2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

GS-UVCE: Gaussian Splatting-Driven Unsupervised Visual Consistency Enhancement for Underwater 3D Scene Reconstruction

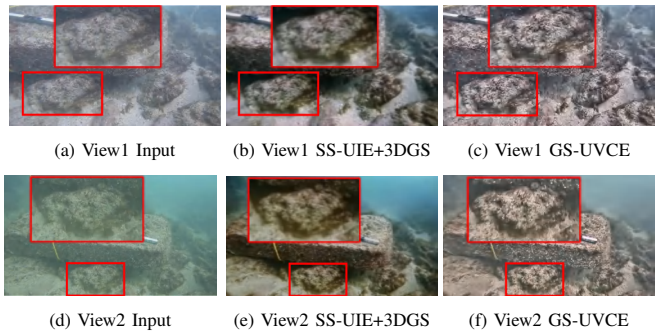
Xiang Li[1], Student Member, IEEE, Chi Li[2], Yiming Xu[3] and Yan Zhuang[4*], Senior Member, IEEE

***Abstract*— Underwater 3D scene reconstruction is critical for the operation of underwater robotics, yet remains highly challenging due to the semi-transparent water medium, which introduces optical distortions, light scattering, and severe visibility degradation. Therefore, effective underwater image enhancement is a prerequisite for reliable reconstruction. However, existing approaches typically enhance individual views with pre-trained models before reconstruction, leading to poor generalization and inconsistent multi-view results. To address these limitations, we propose GS-UVCE, an end-to-end framework for Gaussian Splatting-driven Unsupervised Visual Consistency Enhancement. GS-UVCE incorporates a Medium-MLP to model water-medium effects and a Light-MLP to adaptively correct illumination, ensuring illumination consistency. Furthermore, depth regularization is introduced to preserve geometric consistency under varying scene conditions. Extensive experiments on multiple underwater datasets show that GS-UVCE consistently outperforms SOTA methods, achieving superior reconstruction fidelity and visual consistency enhancement. Code: https://github.com/CharlyX-Lee/GS-UVCE**

I. INTRODUCTION

In underwater environments, high-precision 3D reconstruction is crucial for autonomous underwater vehicle (AUV) tasks such as navigation, localization, and marine archaeology. Recently, 3D Gaussian Splatting (3DGS) [1] has demonstrated remarkable performance in view synthesis and geometric reconstruction in terrestrial scenes. Its explicit, parallelizable, and efficient rendering paradigm makes it well-suited for real-time 3D reconstruction on resource-constrained platforms. These characteristics highlight its promise for underwater applications, where robust and accurate 3D reconstruction remains an open challenge.

However, the 3DGS method cannot be directly applied underwater due to the semi-transparency, anisotropic scattering, and background light of the water medium, which destabilize the representation of primitive color and opacity and hinder reliable reconstruction of distant structures. Underwater imaging further suffers from absorption and scattering, causing color cast, low light, and blur that degrade reconstruction accuracy. These issues necessitate underwater image enhancement to mitigate medium effects and improve visual discernibility. As shown in Fig. 1, directly combining existing single-frame enhancement methods with 3DGS often results in difficulties in representing the water medium and inconsistent multi-view lighting enhancement. Moreover, most enhancement methods depend on supervised learning, but ground truth is difficult to obtain underwater, limiting their generalizability and practicality.



(a) View1 Input (b) View1 SS-UIE+3DGS (c) View1 GS-UVCE
(d) View2 Input (e) View2 SS-UIE+3DGS (f) View2 GS-UVCE

Fig. 1: A comparison between GS-UVCE and the SOTA underwater enhancement model SS-UIE [2] with the baseline 3DGS, using multi-view image (a) and (d) inputs, shows that images (b) and (e) exhibit light inconsistencies after SS-UIE processing, while images (c) and (f) underscore the robust light consistency maintained by GS-UVCE, as emphasized in the View Compare image through detailed magnification.

To address the above challenges, we propose GS-UVCE, an end-to-end underwater 3D scene reconstruction framework that utilizes Gaussian Splatting-driven Unsupervised Visual Consistency Enhancement. Specifically, our method integrates medium-aware modeling and illumination correction to jointly handle water-medium effects and viewpoint-dependent lighting, while depth regularization further mitigates distance-dependent attenuation. This design enables multi-view, illumination, and geometry consistency enhancement across perspectives and structurally accurate 3D reconstruction in degraded underwater environments. The contributions of this paper are summarized as follows:

- We propose GS-UVCE, the first end-to-end framework that unifies unsupervised underwater visual consistency enhancement with 3DGS-based reconstruction, enabling high-fidelity underwater 3D scene reconstruction with multi-view consistency.
- We design two novel modules: Medium-MLP, which models the water medium, and Light-MLP, which adaptively corrects illumination for illumination consistency.

This work was supported in part by the National Natural Science Foundation of China under Grant U22B2041, and the Natural Science Foundation of Liaoning Province, China (No. 2024-BSLH-004).

[1]Xiang Li is with the School of Computer Science and Technology. Dalian University of Technology, Dalian, 116024, China (e-mail: 15140585150@mail.dlut.edu.cn)

[2]Chi Li is with the the School of Information Engineering, Dalian University, Dalian, 116622, China (e-mail: lichi@dlu.edu.cn)

[3]Yiming Xu is with the International School of Information Science Engineering, Dalian University of Technology, Dalian, 116600, China (e-mail: 127991150@mail.dlut.edu.cn)

[4]Yan Zhuang is with the School of Control Science and Engineering. Dalian University of Technology, Dalian, 116024, China (e-mail: zhuang@dlut.edu.cn) (*Corresponding author)

Combined with depth regularization, the framework effectively mitigates attenuation and scattering, ensuring geometry consistency enhancement.

- We validate GS-UVCE through extensive experiments and ablation studies on four public underwater datasets, demonstrating its superior performance over SOTA methods in both high-fidelity 3D reconstruction and underwater image enhancement.

II. RELATED WORK

A. Underwater Image Enhancement

Underwater image enhancement aims to correct the color cast and low light caused by water, thereby recovering the true appearance of underwater scenes, and improving the reliability of subsequent 3D reconstruction. Existing underwater image enhancement methods can be classified into learning-based and physics-based methods.

Learning-based methods leverage deep learning to model the degradation process of underwater images. U-shape [3] and DM-Water [4] employ supervised learning with synthetic data to enhance image quality by learning the underwater degradation patterns, aiming to improve the restoration of underwater features. SS-UIE [2] introduces a semi-supervised framework that combines labeled and unlabeled data to improve the generalization ability of underwater enhancement models. However, obtaining underwater ground truth is challenging, and the performance of learning-based methods varies significantly across different environments, which limits their practicality in real-world underwater scenes.

Physics-based methods model the underwater imaging process by leveraging the physical properties of light, eliminating the need for ground truth data. Berman et al. [5] propose a method for underwater single-image color restoration using haze-lines, which can effectively handle light absorption and scattering. Sucre [6] enhances underwater images by utilizing 3D scene information, offering more robust restoration. SeaThru [7] models the water medium's optical properties to restore underwater images, offering improved color correction and depth restoration in underwater scenes. This approach has been widely adopted, as it demonstrates strong robustness in real-world underwater applications.

B. Underwater 3D Scene Reconstruction

Underwater 3D scene reconstruction aims to generate 3D models of underwater scenes by recovering spatial structures and geometries from images or sensor data. Recent methods have incorporated novel 3D representations, such as NeRF [8] and 3DGS [1], to improve the robustness and quality of underwater 3D scene reconstruction.

WaterNeRF [9] introduces an underwater neural renderer that estimates the medium parameters solely based on histogram-equalized images as input to the network, decoupling it from the rendering process. SeaThruNeRF [10] extends NeRF by incorporating SeaThru's underwater imaging theory, integrating the underwater imaging model into a Multilayer Perceptron (MLP). However, the implicit representation of NeRF leads to substantial computational inefficiency, rendering the reconstruction process too slow for real-time underwater detection.

3DGS [1] overcomes the limitations of NeRF-based methods by providing a more flexible and adaptive representation of 3D objects. However, its explicit representation restricts its ability to model translucent media, such as water. Water Splatting [11] enhances performance using a new 3D Gaussian rendering model. SeaSplat [12] learns medium parameters through the DeepSeeColor [13] method and optimizes rendering with a specific loss function. Despite these advancements, these systems struggle with common underwater challenges, such as uneven lighting, color cast, and low light conditions, highlighting the need for methods capable of robust, high-quality underwater 3D scene reconstruction.

III. METHODOLOGY

The system overview is shown in Fig. 2. The system jointly performs high-fidelity 3D reconstruction and unsupervised underwater visual consistency enhancement within a Gaussian Splatting framework. The Medium-MLP learns the parameters of the underwater imaging model, which constrains the rendering of the medium, allowing the system to depict both the medium and the target object accurately. Additionally, the Light-MLP is proposed to drive the unsupervised enhancement by estimating viewpoint-dependent illumination, modeling anisotropic light fields, and ensuring consistent visual appearance across views. This design supports both mixed and separate medium-object rendering in degraded underwater environments. Optimization is guided by a loss function that accounts for the scattering medium, while depth regularization using an initial monocular depth map preserves high-fidelity local reconstruction.

A. Underwater Imaging Model

Underwater imaging presents distinct challenges, primarily due to the optical properties of water, which significantly degrade image quality. The Jaffe-McGlamery underwater imaging model [14] is a well-regarded framework that effectively characterizes this degradation. The model decomposes underwater imaging into direct transmission, forward scatter, and backscatter components. However, forward scatter is usually ignored when the camera acquires images of the underwater scene. According to a revised underwater image formation model [15] and SeaThru [7], the attenuation coefficient is extended to the attenuation coefficient σ^{attn} and the backscatter coefficient σ^{bs} of the water medium, respectively. Consequently, an underwater degraded image $I(x)$ captured by the camera can be represented as:

$$I(x) = J(x)e^{-\sigma^{attn}D(x)} + B_c(1 - e^{-\sigma^{bs}D(x)}) \quad (1)$$

where $J(x)$ represents the image of objects that the camera would capture in the absence of the water medium, x represents the image pixel, $D(x)$ represents the distance between the objects and the camera corresponding to each pixel in the captured image $I(x)$, and B_c is the background light which can be considered as the backscatter color of water at infinity.

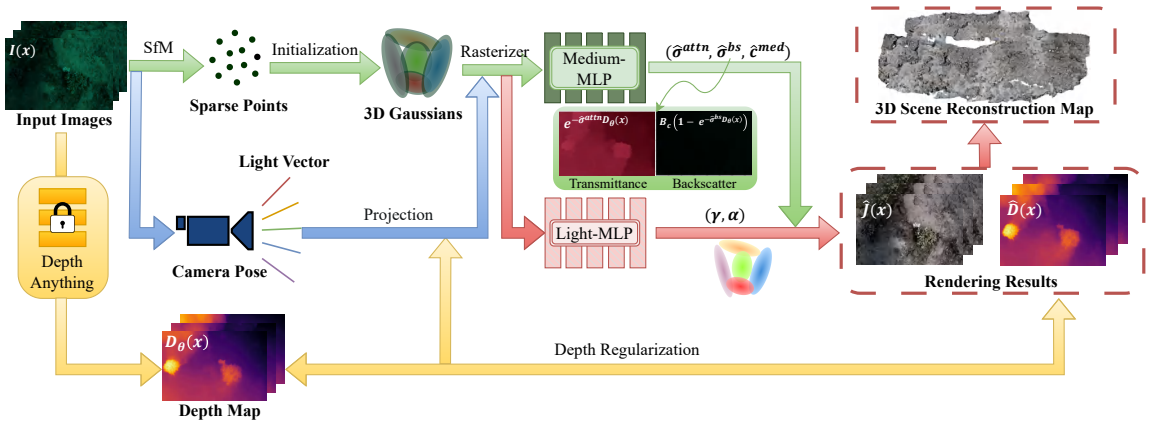


Fig. 2: The overview of GS-UVCE. We utilize an offline system to acquire a series of sparse point clouds, keyframes, and camera poses for initialization purposes. The relevant attributes of these 3D Gaussians are encoded and input into Medium-MLP and Light-MLP to determine the medium and light parameters. A differentiable rasterization technique facilitates end-to-end optimization by aligning rendered outputs with ground truth data. During scene optimization, the depth map predicted by Depth Anything supervises the rendered depth, and the depths of adjacent frames are smoothed. By comparing the input images and rendering results, our approach demonstrates clear improvements in degraded underwater environments.

B. GS-UVCE

1) Rendering with Medium-MLP: We propose a method to enhance underwater image rendering by incorporating the effects of the water medium, such as light attenuation and backscattering. Specifically, a rendering model with Medium-MLP is introduced, which accounts for both object and medium components in underwater scenes, utilizing a series of unstructured, explicit 3D Gaussian primitives for efficient GPU-based computation.

The core of our method is based on a point-based α-blending technique that calculates pixel colors by blending N overlapping Gaussian volumes. The blending is defined as follows:

$$C(r) = \sum_{i \in N} c_i \alpha_i T_i, \quad \text{where} \quad T_i = \prod_{j=1}^{i-1} (1 - \alpha_j) \tag{2}$$

where c_i is the base color of each Gaussian, r is the ray, and α_i is related to the explicit features of the Gaussian volume, which can be obtained from the 2D projection of the Gaussian:

$$\alpha_i = O_i \exp\left(-\frac{1}{2} (p - \hat{\mu}_i)^{\mathsf{T}} \hat{\Sigma}_i^{-1} (p - \hat{\mu}_i) \right) \tag{3}$$

where p is the pixel position, $\hat{\mu}_i$ represents the mean center of the 2D Gaussian distribution, $\hat{\Sigma}_i$ represents the covariance matrix of the 2D Gaussian distribution, and the initial opacity is $O_i \in [0, 1]$.

For underwater scenarios, the final volume rendering equation, incorporating the object and medium components, with:

$$\hat{C}(r) = \sum_{i=1}^{N} \hat{C}_i^{\text{obj}}(r) + \sum_{i=1}^{N} \hat{C}_i^{\text{med}}(r) \tag{4}$$

We adopt principles from Eq. 1 to separate the medium scattering effects, defining the medium attenuation σ_{attn} and backscattering σ_{bs} coefficients:

$$\hat{C}_i^{\text{obj}}(r) = T_i^{\text{obj}} \, e^{-\hat{\sigma}^{\text{attn}} \hat{D}_i} \alpha_i c^{\text{obj}}(r) \tag{5}$$

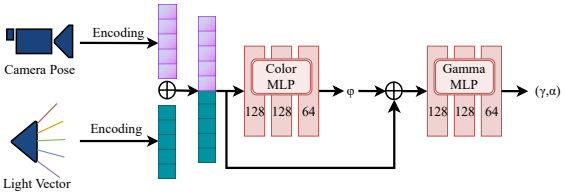


Fig. 3: The detailed architecture of the Light-MLP. Numerical labels within each layer represent the output vector dimensions. Sigmoid activation is utilized post each MLP layer to maintain absorption coefficients between 0 and 1.

$$\hat{C}_i^{\text{med}}(r) = T_i^{\text{obj}} \, e^{-\hat{\sigma}^{\text{bs}} \hat{D}_i} \alpha_i c^{\text{med}}(r) \tag{6}$$

where $T_{\text{obj}}^i = \prod_{j=1}^{i} (1 - \alpha_j)$, c^{med} is color of medium related to ray direction r, and c_i^{obj} is the color of each Gaussian.

For the Medium-MLP, we follow a similar network structure to SeaThru-NeRF [10]. The network is expected to learn the three attributes of the medium: $\hat{\sigma}^{\text{attn}}$, $\hat{\sigma}^{\text{bs}}$, and $\hat{c}^{\text{med}}$.

2) Scene Enhancement with Light-MLP : Our enhancement method employs Medium-MLP to distinguish between the medium and object components. Specifically, Light-MLP enhances the properties of the object without altering the properties of the medium.

Following [16], the light field is modeled as a function of ray direction r, which is enhanced using a dynamic gamma correction:

$$\phi(r) = \theta(\varphi(r)) = \left(\frac{\varphi(r)}{\alpha(r)} \right)^{\frac{1}{\gamma_0 + \gamma(r)}} \tag{7}$$

where γ and α are learned through the Light-MLP shown in Fig. 3, which dynamically outputs γ and enhances φ, γ_0 acts as a base γ value and θ is the enhancement function.

Inspired by Retinex theory [17] which can effectively represent non-uniform light fields, we extend the formula in Sec. III-B.1, and decompose the color of the i-th Gaussian volume, into the base color of the Gaussian volume, expressed using lower-order spherical harmonics SH_i^{n}, and

the illumination color $\phi^{obj}(r)$ which effectively represent non-uniform light fields:

$$c_i^{\text{obj}}(r) = \text{SH}_i^{\text{n}} \cdot \phi^{\text{obj}}(r) \tag{8}$$

The goal of the Light-MLP is to learn and enhance the illumination pattern of the Gaussian volumes, with $\phi_{obj}(r)$ as the learned quantity output by the Enhancement Network in Eq. 7.

For our system, we modify $\hat{C}_i^{\text{obj}}$ in Eq. 5 by introducing illumination-related terms. The resulting object color component is expressed as:

$$\hat{C}_i^{\text{obj}}(r) = T_i^{\text{obj}}\, e^{-\hat{\sigma}^{\text{attn}}\hat{D}_i}\alpha_i SH_i^n \phi(r) \tag{9}$$

where SH_i^n represents the spherical harmonics which encodes low-frequency object colors, and SH_i^n models the illumination field.

3) Depth Regularization: In underwater scenes, complex illumination and moving objects introduce noise and discontinuities into depth-map estimation. Accurately capturing the boundaries and fine structural details in such environments is a crucial challenge. We utilize depth supervision similar to other solutions of radiance fields that have demonstrated their utility [18], [19]. Instead of relying on computationally expensive multi-view stereo (MVS) depth, we utilize recent deep learning-based monocular depth estimation and employ the Depth-Anything-V2 [20] pre-trained model as our monocular depth estimation model to provide dense guidance for optimization.

We employ the rasterization pipeline to render the depth map of Gaussian splats as follows:

$$\hat{D} = \sum_{i \in N} d_i \alpha_i T_i, \tag{10}$$

where $\hat{D}$ is the rendered depth and d_i is the depth of each splat from the camera. Eq. 10 enables the direct use of α_i and T_i calculated in Eq. 2, facilitating rapid depth rendering with minimal computational load.

To integrate the depth information effectively, we calculate the depth loss $\mathcal{L}_{depth}$ as follows:

$$\mathcal{L}_{depth} = \frac{1}{N}\sum_{i=1}^{N} \left\| \hat{D}(I_i) - D_\theta(I_i) \right\|_1 \tag{11}$$

where N is the number of pixels, $\hat{D}(I_i)$ is the estimated depth map from the Gaussian splatting model, and $D_\theta(I_i)$ is the depth map provided by the Depth Anything model.

Inspired by [21], we use an unsupervised depth smoothness constraint that minimizes the depth value differences between adjacent pixels. To avoid incorrect regularization in boundary areas, we use the Canny edge detector [22] as a mask, ensuring that regions with significant depth differences along boundaries remain unregularized. For a depth d_i and its adjacent depth d_j, we regularize the difference between them:

$$\mathcal{L}_{smooth} = \sum_{d_j \in \text{adj}(d_i)} \mathbf{1}_{ne}(d_i, d_j) \cdot \left\| d_i - d_j \right\|^2 \tag{12}$$

where $\mathbf{1}_{ne}$ is an indicator function that signifies whether both depths are not within edge regions.

C. Loss Functions

To enhance the model's ability to learn medium and illumination parameters, we assign distinct loss function constraints to different output components. The total loss function is defined as:

$$\mathcal{L} = \mathcal{L}_{\text{gs}} + \mathcal{L}_{\text{dep}} + \mathcal{L}_{\text{bs}} + \mathcal{L}_{\text{exp}} \tag{13}$$

where $\mathcal{L}_{\text{gs}}$ is the Gaussian Loss, $\mathcal{L}_{\text{dep}}$ is the Depth Loss, $\mathcal{L}_{\text{bs}}$ is the Backscatter Loss, and $\mathcal{L}_{\text{exp}}$ is the Exposure Loss.

1) Gaussian Loss: The Gaussian Loss enhances the optimization of the rendered underwater scene by accounting for both pixel intensity and structural similarity. In 3DGS[1], the loss function is combined with an $\mathcal{L}_1$ Loss and a $\mathcal{L}_{\text{D-SSIM}}$ Loss, which suits primitives without shared parameters. For the final render image $\hat{I}$ and the input image I, we adopt the weighted loss function $\mathcal{L}_{\text{Reg}}$ from WaterSplatting [11] to boost the weight of the dark regions in optimization to align with how humans perceive dynamic range. To be more specific, we apply a pixel-wise weight $W = \{w_{i,j}\}$ on both rendered estimate $\hat{I}$ and target image I, where $w_{i,j} = \left(sg(\hat{I}_{i,j}) + \epsilon\right)^{-1}$ with pixel coordinate (i, j) and $sg()$ denotes stopping gradient of its argument, which backpropagates zero derivative.

Thus, we have our regularized $\mathcal{L}_1$ Loss:

$$\mathcal{L}_{\text{Reg-1}} = \left| W \odot (\hat{I} - I) \right| \tag{14}$$

which results in less blur and sharper edges, and our regularized $\mathcal{L}_{\text{D-SSIM}}$ Loss:

$$\mathcal{L}_{\text{Reg-DSSIM}} = \mathcal{L}_{\text{D-SSIM}}(W \odot I, W \odot \hat{I}) \tag{15}$$

which encourages high structural similarity between I and $\hat{I}$. The Gaussian Loss function is:

$$\mathcal{L}_{gs} = (1 - \lambda_{gs})\mathcal{L}_{\text{Reg-1}} + \lambda_{gs}\mathcal{L}_{\text{Reg-DSSIM}} \tag{16}$$

where λ_{gs} is a balancing parameter that takes values between 0 and 1. This setting ensures that the dynamic range aligns more effectively with human eye perception.

2) Depth Loss: The method of depth regularization has already been discussed in Sec. III-B.3. The final Depth Loss terms are concluded by incorporating the depth loss Eq. 11 and smoothness loss Eq. 12 with their respective hyperparameters λ_{depth} and λ_{smooth}:

$$\mathcal{L}_{dep} = \lambda_{depth}\mathcal{L}_{depth} + \lambda_{smooth}\mathcal{L}_{smooth} \tag{17}$$

3) Backscatter Loss: Backscatter Loss is to improve the rendering of underwater scenes by accounting for the effects of light backscattering in the water medium. We optimize backscatter using the direct component D_c of the object. Specifically, for images from which the medium is removed, the DCP dark channel prior assumption generally holds, meaning that there is always a color channel in the object part

TABLE I: **The quantitative comparison results between ours and existing methods of the underwater rendering task on the SeaThru-NeRF dataset.** We show PSNR↑, SSIM↑, LPIPS↓, and Avg. FPS↑.

Dataset/Metric	Curaçao			IUI3-RedSea			J.G. Red Sea			Panama			Avg.
Method	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	FPS
Mip-NeRF 360 [23]	20.817	0.685	0.369	20.832	0.694	0.377	20.554	0.778	0.301	25.490	0.751	0.322	0.44
3DGS [1]	26.832	0.771	0.279	25.412	0.715	0.364	21.537	0.802	0.252	28.340	0.796	0.254	201.2
NeuralSea [24]	26.775	0.694	0.376	21.977	0.709	0.461	20.582	0.779	0.381	26.468	0.751	0.371	0.13
SeaThru-NeRF [10]	28.128	0.880	0.172	28.053	0.834	0.250	22.451	0.857	0.247	30.489	0.922	0.223	0.14
WaterSplatting [11]	29.469	0.826	0.149	29.959	0.915	0.201	23.308	0.868	0.116	32.116	0.932	0.082	40.5
Ours (w/o L-MLP)	36.643	0.919	0.116	35.877	0.948	0.113	34.474	0.892	0.111	37.947	0.945	0.081	190.3

*The best and second-best results are highlighted in red and blue.

whose pixel value is close to zero. We adopt the Backscatter Loss proposed by DeepSeeColor [13]:

$$\mathcal{L}_{bs} = \lambda_{bs} \sum_{i,j} \sum_{c} \left(\max\{\hat{D}_c(i,j), 0\} + k \cdot \min\{\hat{D}_c(i,j), 0\} \right) \tag{18}$$

with hyperparameter λ_{bs}, $k > 1$ and $\hat{D}_c = Je^{-\sigma^{attn} D}$.

4) Exposure Loss: The Exposure Loss controls the brightness of the rendered underwater image, ensuring that it aligns with the expected exposure level for accurate visual representation. Similar to many underwater works [12], [13] and low-light enhancement works [25], [26], we apply constraints on the exposure of $\hat{J}$, controlling the image brightness to match the expected level:

$$\mathcal{L}_{exp} = \lambda_{exp} \frac{1}{N} \sum_{i=1}^{n} \left(\frac{1}{M} \sum_{(x,y) \in \text{block}_i} \text{gray}(x,y) - \text{target} \right)^2 \tag{19}$$

where λ_{exp} is the hyperparameter, N is the number of blocks in the image, and M is the number of pixels per block. $block_i$ denotes the i-th block, $gray(x,y)$ is the grayscale value of the pixel located at coordinates (x,y) in the i-th block, and $target$ is the target mean exposure.

IV. EXPERIMENTS

A. Experimental Setup

1) Implementation details: We utilized the renderer implementation of 3DGS [1] and NerfStudio [27] as the backbone and extended their CUDA-based differentiable rasterization pipeline to support depth-map and accumulated-opacity rendering, as well as computations and gradient backpropagation for $\mathcal{L}_{bs}$, $\mathcal{L}_{dep}$, and $\mathcal{L}_{exp}$. During training, the loss weights were set as: λ_{gs}=0.2, λ_{depth}=0.05, λ_{smooth}=0.1, λ_{bs}=0.1 and λ_{exp}=0.01. The value k used to compute λ_{bs} is set to 1000. The attenuation coefficient σ^{attn} and backscatter coefficient σ^{bs} are both limited between 0 and 5. All other optimization parameters followed the original 3DGS settings. The proposed method was implemented using the Pytorch 2.5.1 framework with CUDA 12.1. All datasets were utilized for training runs for 30,000 iterations on a single NVIDIA GeForce RTX 4090 GPU.

2) Datasets: We evaluated the rendering quality of our method in a rendering experiment. The experiment uses the SeaThru-NeRF dataset [10], which includes underwater photographs from four regions: Curacao, IUI3-RedSea, JapaneseGardens-RedSea, and Panama. To better simulate the real working environment of AUVs and evaluate the visual effectiveness of our method, the enhancement experiments incorporate the DRUVA [28], RUIE [29], and SeaThru [7] datasets. The DRUVA dataset provides videos of 20 diverse real underwater scenes. The RUIE dataset contains unbalanced and uncorrected continuous-frame images. The SeaThru dataset offers continuous-frame images of five real low-light underwater scenes with ground truth as reference. For diversity, specific scenes from these datasets were selected for testing.

3) Metrics: Rendering quality was evaluated using Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM) [30], and Learned Perceptual Image Patch Similarity (LPIPS) [31]. To assess computational efficiency, frames per second (FPS) were recorded to compare the processing speed of our method against baseline methods. Moreover, visual inspection remains essential alongside quantitative metrics. All reported results are averaged over three runs.

B. Experimental results

1) Rendering Experiments: To evaluate our method for the underwater rendering task, we compared our method with five existing methods, including two general scene 3D representation methods, Mip-NeRF 360 [23] and 3DGS [1], and three underwater scene 3D representation methods NeuralSea [24], SeaThru-NeRF [10], and WaterSplatting [11]. Because existing methods do not perform underwater scene enhancement and due to the specific requirements of NeuralSea, we only compared the rendering quality assessment on the SeaThru-NeRF dataset with our method without Light-MLP (w/o L-MLP), while the Light-MLP-enabled method (GS-UVCE) is shown in Fig. 4 as a reference.

To quantitatively evaluate the proposed method, we employ four metrics to measure the results of each method as detailed in Table I. The table indicates that the proposed method consistently demonstrates superior performance across all metrics and datasets, achieving the highest PSNR and SSIM while exhibiting the lowest LPIPS. On the SeaThru-NeRF dataset, our method is comparable to the SOTA method WaterSplatting with a clear advantage. Compared to 3DGS, PSNR improves by around 40%.

Figure 4 presents the results of our method and the comparison method on the SeaThru-NeRF dataset. The lower left corner of each group of images shows the depth map of the rendered image, and the lower right corner shows the details within the red bording box in the enlarged rendered image.

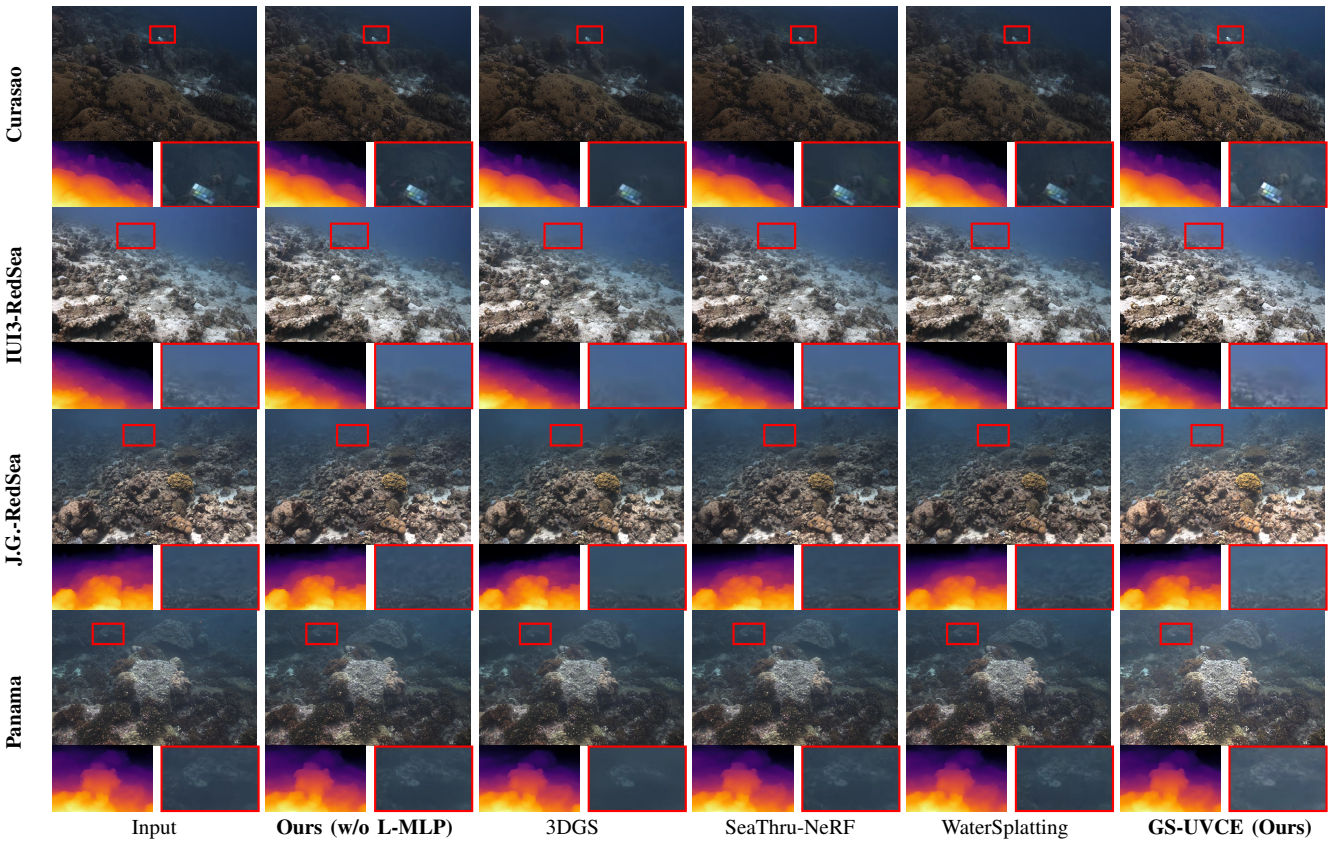


Fig. 4: Detailed comparison of reconstruction results for SeaThru-NeRF datasets. 3DGS fails to handle the medium, causing significant blurring of distant coral details. Our method outperforms Water Splatting and SeaThru-NeRF in maintaining distant details and accurately rendering underwater scenes.

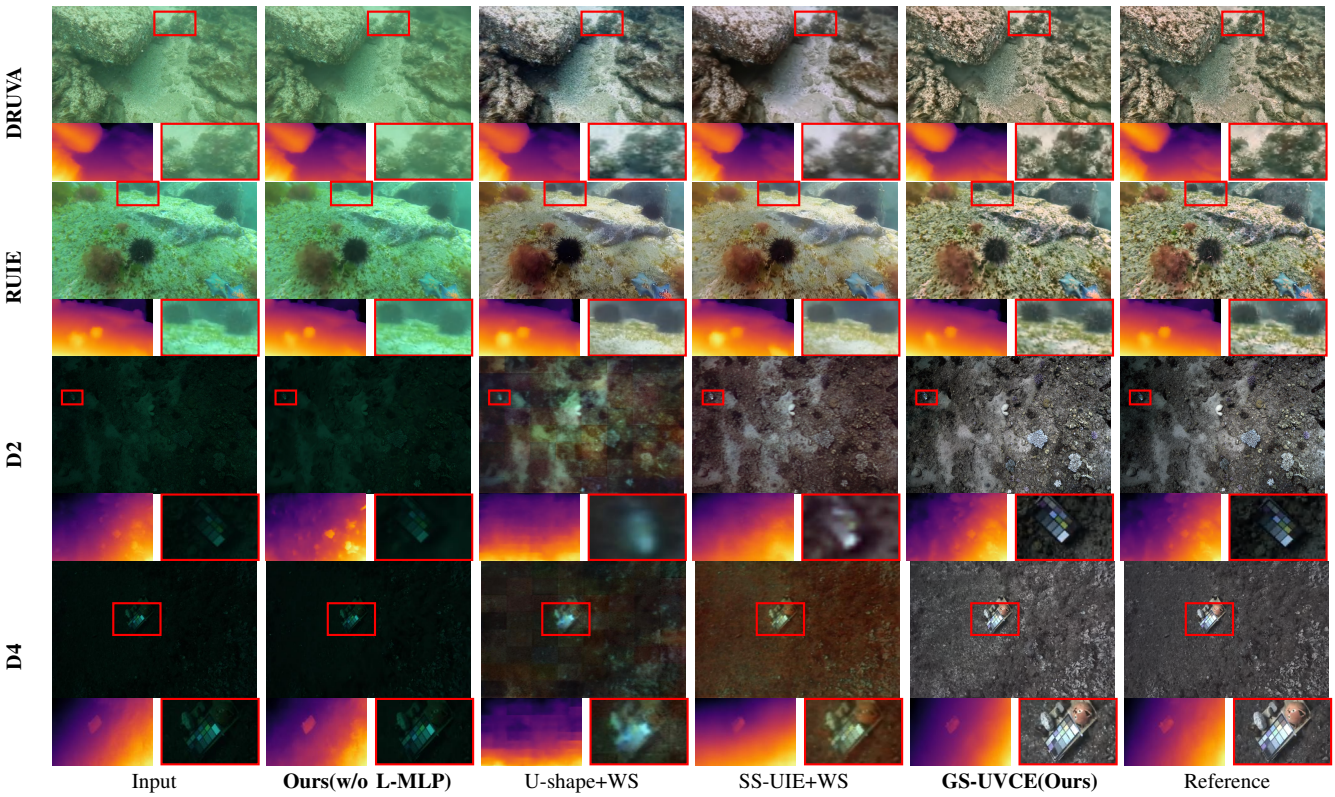


Fig. 5: Rendering results for underwater color cast scenes, and low-light scenes. We compare our method with WaterSplatting combined with the UIE SOTA methods U-shape and SS-UIE. Our method achieves better color restoration after medium removal, with overexposure and color distortion observed in pre-processing methods as highlighted in the red boxes.

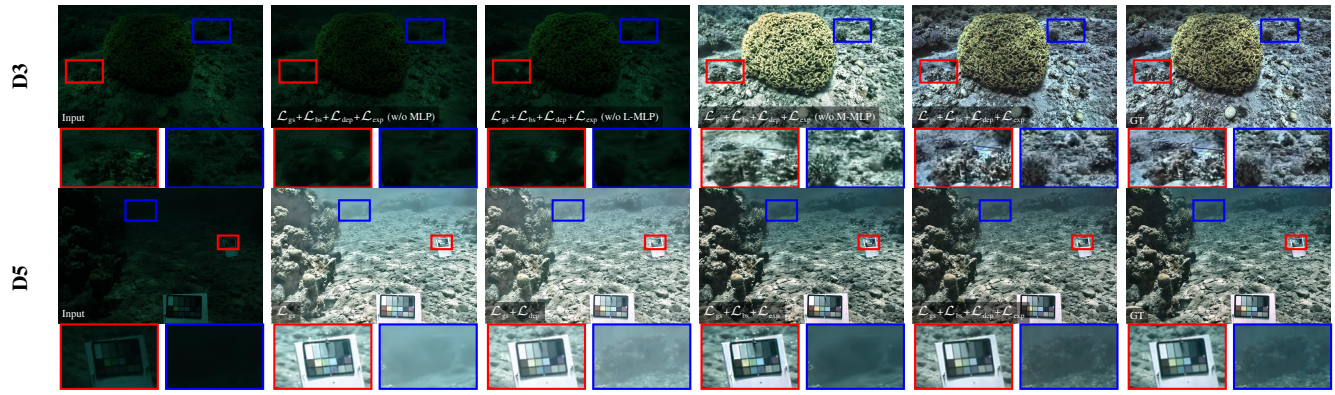


Fig. 6: Results of image rendering in ablation studies. The L-MLP method significantly improves the reconstruction quality of low light images, while the M-MLP method significantly improves the rendering representation of the water medium. Each loss function further enhances the fineness of the reconstruction results.

TABLE II: Processing capabilities of existing methods for underwater image enhancement.

Method	Low Light	Reconstruction	Unsupervised
U-shape [3]	×	×	×
DM-Water [4]	×	×	×
SS-UIE [2]	✓	×	×
SeaThru [7]	✓	×	✓
U-DNet [32]	×	×	✓
U2NeRF [33]	×	✓	✓
GS-UVCE (Ours)	✓	✓	✓

From Fig. 4, it can be observed that 3DGS fails to reconstruct distant areas, resulting in blurred details. In the Curacao sequence, WaterSplatting and SeaThru-NeRF blur seaweed details near the distant color checker. In the IUI3-RedSea, J.G.-RedSea, and Panama sequences, existing methods have difficulty restoring distant corals while retaining details. In contrast, our method can still retain distant details. With the help of L-MLP, GS-UVCE presents visual effects of distant local details that are superior to those of the input image.

2) Enhancement Experiments: To evaluate our method for underwater image enhancement, we compared our method with WaterSplatting combined with the SOTA UIE methods, U-shape [3] and SS-UIE [2] on DRUVA, RUIE, and SeaThru datasets. Ground truth images were used for supervised learning-based enhancement and rendering quality assessment. Figure 5 presents rendered images generated from the same pose as the input views for visual inspection. Compared to WaterSplatting combined with UIE, our method achieves better color restoration after medium removal, with overexposure and color distortion observed in processing methods as highlighted in the red boxes. U-shape cannot effectively handle low light scenes, resulting in severe blurring of details. Our method outperforms image preprocessing methods combined with WaterSplatting for rendering underwater scenes in terms of preserving distant details and accurately rendering underwater scenes. From Table II, among the existing underwater image enhancement methods, our method is the first to simultaneously achieve low-light processing, scene reconstruction, and unsupervised training.

TABLE III: Ablation Study results on SeaThru Datasets.

Loss Functions	**PSNR↑**	**SSIM↑**	**LPIPS↓**
$\mathcal{L}_{gs}$	20.014	0.526	0.306
$\mathcal{L}_{gs}+\mathcal{L}_{dep}$	20.465	0.556	0.261
$\mathcal{L}_{gs}+\mathcal{L}_{bs}$	22.084	0.725	0.247
$\mathcal{L}_{gs}+\mathcal{L}_{bs}+\mathcal{L}_{dep}$	22.739	0.781	0.249
$\mathcal{L}_{gs}+\mathcal{L}_{bs}+\mathcal{L}_{exp}$	29.930	0.821	0.223
$\mathcal{L}_{gs}+\mathcal{L}_{bs}+\mathcal{L}_{dep}+\mathcal{L}_{exp}$	30.600	0.887	0.129
$\mathcal{L}_{gs}+\mathcal{L}_{bs}+\mathcal{L}_{dep}+\mathcal{L}_{exp}$ (w/o MLP)	8.801	0.356	0.554
$\mathcal{L}_{gs}+\mathcal{L}_{bs}+\mathcal{L}_{dep}+\mathcal{L}_{exp}$ (w/o L-MLP)	8.980	0.368	0.462
$\mathcal{L}_{gs}+\mathcal{L}_{bs}+\mathcal{L}_{dep}+\mathcal{L}_{exp}$ (w/o M-MLP)	17.791	0.646	0.229

*The best and second-best results are highlighted in red and blue.

C. Ablation studies

To assess the contribution of each loss term and the proposed MLP, we designed nine control groups by selectively relaxing constraints from the loss function and the MLP module. To present our results objectively, we also provide the same image metrics with the given ground truth for reference. As shown in Fig. 6, removing specific terms from the proposed loss function or the MLP module consistently resulted in performance degradation in the outputs of our model. As shown in TABLE III, removing the L-MLP in GS-UVCE leads to a significant loss in the visual quality of images rendered in a low-light scene. The proposed method demonstrates good performance in both image rendering and image enhancement with complete losses.

V. CONCLUSIONS

This paper introduces GS-UVCE, a Gaussian Splatting-driven framework that unifies unsupervised visual consistency enhancement for underwater 3D scene reconstruction. By jointly modeling medium scattering through a Medium-MLP, correcting anisotropic illumination with a Light-MLP, and enforcing depth regularization, GS-UVCE achieves multi-view consistency enhancement and high-fidelity 3D reconstruction without ground truth supervision. Experiments on multiple real-world datasets show that GS-UVCE significantly outperforms state-of-the-art baselines in both image quality and reconstruction accuracy, while maintaining real-time rendering efficiency. These findings highlight the effectiveness of unifying enhancement and

reconstruction within a single framework and pave the way for more robust perception and navigation in challenging underwater environments.

REFERENCES

[1] B. Kerbl, G. Kopanas, T. Leimkuehler, and G. Drettakis, "3d gaussian splatting for real-time radiance field rendering," *ACM Transactions on Graphics (TOG)*, vol. 42, pp. 1 –14, 2023.

[2] L. Peng and L. Bian, "Adaptive dual-domain learning for underwater image enhancement," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, 2025, pp. 6461–6469.

[3] L. Peng, C. Zhu, and L. Bian, "U-shape transformer for underwater image enhancement," *IEEE Transactions on Image Processing*, vol. 32, pp. 3066–3079, 2023.

[4] Y. Tang, H. Kawasaki, and T. Iwaguchi, "Underwater image enhancement by transformer-based diffusion model with non-uniform sampling for skip strategy," in *Proceedings of the 31st ACM International Conference on Multimedia*, 2023, pp. 5419–5427.

[5] D. Berman, D. Levy, S. Avidan, and T. Treibitz, "Underwater single image color restoration using haze-lines and a new quantitative dataset," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 8, pp. 2822–2837, 2020.

[6] C. Boittiaux, R. Marxer, C. Dune, A. Arnaubec, M. Ferrera, and V. Hugel, "Sucre: Leveraging scene structure for underwater color restoration," in *2024 International Conference on 3D Vision (3DV)*, 2024, pp. 1488–1497.

[7] D. Akkaynak and T. Treibitz, "Sea-thru: A method for removing water from underwater images," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 1682–1691.

[8] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," *Commun. ACM*, vol. 65, no. 1, 99–106, Dec. 2021.

[9] A. V. Sethuraman, M. S. Ramanagopal, and K. A. Skinner, "Waternerf: Neural radiance fields for underwater scenes," in *OCEANS 2023-MTS/IEEE US Gulf Coast*, IEEE, 2023, pp. 1–7.

[10] D. Levy, A. Peleg, N. Pearl, *et al.*, "Seathru-nerf: Neural radiance fields in scattering media," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 56–65.

[11] H. Li, W. Song, T. Xu, A. Elsig, and J. Kulhanek, "WaterSplatting: Fast underwater 3D scene reconstruction using gaussian splatting," *3DV*, 2025.

[12] D. Yang, J. J. Leonard, and Y. Girdhar, "Seasplat: Representing underwater scenes with 3d gaussian splatting and a physically grounded image formation model," *arxiv*, 2024.

[13] S. Jamieson, J. P. How, and Y. Girdhar, "Deepseecolor: Realtime adaptive color correction for autonomous underwater vehicles via deep learning methods," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 3095–3101.

[14] J. Jaffe, "Computer modeling and the design of optimal underwater imaging systems," *IEEE Journal of Oceanic Engineering*, vol. 15, no. 2, pp. 101–111, 1990.

[15] D. Akkaynak and T. Treibitz, "A revised underwater image formation model," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[16] H. Wang, X. Xu, K. Xu, and R. W. Lau, "Lighting up nerf via unsupervised decomposition and enhancement," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2023, pp. 12 632–12 641.

[17] E. H. Land, "The retinex theory of color vision," *Scientific american*, vol. 237, no. 6, pp. 108–129, 1977.

[18] C. Gao, A. Saraf, J. Kopf, and J.-B. Huang, "Dynamic view synthesis from dynamic monocular video," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 5712–5721.

[19] W. Xian, J.-B. Huang, J. Kopf, and C. Kim, "Space-time neural irradiance fields for free-viewpoint video," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 9421–9431.

[20] L. Yang, B. Kang, Z. Huang, *et al.*, "Depth anything v2," *Advances in Neural Information Processing Systems*, vol. 37, pp. 21 875–21 911, 2024.

[21] J. Chung, J. Oh, and K. M. Lee, "Depth-regularized optimization for 3d gaussian splatting in few-shot images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 811–820.

[22] J. Canny, "A computational approach to edge detection," *IEEE Transactions on pattern analysis and machine intelligence*, no. 6, pp. 679–698, 2009.

[23] J. T. Barron, B. Mildenhall, D. Verbin, P. P. Srinivasan, and P. Hedman, "Mip-nerf 360: Unbounded anti-aliased neural radiance fields," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 5460–5469.

[24] T. Zhang and M. Johnson-Roberson, "Beyond nerf underwater: Learning neural reflectance fields for true color correction of marine imagery," *IEEE Robotics and Automation Letters*, vol. 8, no. 10, pp. 6467–6474, 2023.

[25] L. Ma, T. Ma, R. Liu, X. Fan, and Z. Luo, "Toward fast, flexible, and robust low-light image enhancement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 5637–5646.

[26] Y. Cai, H. Bian, J. Lin, H. Wang, R. Timofte, and Y. Zhang, "Retinexformer: One-stage retinex-based transformer for low-light image enhancement," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 12 504–12 513.

[27] M. Tancik, E. Weber, E. Ng, *et al.*, "Nerfstudio: A modular framework for neural radiance field development," in *ACM SIGGRAPH 2023 Conference Proceedings*, ser. SIGGRAPH '23, 2023.

[28] N. Varghese, A. Kumar, and A. N. Rajagopalan, "Self-supervised monocular underwater depth recovery, image restoration, and a real-sea video dataset," in *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 12 214–12 224.

[29] R. Liu, X. Fan, M. Zhu, M. Hou, and Z. Luo, "Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light," *IEEE transactions on circuits and systems for video technology*, vol. 30, no. 12, pp. 4861–4875, 2020.

[30] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[31] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," *IEEE*, 2018.

[32] A. Saleh, M. Sheaves, D. Jerry, and M. Rahimi Azghadi, "Adaptive deep learning framework for robust unsupervised underwater image enhancement," *Expert Systems with Applications*, vol. 268, p. 126 314, 2025.

[33] V. Gupta, S Manoj, M. Varma, and K. Mitra, "U2nerf: Unsupervised underwater image restoration and neural radiance fields," in *The Second Tiny Papers Track at ICLR 2024*, 2024.