrix Σ_i. Both $\boldsymbol{\mu}_i$ and $\boldsymbol{x}$ lie on the 2D image plane. Each Gaussian i is further characterised by an opacity α_i and a view-dependent colour appearance c_i, parameterised using spherical harmonics (SH). The final pixel colour C is computed via α-blending as follows:

$$C = \sum_{i=1}^{N} \alpha_i c_i(\mathbf{v}) \prod_{j=1}^{i-1}(1-\alpha_j), \quad \text{where } c_i(\mathbf{v}) = \sum_{l=0}^{L}\sum_{m=-l}^{l} c_{i,lm}\cdot Y_{lm}(\mathbf{v}). \tag{2}$$

Here, $c_{i,lm}$ are the learned SH coefficients for each RGB channel, and $Y_{lm}(\mathbf{v})$ are the real SH basis functions in the viewing direction $\mathbf{v}$, where l and m denote the degree and order, respectively. The α-blending operation accounts for occlusions among Gaussians, resulting in a more accurate and detailed final colour composition.

Underwater image formation. Compared to clear, onshore scenes, where scattering and absorption caused by the medium (i.e., air) can often be neglected, underwater scenes are heavily affected by these parameters, since they significantly degrade reconstruction quality [26]. Therefore, it is crucial to consider these medium parameters during reconstruction. As modelled in [1], the intensity of the underwater image is defined as follows:

$$I_c = J\cdot T^D + B^\infty \cdot (1-T^B), \tag{3}$$

where I_c denotes the colour captured by the camera, and J represents the true object colour, free from any medium-induced effects. B^∞ refers to the background light from an infinite distance. T^D and T^B are transmission terms representing the attenuation of direct and backscattered light, respectively. The T^D and T^B are defined as,

$$T^D = \exp(-\beta^d \cdot z) \quad \text{and} \quad T^B = \exp(-\beta^b \cdot z), \tag{4}$$

Here β^d and β^b are two colour medium coefficients, and z is the line-of-sight distance from the camera to the 3D scene point rather than the water depth.

4 Methodology

The detailed structure of the proposed framework is presented in Fig. 1. The model estimates underwater medium parameters using an MLP with encoded depth and viewpoint position information as inputs. The MLP $f_\theta(\gamma_z(z), \gamma_\theta(\mathbf{v}))$ predicts, for each Gaussian and channel, the parameters $\{T^D, T^B, \beta^d, \beta^b, b\}$, which are jointly optimised under the main loss function described in Section. 4.4. Accordingly, the colours of the spatial 3D Gaussians are modified through the affine transformation. Then, the modified Gaussians, including additional Gaussians introduced by the frame interpolation process with learnable uncertainty weights, are projected onto the 2D plane. During training, an edge-aware smoothness loss is integrated into the model as a part of the loss function to remove noise from underwater images, and in the meantime maintaining geometric details.

4.1 Colour-Channel Decoupling

As discussed in numerous underwater studies [10, 19, 20], the attenuation rates differ across red, green, and blue light channels. Therefore, the medium parameters for the RGB channels cannot be treated identically. To overcome this and achieve more accurate colour reconstruction, we decouple the channel-wise learning of the underwater medium. Consequently, the attenuation (T^D) and backscatter (T^B) factors for each colour channel $ch \in \{r, g, b\}$ are defined as:

$$T^D_{ch} = \exp(-\beta^d_{ch} \cdot z), \quad T^B_{ch} = \exp(-\beta^b_{ch} \cdot z). \tag{5}$$

Previous works use a 2-layer MLP [16] to estimate the complex underwater medium parameters, however, it may fall short in capturing such complicated patterns as the study indicated in [9]. Thus, we employ a deeper MLP with 5 layers, selected based on a hyperparameter sensitivity analysis (see supplementary material for details). The final corrected colour of Gaussian i, $c^m_{i,ch}(v)$, is obtained through an affine transformation, as presented below, where $c_{i,ch}(v)$ is the observed view-dependent colour.

$$c^m_{i,ch}(v) = T^D_{i,ch} \cdot c_{i,ch}(v) + (1 - T^B_{i,ch}) \cdot b_{i,ch}. \tag{6}$$

Here, $b_{i,ch}$ denotes a bias term specific to each Gaussian and channel, modelling the background light. It is assumed view-independent within the views observing Gaussian i, while view-dependent effects are captured by $c_{i,ch}(v)$. As Eq. 6 follows directly from Eq. 3, the observed channel-wise colour equals a scaled direct term plus an additive background light term, hence the mapping is affine.

4.2 Intermediate Frame Interpolation (IFI)

To address the limitations introduced by sparse views, we adopt a frame interpolation technique to synthesise intermediate frames between adjacent input images, leveraging them in two key aspects: i) enhancing the initial point cloud and ii) improving the RUSplatting optimisation, where we propose Adaptive Frame Weighting (AFW). Although various methods are available and should not significantly affect the final 3DGS results, we use Real-time Intermediate Flow Estimation (RIFE) [11], a state-of-the-art deep learning-based approach that employs bidirectional optical flow and knowledge distillation to enable real-time processing.

Initial point cloud. The initial point cloud (obtained using COLMAP [25]) acts as a rough geometric framework for the final model. Its quality and density directly influence the accuracy of the rendering. By incorporating interpolated frames and re-running COLMAP on the augmented dataset, additional initialisation points are introduced, enabling a denser distribution of Gaussians and improving the overall reconstruction precision and detail. Without IFI, the model accesses an average of 17,635 initialisation points across all datasets. After adopting IFI, this number increases significantly by 55.7% to an average of 27,461 points. Detailed results for each dataset are provided in the supplementary material.

Adaptive Frame Weighting (AFW) γ. In sparse view scenarios, adjacent frames are significantly spaced apart, making frame interpolation challenging. The large spatial shifts between frames may introduce artefacts in certain geometric structures of the interpolated frames, which may degrade the quality of 3D reconstruction during optimisation when these frames are used. Inspired by [13, 14], we propose an AFW mechanism, introducing an uncertainty learnable parameter γ for each interpolated frame f that is optimised along with the training. Accordingly, the loss function for the interpolated frame L_f is updated as follows.

$$L'_f = \frac{1}{2} \cdot \gamma_f \cdot L_f - \frac{1}{2} \cdot \alpha \cdot \log(\gamma_f), \tag{7}$$

where α controls the relative weight of the uncertainty regularisation term. α is a hyperparameter selected within the range $[0,1]$. Adapting such weights ensures that interpolated frames aid the training while alleviating the effect caused by the potential artefacts.

4.3 Edge-Aware Smoothness Loss (ESL)

Inspired by bilateral filtering, which smooths noisy images while preserving edges, we propose a novel loss function that behaves dually, removing noise and preserving geometric edges. A key advantage of our approach is that the functionality is embedded into the end-to-end training framework without relying on multiple predefined parameters like traditional bilateral filters. Instead, we utilise the underlying 3D geometry to guide edge-preserving smoothness in the rendered images.

Our edge-aware smoothness loss is defined as,

$$L_{\text{Smooth}}(I,D) = \sum_{i,j} \left(w^x_{i,j} \cdot \left| I_{i,j} - I_{i,j+1} \right| + w^y_{i,j} \cdot \left| I_{i,j} - I_{i+1,j} \right| \right), \tag{8}$$

where (i,j) indexes the spatial coordinates in the image plane, I denotes the input image and the D denotes the pseudo depth map generated from the Depth-Anything-V2 [31]. The edge-aware weights $w^x_{i,j}$ and $w^y_{i,j}$, corresponding to the horizontal and vertical directions, respectively, are computed using depth discontinuities,

$$w^x_{i,j} = \exp\left(-\lambda_b \cdot \left| D_{i,j} - D_{i,j+1} \right|\right), \quad w^y_{i,j} = \exp\left(-\lambda_b \cdot \left| D_{i,j} - D_{i+1,j} \right|\right). \tag{9}$$

Here, λ_b controls the edge sensitivity and is selected via grid search in the range $[0,5]$.

4.4 Loss Function

The entire model is trained end-to-end using a combination of four loss terms, including three common losses (described below) and our proposed edge-aware smoothness loss, L_{Smooth}

(a) Cormoran (b) Isro (c) Kwaj (d) Tokai

Figure 2: Sample images from Submerged3D dataset.

(described in Section 4.3), with a weight λ_s of 0.2. Additionally, we integrate the adaptive frame weighting for interpolated frames as described in Section 4.2. Consequently, the final loss function is in the form of,

$$L_{final} = L_{\text{Rec}} + L_{\text{Depth}} + L_{\text{g}} + \lambda_s L_{\text{Smooth}}, \tag{10}$$

for the original frame, while for the interpolated frame, the loss function turns to,

$$L'_{final} = \frac{1}{2} \cdot \gamma \cdot L_{final} - \frac{1}{2} \cdot \alpha \cdot \log(\gamma). \tag{11}$$

Reconstruction Loss L_{Rec}. To evaluate the differences between the reconstructed image and the ground truth image, the L_{Rec} is defined as $L_{\text{Rec}} = \lambda_r \cdot \omega \cdot L_1 + (1 - \lambda_r) \cdot \omega \cdot L_{\text{SSIM}}$, where the λ_r is a loss weight to balance the L_1 and L_{D-SSIM}. ω denotes the Binary Motion Mask (BMM) [27] that remove the distractor and marine snow artefacts. L_1 refers to the absolute error loss and L_{SSIM} is the SSIM loss.

Depth-supervised Loss L_{Depth}. The depth information plays a vital role as the depth z information is involved in the estimation of the underwater medium parameters, as illustrated in Eq. 5. Thus, the L_{Depth} is included and defined as, $L_{\text{Depth}} = \lambda_d \cdot \|\hat{D} - D\| + \lambda_{ca} \cdot (\sum_{ch \in \{r,g,b\}} \sum_{n \in \{d,b\}} \|z^n_{ch} - D\|)$, where λ_d and λ_{ca} are loss weights, D is the pseudo depth map, obtained from Depth-Anything-V2 [31], and $\hat{D}$ is the rendered depth map from the RUSplatting. z^n_{ch} denotes the inferred depth under direct (d) and backscatter (b) lighting conditions for each colour channel ch.

Grey-world Assumption Loss L_g. Following [6], L_g is built based on the assumption that in a natural scene, the average value of each colour channel of the image should be equal. With this restriction, it can promote the colour balance and aid in the accurate estimation of the medium parameters. $L_g = \sum_{ch \in \{r,g,b\}} \|\mu(J_{ch}) - 0.5\|_2^2$. Here, J_{ch} denotes the restored pixel values for all three colour channels, ranging between 0 to 1. $\mu(\cdot)$ denotes the mean value of a channel. It is worth noting that the 0.5 here corresponds to the ideal mean intensity for the colour channel under the grey-world assumption.

5 Experiments

5.1 Implementation details

Datasets. Three datasets were used: SeaThru-NeRF [17], S-UW [27] and our newly collected dataset, Submerged3D. Submerged3D consists of four scenes, each containing 20

Dataset	SeaThru-NeRF			S-UW			Submerged3D		
Method	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
Instant-NGP [22]	23.7944	0.6317	0.4434	18.9001	0.4693	0.3966	19.0266	0.5107	0.4977
SeaThru-NeRF [17]	26.0604	0.7806	0.3398	21.2253	0.5948	0.4104	19.1967	0.4967	0.5685
3DGS [15]	26.8353	0.9011	0.1826	23.7563	0.7745	0.2231	22.6551	0.7544	0.3288
WaterSplatting [18]	27.3924	0.8475	0.1870	24.8788	0.7915	0.2159	24.5832	0.7517	0.3172
UW-GS [27]	28.2557	0.9105	0.1750	24.1833	0.7838	0.2050	24.3409	0.7723	0.3107
RUSplatting (ours)	**29.2928**	**0.9163**	**0.1665**	**25.5955**	**0.8138**	**0.1964**	**25.7990**	**0.7724**	**0.3008**

Table 1: Performance comparison between RUSplatting and five baseline models on three datasets. The results are the average values over all four scenes in each dataset. ↑ indicates that higher values are better, while ↓ indicates that lower values are better. Red colour with bold denotes the best result and yellow denotes the second best.

RGB 720p images focusing on shipwrecks captured in the deep sea. It is challenging for 3D reconstruction due to sparse viewpoints and low-light environment conditions. Examples of the scenes in Submerged3D are shown in Fig. 2. To create the training and test splits, we assign every 8th image in the dataset to the test set and use the rest for training. SeaThru-NeRF, S-UW, and our Submerged3D provide no ground-truth depth, therefore we use Depth-Anything-V2 [31] pseudo-depth as weak supervision.

Experimental Setup. We utilised COLMAP [25] to initialise the point cloud and estimate the camera poses for the image sequences. For each pair of adjacent images, only one interpolated frame is generated. The experiments are conducted on a single RTX 4090 GPU. All experiments are run for 20,000 iterations. λ_r, λ_d, and λ_{ca} are set to 0.8, 0.1, and 1, respectively.

5.2 Evaluation

We compared our results with the Instant-NGP [22], SeaThru-NeRF [17], 3DGS [15], WaterSplatting [18] and UW-GS [27]. For a fair comparison, we used their official implementations and trained each method on the same image sequences using our dataset split strategy. The quality of the synthesised views was evaluated using PSNR, SSIM, and LPIPS metrics.

Quantitative Comparisons. Tab. 1 presents comparisons with prior state-of-the-art methods, reporting the average rendering quality across all scenes in each dataset. We can observe that our RUSplatting consistently outperforms all baseline methods over three datasets across all metrics. Specifically, RUSplatting achieves PSNR improvements of 2.83, 3.01, and 3.84 dB over baseline methods on SeaThru-NeRF, S-UW and Submerged3D, respectively. It also yields SSIM gains of 12.53%, 19.19%, and 17.53%, and reduces LPIPS by 37.29%, 32.32%, and 25.66% on the same datasets.

Qualitative Comparisons. The visual results shown in Fig. 3 further demonstrate the superior performance of our method. Specifically, in the first row, all other models fail to produce a clear and accurate image and contain artefacts, as illustrated by the yellow and blue-outlined regions, whereas our method preserves fine details. Additionally, RUSplatting accurately reconstructs the gully structure in the second row, highlighted in red, which the

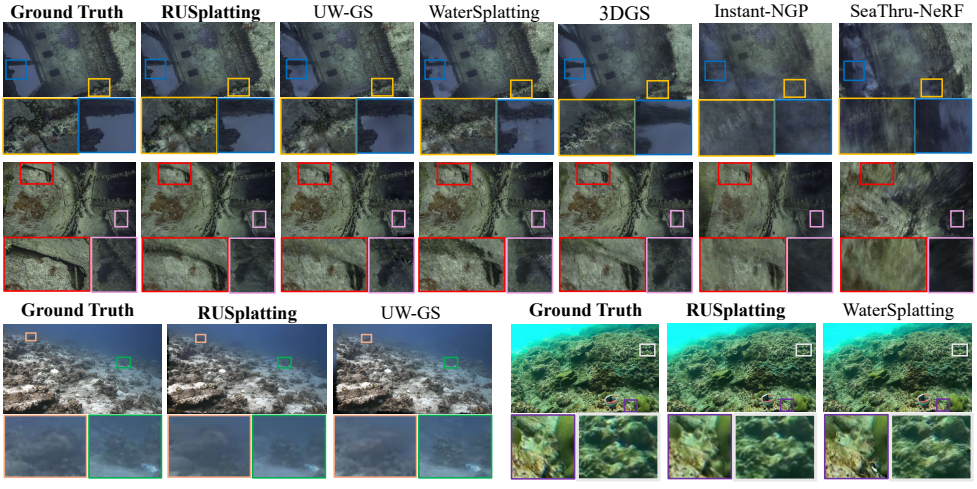

Figure 3: Novel view rendering comparison on the Submerged3D, SeaThru-NeRF and S-UW datasets. The first two rows show results from the Cormoran, and Tokai scenes from Submerged3D, respectively. The left side of the third row shows the result from IUI-Reasea (SeaThru-NeRF), while the right side shows the Seabed from S-UW

other models fail to capture. Moreover, according to the third row, only our model renders high-quality results for both datasets. In contrast, the rendered images by UW-GS contain blurry artefacts (as shown in the green outlined region) and fail to capture the geometric structure in distant as displayed in the orange outlined area. Moreover, the outputs from WaterSplatting are noisy, as highlighted in the purple and grey outlined regions.

5.3 Ablation Study

To evaluate the impact of each component, we conducted ablation studies, with configurations and results presented in Fig. 4. According to the right subfigure in Fig. 4, the following conclusions can be drawn.

Deeper MLP. The comparison between M1 and M5 demonstrates the effectiveness of a deeper architecture, as the SSIM increases significantly from 0.7192 to 0.7491, and PSNR rises notably from 24.68 to 25.03.

RGB Decoupling. We observe a substantial gain in SSIM (from 0.7284 to 0.7491) and PSNR (from 24.44 to 25.03) through the evaluation between M2 and M5. This indicates that RGB decoupling positively contributes to the overall performance.

ESL. Comparing M3 and M5, there is a notable improvement in SSIM from 0.7306 to 0.7491, and a substantial increase in PSNR from 24.67 to 25.03. This clearly highlights ESL's essential role in preserving structural fidelity.

Modules	M1	M2	M3	M4	M5	RUSplatting
Deeper MLP	✗	✓	✓	✓	✓	✓
RGB Decoupling	✓	✗	✓	✓	✓	✓
ESL	✓	✓	✗	✓	✓	✓
IFI w/o AFW	✓	✓	✓	✗	✓	✓
IFI w/ AFW	✗	✗	✗	✗	✗	✓

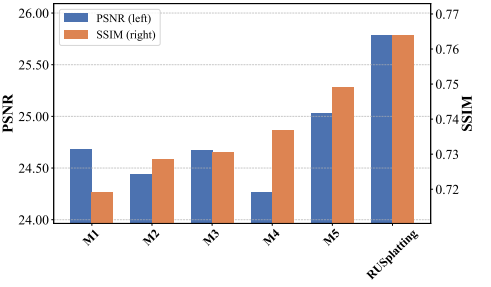

Figure 4: Left: Ablation study configurations. Right: Average results of all scenes. We construct five model variants (i.e., M1–M5) and compare them to RUSplatting.

IFI. A 1.84 dB decrease from M5 to M4 regarding the PSNR, significantly evidences the importance of the IFI in terms of achieving high-quality pixel-wise reconstruction. Also, the SSIM drops from 0.7491 to 0.7368, indicating that the IFI is also critical in capturing sophisticated structure. It facilitates the model in obtaining more initialisation points and alleviates the impact of sparse perspectives.

AFW. Employing AFW (the full model, RUSplatting) results in the highest performance, achieving an SSIM of 0.7640 and a PSNR of 25.7874 on average across all datasets. By adjusting the uncertainty parameters of each interpolated frame, the model can capture more scene information for training by increasing the weight of high-quality interpolated frames, while reducing the weight of low-quality interpolated frames to reduce their potential impact on training. The improvement demonstrates the value of such adaptive weighting features for optimal image reconstruction quality.

6 Conclusion

This paper presents a new GS-based framework for deep underwater scene reconstruction. Specifically, we propose decoupled learning for the RGB channels, leveraging wavelength-dependent attenuation and scattering properties of underwater light. This approach enables more accurate colour reconstruction. To mitigate challenges posed by sparse viewpoints, we incorporate a frame interpolation mechanism that facilitates denser and more robust scene capture, along with adaptive frame weighting during training. Finally, we propose a new loss function that dynamically adjusts to image characteristics, effectively suppressing noise while preserving geometric structure. Collectively, our RUSplatting framework significantly enhances the visual fidelity, structural consistency, and colour accuracy of underwater reconstructions. However, as RIFE and Depth-Anything-V2 are not specifically designed for underwater conditions, future work could improve their robustness through fine-tuning or physics-based priors, thereby further enhancing the colour fidelity and geometric accuracy of our framework.

Acknowledgements

This work was supported by the UKRI MyWorld Strength in Places Programme(SIPF00006/1) and EPSRC ECR(EP/Y002490/1).

References

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

[2] Guikun Chen and Wenguan Wang. A survey on 3d gaussian splatting. *arXiv preprint arXiv:2401.03890*, 2024.

[3] John Y Chiang and Ying-Ching Chen. Underwater image enhancement by wavelength compensation and dehazing. *IEEE transactions on image processing*, 21(4):1756–1769, 2011.

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

[5] Y. Du, Z. Zhang, P. Zhang, F. Sun, and X. Lv. UDR-GS: Enhancing Underwater Dynamic Scene Reconstruction with Depth Regularization. *Symmetry*, 16(8):1010, 2024. doi: 10.3390/sym16081010.

[6] Zhenqi Fu, Huangxing Lin, Yan Yang, Shu Chai, Liyan Sun, Yue Huang, and Xinghao Ding. Unsupervised underwater image restoration: From a homology perspective. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pages 643–651, 2022.

[7] Salma P González-Sabbagh and Antonio Robles-Kelly. A survey on underwater computer vision. *ACM Computing Surveys*, 55(13s):1–39, 2023.

[8] Luca Gough, Adrian Azzarelli, Fan Zhang, and Nantheera Anantrasirichai. AquaNeRF: Neural Radiance Fields in Underwater Media with Distractor Removal. In *Proceedings of the 2025 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2025.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[10] Guoxi Huang, Haoran Wang, Brett Seymour, Evan Kovacs, John Ellerbrock, Dave Blackham, and Nantheera Anantrasirichai. Visual enhancement and 3d representation for underwater scenes: A review. *arXiv preprint arXiv:2505.01869*, 2024.

[11] Zhewei Huang, Tianyuan Zhang, Wen Heng, Boxin Shi, and Shuchang Zhou. Real-time intermediate flow estimation for video frame interpolation. In *European Conference on Computer Vision*, pages 624–642. Springer, 2022.

[12] Fushuo Huo, Bingheng Li, and Xuegui Zhu. Efficient wavelet boost learning-based multi-stage progressive refinement network for underwater image enhancement. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1944–1952, 2021.

[13] Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? *Advances in neural information processing systems*, 30, 2017.

[14] Alex Kendall, Yarin Gal, and Roberto Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7482–7491, 2018.

[15] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4): 139–1, 2023.

[16] Joo Chan Lee, Daniel Rho, Xiangyu Sun, Jong Hwan Ko, and Eunbyung Park. Compact 3d gaussian representation for radiance field. In *the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21719–21728, 2024.

[17] Deborah Levy, Amit Peleg, Naama Pearl, Dan Rosenbaum, Derya Akkaynak, Simon Korman, and Tali Treibitz. Seathru-nerf: Neural radiance fields in scattering media. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 56–65, 2023.

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

[19] Yini Li and Nantheera Anantrasirichai. Zero-TIG: Temporal consistency-aware zero-shot illumination-guided low-light video enhancement, 2025. URL https://arxiv.org/abs/2503.11175.

[20] Shaohua Liu, Junzhe Lu, Zuoya Gu, Jiajun Li, and Yue Deng. Aquatic-gs: A hybrid 3d representation for underwater scenes. *arXiv preprint arXiv:2411.00239*, 2024.

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

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

[23] Yan-Tsung Peng and Pamela C Cosman. Underwater image restoration based on image blurriness and light absorption. *IEEE transactions on image processing*, 26(4):1579–1594, 2017.

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

[25] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

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

[27] Haoran Wang, Nantheera Anantrasirichai, Fan Zhang, and David Bull. UW-GS: Distractor-aware 3D gaussian splatting for enhanced underwater scene reconstruction. In *Proceedings of the Winter Conference on Applications of Computer Vision (WACV)*, pages 3280–3289, February 2025.

[28] Tong Wu, Yu-Jie Yuan, Ling-Xiao Zhang, Jie Yang, Yan-Pei Cao, Ling-Qi Yan, and Lin Gao. Recent advances in 3d gaussian splatting. *Computational Visual Media*, 10 (4):613–642, 2024.

[29] Daniel Yang, John J. Leonard, and Yogesh Girdhar. SeaSplat: Representing underwater scenes with 3d gaussian splatting and a physically grounded image formation model. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2025.

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

[31] Lihe Yang, Bingyi Kang, Zilong Huang, Zhen Zhao, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth anything v2. *Advances in Neural Information Processing Systems*, 37:21875–21911, 2024.