

Visual enhancement and 3D representation for underwater scenes: a review

Received: 5 January 2026

Accepted: 19 May 2026

Published online: 04 June 2026

Cite this article as: Huang G., Wang H., Seymour B. *et al.* Visual enhancement and 3D representation for underwater scenes: a review. *Artif Intell Rev* (2026). <https://doi.org/10.1007/s10462-026-11597-4>

Guoxi Huang, Haoran Wang, Brett Seymour, Evan Kovacs, John Ellerbroc, Dave Blackham & Nantheera Anantrasirichai

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

Visual enhancement and 3D representation for underwater scenes: a review

Guoxi Huang^{1*}, Haoran Wang¹, Brett Seymour², Evan Kovacs³, John Ellerbroc⁴,
Dave Blackham⁵, Nantheera Anantrasirichai^{1*}

¹Visual Information Laboratory, University of Bristol, UK.

²Submerged Resources Center, National Park Service, USA.

³Marine Imaging Technologies, LLC, USA.

⁴Gates Underwater Products, Inc, USA.

⁵Esprit film and television Ltd, UK.

*Corresponding author(s). E-mail(s): guoxi.huang@bristol.ac.uk; n.anantrasirichai@bristol.ac.uk;

Abstract

Underwater visual enhancement (UVE) and underwater 3D reconstruction pose significant challenges in computer vision and AI-based tasks due to complex imaging conditions in aquatic environments. Despite the development of numerous enhancement algorithms, a comprehensive and systematic review covering both UVE and underwater 3D reconstruction remains absent. To advance research in these areas, we present an in-depth review from multiple perspectives. First, we introduce the fundamental physical models, highlighting the peculiarities that challenge conventional techniques. We survey advanced methods for visual enhancement and 3D reconstruction specifically designed for underwater scenarios. The paper assesses various approaches from non-learning methods to advanced data-driven techniques, including Neural Radiance Fields and 3D Gaussian Splatting, discussing their effectiveness in handling underwater distortions. Finally, we conduct both quantitative and qualitative evaluations of state-of-the-art UVE and underwater 3D reconstruction algorithms across multiple benchmark datasets. Finally, we highlight key research directions for future advancements in underwater vision.

Keywords: Underwater Image Enhancement , 3D Gaussian Splatting, NeRF, Underwater 3D Reconstruction

Statements and Declarations

Competing Interests: The authors declare no competing interests.

Funding This work was funded by the EPSRC ECR International Collaboration Grants (EP/Y002490/1) and the UKRI MyWorld Strength in Places Programme (SIPF00006/1).

Author contributions G.H.: Conceptualization, Visualization, Writing - original draft, Writing - review & editing; H.W.: Writing - original draft, Writing - review & editing; B.S: Supervision, Writing - review & editing;

E.K.: Supervision, Writing - review & editing; J.E.:Supervision, Writing - review & editing; D.B.:Supervision, Writing - review & editing; N.A.:Supervision, Funding acquisition, Writing - original draft, Writing - review& editing.

Data Availability. No datasets were generated or analyzed during the current study.

ARTICLE IN PRESS

Contents

1 Introduction

1.1 Peculiarities of Underwater Environments	
1.2 Motivation for Enhanced Underwater Imaging	
1.3 Scope of This Review and Literature Coverage	
1.4 Contributions	
1.5 Organization of the Paper	

2 Underwater Light Propagation and Image Formation

2.1 Absorption and Attenuation of Light	
2.2 Scattering Phenomena	
2.3 Jaffe–McGlamery Underwater Image Formation Model	
2.4 Simplified Underwater IFMs	

3 Underwater Visual Enhancement

3.1 Conventional Methods	
3.1.1 Statistical Approaches	
3.1.2 IFM-based Methods	
3.1.3 Retinex-based Methods	
3.1.4 Fusion-based Methods	
3.2 Data-Driven Approaches	
3.2.1 CNN-Based Methods	
3.2.2 Transformer-Based Methods	
3.2.3 Mamba-based Methods	
3.2.4 Diffusion Model-based Methods	
3.2.5 Learning Under Limited or No Paired Supervision	
3.3 Hybrid Approaches	
3.4 Evaluation and Benchmark	
3.4.1 Evaluation Metrics	
3.4.2 Performance Evaluation of various Enhancement methods	
3.5 Discussion and Open Challenges	

4 3D Reconstruction for Underwater Scenes

4.1 Photogrammetry	
4.1.1 Photogrammetry Approaches for Underwater Scenes.	
4.1.2 Real-Time Visual SLAM	
4.1.3 End-to-End Underwater MVS	

4.2	Neural Radiance Fields (NeRF)	
4.2.1	Principles and Volume Rendering	
4.2.2	NeRF Variants	
4.2.3	Underwater NeRF Applications	
4.3	3D Gaussian Splatting	
4.3.1	3DGS Variants	
4.3.2	Underwater 3DGS Applications	
4.4	Performance Evaluation of Underwater NeRF/3DGS Models	
4.5	Hybrid and Multi-Sensor Systems	
4.6	Discussion and Open Challenges	
5	Pipeline-Level Evaluation	
5.1	Pipeline-Level Reconstruction Case Studies	
6	Overall challenges and future work	
6.1	Cross-cutting Data, Supervision, and Deployment Challenges	
6.2	Foundation models for underwater imagery	
6.3	Perspectives on Datasets, Modeling Tools, and Evaluation	
6.4	Ethical Issues and Bias	
7	Conclusions	

1 Introduction

Underwater imaging plays an increasingly important role in scientific exploration, industrial inspection, and environmental monitoring. Because more than 70% of the Earth's surface is covered by water, vast biological resources, geological structures, archaeological remains, and critical infrastructure remain hidden beneath oceans, seas, lakes, and rivers. Visual observation of these submerged environments is therefore essential for applications such as marine biology (Shaker et al. 2023), archaeology (Ling et al. 2024), geological surveying (Yaqoob et al. 2025), and the inspection of subsea assets (Rahnama et al. 2025), including pipelines and offshore platforms. As the need for long-term ecosystem monitoring, resource management, hazard mitigation, and digital preservation continues to grow, underwater imaging has become a key enabling technology for both scientific understanding and operational decision-making.

Despite its importance, underwater visual sensing remains substantially more challenging than imaging in air. In addition to the practical difficulty and cost of acquiring data in submerged environments, underwater images are severely degraded by wavelength-dependent absorption, scattering, non-uniform illumination, suspended particles, and refractive distortions introduced by camera housings. These factors reduce contrast, distort colour, blur fine structures, and complicate reliable visual analysis. As a result, raw underwater imagery often lacks the visual fidelity required not only for human interpretation but also for downstream computer vision tasks.

For this reason, underwater visual enhancement has become an important and widely adopted pre-processing step. By improving visibility, restoring colour balance, and recovering local contrast, enhancement methods can make underwater data more suitable for subsequent analysis, including inspection, scene understanding, and 3D reconstruction. In practice, many underwater reconstruction pipelines benefit from enhanced inputs because improved image quality can facilitate feature detection, matching, camera pose estimation, and dense reconstruction. Consequently, underwater visual enhancement is not merely a perceptual refinement step, but often a practical means of improving the usability of underwater imagery for downstream geometric processing.

Alongside enhancement, underwater 3D reconstruction has also attracted substantial attention. Traditional 3D mapping pipelines frequently rely on photogrammetry, where Structure from Motion (SfM), visual Simultaneous Localization and Mapping (visual SLAM), and Multi-View Stereo (MVS) are used to estimate camera motion and recover scene geometry from overlapping images (Zhang et al. 2022b; Teague and Scott 2017; Storlazzi et al. 2016). Compared with laser scanning, photogrammetry is often more cost-effective for covering larger areas while preserving visually rich texture. More recently, advances in machine learning have introduced powerful alternatives for both enhancement and reconstruction. Deep networks have been used to learn underwater restoration from paired, synthetic, or unpaired data, while modern 3D representations such as Neural Radiance Fields (NeRF) (Mildenhall et al. 2020) and 3D Gaussian Splatting (3DGS) (Kerbl et al. 2023a) have shown strong potential for high-fidelity scene modelling from image collections.

Nevertheless, underwater conditions continue to challenge both enhancement and reconstruction. Severe degradation, unstable illumination, domain shifts, and limited high-quality reference data restrict the robustness of current methods. Moreover, although enhancement as a pre-processing step is often effective in practice, it is not always neutral with respect to downstream geometry. Enhancement may alter image statistics, local textures, edge structures, or cross-view photometric consistency, which can in turn influence matching, calibration, or reconstruction quality. This does not diminish the practical value of two-stage pipelines; rather, it suggests that the interaction between enhancement and reconstruction deserves closer examination. In this review, we therefore consider underwater visual enhancement primarily as an enabling pre-processing component for downstream reconstruction, while also identifying the analysis of cumulative error in sequential enhancement–reconstruction pipelines as an important future research direction.

1.1 Peculiarities of Underwater Environments

The major factors that make underwater imagery fundamentally different from terrestrial imagery. These include light absorption and scattering, non-uniform illumination, dynamic water conditions, marine snow, optical and geometric distortions, and dynamic scenes. Importantly, these factors degrade not only visual appearance but also the reliability of feature extraction, correspondence estimation, calibration, and geometric reconstruction.



Fig. 1 Examples of underwater images exhibiting wavelength-dependent colour casts and veiling effects (Liu et al. 2020).



Fig. 2 Example underwater images with non-uniform lighting: the top row shows images from the UIEB dataset (Li et al. 2020), while the bottom row presents images from the LSUI dataset (Peng et al. 2023).

Light Absorption and Scattering. Light attenuation in water is far more severe than in air and is governed by both absorption and scattering. Absorption, predominantly caused by water molecules and dissolved organic matter, is wavelength-dependent: red, orange, and yellow wavelengths are attenuated first, often giving underwater scenes a bluish or greenish appearance, as shown in Figure 1. In addition, suspended particles such as plankton and marine snow scatter the remaining light, creating a veil that reduces contrast and obscures fine details. These degradations make it significantly harder to detect reliable features and establish stable correspondences across views, both of which are essential for photogrammetric reconstruction. Correcting spectral imbalance and improving visibility are therefore central goals of underwater visual enhancement.

Non-Uniform Illumination. Lighting conditions underwater are often highly uneven, whether due to oblique sunlight, attenuation with depth, or artificial light sources mounted on cameras or vehicles. This leads to bright hotspots, dark shadow regions, and spatially varying colour responses within the same image. Figure 2 illustrates representative examples from the UIEB and LSUI datasets. Such non-uniform illumination complicates both enhancement and reconstruction: local contrast correction may over-amplify noise or saturate bright areas, while changing illumination patterns violate the photometric consistency assumptions commonly used in feature matching and multi-view geometry.

Dynamic Water Conditions. Oceanic and freshwater environments are rarely stable. Currents, waves, suspended matter, and biological activity produce rapid temporal changes in visibility and appearance, as shown in Figure 3. These fluctuations complicate frame-to-frame correspondence estimation and often lead to sparse or unstable reconstructions. For enhancement, temporal inconsistency may cause flickering or colour instability in video restoration. For reconstruction, dynamic appearance changes reduce the reliability of tracking, matching, and multi-view aggregation.

Marine Snow. Marine snow consists of suspended particulate matter such as organic debris, phytoplankton shells, and other drifting material that appears as snow-like structures in underwater imagery. As illustrated in Figure 4, these particles vary greatly in size, density, and reflectance, and can significantly degrade image clarity. Marine snow is particularly problematic for 3D reconstruction because it may create transient artefacts, false correspondences, and incomplete surface recovery (Malyugina et al. 2025). Effective enhancement methods must therefore suppress particulate interference while retaining scene details that are useful for later analysis and reconstruction.



Fig. 3 Examples of underwater images under dynamic environmental conditions (Xie et al. 2024).

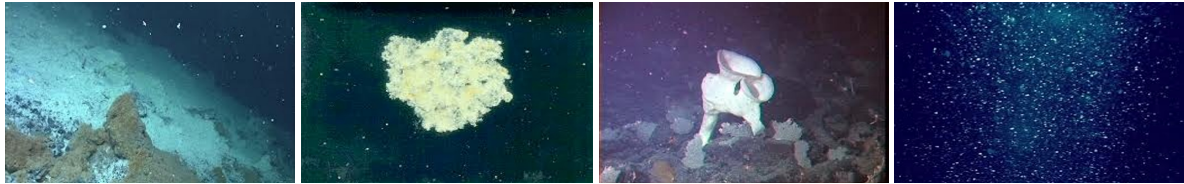


Fig. 4 Examples of underwater images with marine snow (Banerjee et al. 2014).

Optical and Geometric Distortions. Refraction at the interfaces between water, housing glass, and air causes projection behaviour that deviates significantly from the classic pinhole camera model. As a result, underwater images may exhibit nonlinear distortions that are not adequately captured by standard radial distortion correction alone. These distortions complicate camera calibration, feature matching, triangulation, and dense reconstruction, and may lead to systematic errors such as doming or bowing effects over large, approximately planar surfaces such as seafloors (Wright et al. 2020). Correcting refractive distortion is therefore essential for reliable underwater photogrammetry and 3D reconstruction.

Dynamic Scenes. In addition to dynamic water conditions, many underwater scenes contain genuine scene motion, including swaying vegetation, drifting schools of fish, and moving particulate layers. Furthermore, the camera platform itself, such as an ROV or AUV, may experience unpredictable motion due to currents. These effects violate the static-scene assumptions underlying many SfM, SLAM, and MVS pipelines, and often generate spurious correspondences. Robust underwater reconstruction must therefore identify stable scene structure while remaining tolerant to motion-induced outliers.

1.2 Motivation for Enhanced Underwater Imaging

The demand for robust underwater imaging solutions is driven by a combination of scientific, industrial, and environmental requirements. In marine science and environmental monitoring, image quality directly affects the ability to identify species, assess habitats, estimate populations, and track ecological change over time. In underwater archaeology, clear visual observations and accurate reconstructions support non-invasive documentation of submerged cultural heritage, where incomplete or distorted imagery can lead to misinterpretation of structures or artefacts. In industrial inspection, enhanced imagery and reliable 3D models are essential for assessing the condition of offshore infrastructure under poor visibility and for identifying defects that may compromise operational safety.

Across these applications, underwater visual enhancement serves an important practical role by improving the usability of degraded imagery before downstream analysis. In particular, enhancement is often employed as a pre-processing step to support reconstruction pipelines that would otherwise struggle with low contrast, colour distortion, and reduced texture visibility. At the same time, the interaction between enhancement and reconstruction deserves closer attention, since improvements in perceptual quality do not always guarantee improvements in geometric accuracy. This review therefore considers both the practical value of enhancement-based pre-processing and the need to better understand its downstream effects.

1.3 Scope of This Review and Literature Coverage

This review provides a methodologically grounded synthesis of techniques for addressing degradation in underwater imagery and enabling reliable 3D reconstruction in subaqueous environments. Works are selected based on their methodological contributions to imaging physics, visual enhancement, and geometric reconstruction, as well as their relevance, methodological clarity, and influence within the field; Selected conference papers are included to reflect fast-moving technical developments, while journal articles carry greater weight in the broader synthesis. We exclude, or cite only peripherally, papers focused primarily on downstream detection, tracking, classification, or segmentation unless they directly help illustrate task utility, evaluation needs, or deployment constraints. Likewise, generic clear-medium computer vision and graphics methods are discussed only when they provide necessary conceptual background or representative context for underwater extensions, and works with insufficient methodological detail are not emphasized in the synthesis.

We begin with underwater light propagation and image formation, establishing the physical basis for colour distortion, contrast attenuation, and visibility loss. Building on this, we examine underwater visual enhancement (UVE) methods spanning classical, physics-based, and learning-based approaches, followed by a survey of 3D reconstruction frameworks, including photogrammetry, structure-from-motion (SfM), visual SLAM, multi-view stereo (MVS), learning-based depth estimation, and neural scene representations such as NeRF and 3D Gaussian Splatting (3DGS). The review emphasizes the coupling between enhancement and reconstruction, particularly the role of UVE in influencing geometric accuracy and robustness. Datasets, evaluation protocols, and deployment constraints are also covered, with attention to cumulative error and pipeline-level robustness. Methods for underwater enhancement can be broadly grouped into three categories:

- **Physics-Based Models:** methods that explicitly model underwater attenuation and scattering (Berman et al. 2017; Li et al. 2021a) to restore colour balance and scene contrast;
- **Traditional Image Processing:** classical approaches such as histogram equalization, Retinex-based enhancement, and contrast adjustment, which are often computationally lightweight but may be less robust under severe degradation;
- **Deep Learning:** data-driven methods, including CNN- and GAN-based models, that learn mappings from degraded to enhanced images using paired, synthetic, or unpaired supervision.

We also examine how mainstream 3D reconstruction pipelines are adapted to underwater conditions:

- **SfM and MVS:** extensions of terrestrial photogrammetry that address refractive effects, unstable correspondences, and inconsistent lighting;
- **Learning-Based 3D:** neural depth estimation, volumetric reconstruction, and representation learning methods adapted to underwater inputs;
- **Sequential Pipelines:** practical workflows in which enhanced images are used as inputs to downstream reconstruction systems, together with discussion of their benefits and limitations.

1.4 Contributions

This review is motivated by the rapid growth of underwater imaging research and by the limitations of existing surveys, which often focus narrowly on either image restoration or 3D mapping. In contrast, we provide a broader and more connected perspective that examines underwater visual enhancement together with underwater 3D reconstruction, while preserving the practical viewpoint that enhancement is frequently used as a pre-processing stage. The main contributions of this paper are as follows:

- **A unified taxonomy of underwater enhancement and reconstruction methods:** we organize the literature across physics-based, traditional, and learning-based paradigms, and summarize their assumptions, strengths, and limitations;
- **A review of enhancement for downstream reconstruction:** we examine how underwater visual enhancement is used in practice as a pre-processing step for 3D reconstruction, and discuss when and why it is beneficial;

- **Open challenges and future directions:** we highlight unresolved issues including domain generalization, benchmark design, robustness under severe degradation, and the cumulative error that may arise in sequential enhancement and reconstruction pipelines.

This review is intended to provide a structured reference for researchers and practitioners working across underwater image restoration, geometric reconstruction, and real-world deployment.

1.5 Organization of the Paper

The remainder of this paper is organized as follows. Section 2 reviews the physics of underwater light propagation and image formation. Section 3 surveys underwater visual enhancement methods, spanning non-learning, data-driven, and hybrid approaches. Section 4 discusses underwater 3D reconstruction techniques, including traditional photogrammetry as well as recent methods based on NeRF and 3D Gaussian Splatting. Section 5 presents pipeline-level evaluation, including dataset context, evaluation criteria, and reconstruction case studies. Finally, Section 7 concludes the paper and outlines future research directions.

Figure 5 provides a compact roadmap of the paper structure and the methodological relationships emphasized throughout this review. In particular, underwater physics defines the degradation process, visual enhancement improves image quality for human interpretation and downstream processing, and modern reconstruction methods can often benefit from enhanced inputs, while future work is needed to better understand the cumulative effects of sequential pipelines.

Overall, the field of underwater visual enhancement and 3D reconstruction is entering a period of rapid development. By addressing the coupled optical, computational, and geometric challenges of subaqueous environments, future research can enable safer, more scalable, and more informative exploration of the underwater world.

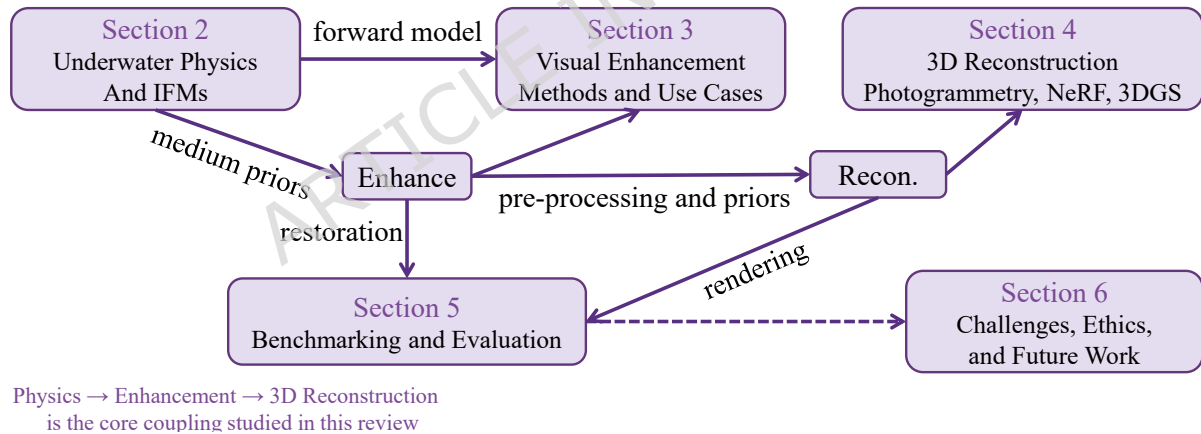


Fig. 5 Roadmap of the paper. The review is organized around the idea that underwater physics provides the forward model, visual enhancement improves image quality for both visual inspection and downstream processing, and modern reconstruction methods such as photogrammetry, NeRF, and 3D Gaussian Splatting can often benefit from enhanced inputs.

2 Underwater Light Propagation and Image Formation

Understanding the physical principles of underwater light propagation is fundamental to developing effective image enhancement and 3D reconstruction algorithms. Compared to terrestrial imaging, underwater photography encounters significantly more complex distortions arising from wavelength-dependent absorption, scattering by suspended particles, refractive effects at media interfaces, and non-uniform illumination. This section provides a detailed overview of these phenomena, highlights the Jaffe–McGlamery underwater image formation model (IFM) and its simplified variants, and discusses specialized calibration procedures required in underwater imaging.

Table 1 Penetration depths of different light wavelengths in clear seawater

Light Color	Wavelength (nm)	Approx. Penetration Depth
Ultraviolet (UV)	<400	<5 m
Blue Light	400–500	50–100 m
Green Light	500–550	30–50 m
Yellow Light	550–600	~20 m
Red Light	600–700	<5 m
Near-Infrared (NIR)	>700	<1 m

2.1 Absorption and Attenuation of Light

When light travels through water, its intensity decays exponentially due to both absorption and scattering. The Beer–Lambert law describes the attenuation of light as a function of the propagation distance:

$$I_d(x) = J(x) e^{-\beta(\lambda) d(x)}, \quad (1)$$

where $I_d(x)$ is the direct transmission component of the observed intensity at pixel x . $J(x)$ represents the ideal intensity from the object (i.e., what would be measured in a clear medium without attenuation). $\beta(\lambda)$ is the wavelength-dependent total attenuation coefficient, combining absorption and scattering effects. $d(x)$ is the object-to-camera distance.

Because attenuation coefficients vary across the visible spectrum, the usable color bandwidth narrows with increasing depth. As shown in Table 1, blue and green wavelengths penetrate more deeply in clear seawater, while red and near-infrared wavelengths attenuate rapidly. Strong attenuation of red wavelengths frequently imparts a bluish or greenish cast to underwater images.

2.2 Scattering Phenomena

In addition to absorption, *scattering* drastically impacts underwater imagery. Particulates (e.g., silt, algae) can deviate light from its path, reducing clarity and contrast. Scattering is typically divided into:

- **Forward scattering:** Light is deflected at small angles, leading to blurring.
- **Backscattering:** Light is scattered back toward the camera, creating a haze-like effect that reduces image contrast.

While both types degrade image quality, backscattering often proves more detrimental, as it adds a veil of background illumination. Although many haze-removal methods in atmospheric imaging (He et al. 2010) inspire underwater dehazing solutions, the scattering coefficients and spectral absorption underwater can differ substantially.

2.3 Jaffe–McGlamery Underwater Image Formation Model

A widely accepted image formation model (IFM) for underwater optical imaging was introduced by Jaffe and McGlamery (McGlamery 1980; Jaffe 1990), offering a comprehensive framework to describe how light interacts with water and suspended particulates. As illustrated in Figure 6, the model decomposes the observed signal into direct transmission I_d , forward scattering I_f , and backscatter I_b , corresponding respectively to signal-preserving object radiance, blur induced by small-angle scattering, and veil-like background illumination accumulated along the camera ray. These three components collectively determine the total irradiance recorded by the camera sensor and are dictated by factors such as water turbidity, imaging depth, and the dominant light wavelengths.

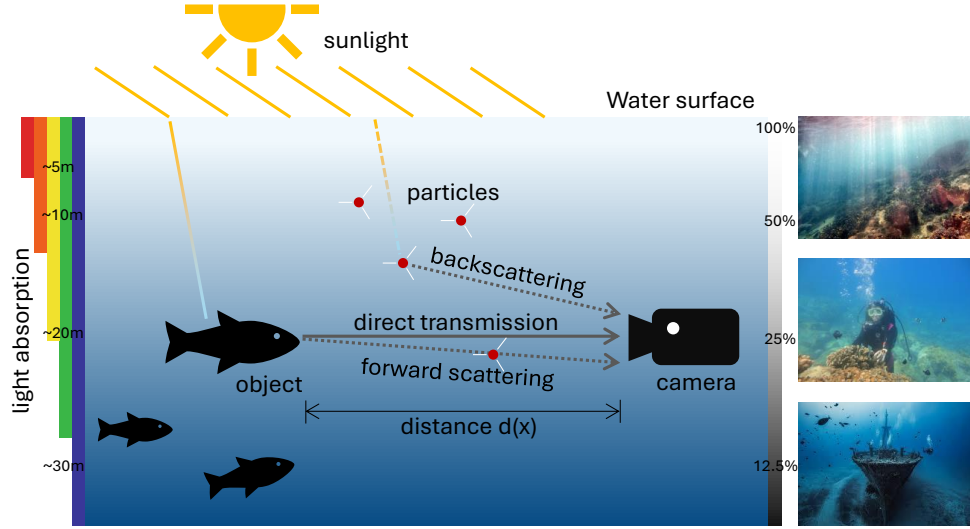


Fig. 6 Jaffe–McGlamery underwater IFM, depicting light absorption and the selective attenuation of underwater illumination. The diagram highlights the effects of direct transmission, forward scattering, and backscattering caused by suspended particles, all of which influence image quality. The color gradient illustrates the depth-dependent absorption of light, while the side images demonstrate varying levels of underwater visibility at different depths

We retain this classical formulation here only as the minimum physical background needed for the later discussions of simplified restoration models, revised IFMs, and physics-guided NeRF/3DGS methods.

Mathematically, the Jaffe–McGlamery model often expresses the captured intensity $I(x)$ of a pixel location x as:

$$I(x) = I_d(x) + I_f(x) + I_b(x), \quad (2)$$

where $I_d(x)$ is the direct transmission from the scene $I_f(x)$ is the forward-scattered term, and $I_b(x)$ is the backscattered term.

Transmission Map and Attenuation. The direct transmission component decays exponentially with distance, following the Beer–Lambert law:

$$I_d(x) = J(x) T(x), \quad (3)$$

where $J(x)$ is the scene radiance from the object and $T(x)$ is the transmission function:

$$T(x) = e^{-\beta(\lambda) d(x)}, \quad (4)$$

where $\beta(\lambda) = a(\lambda) + b(\lambda)$ denotes the *total* attenuation coefficient (absorption $a(\lambda)$ plus scattering $b(\lambda)$), and $d(x)$ is the object-camera distance.

Forward- and Backscattering Components. Forward scattering (I_f) introduces blur by deviating a portion of the light rays, whereas backscattering (I_b) adds an additional haze-like illumination:

$$\begin{aligned} I_f(x) &= \int_0^{d(x)} J(x) S_f(s) e^{-\beta(\lambda)s} ds, \\ I_b(x) &= \int_0^{d(x)} L(\lambda) S_b(s) e^{-\beta(\lambda)s} ds, \end{aligned} \quad (5)$$

where $S_f(s)$ and $S_b(s)$ represent phase functions describing the angular distribution of forward and backscattered light, respectively, and $L(\lambda)$ denotes the ambient light.

2.4 Simplified Underwater IFMs

Considering that the full Jaffe–McGlamery IFM is often too complex for real-time or large-scale applications, many practical systems adopt simplified assumptions (Bryson et al. 2016; Schechner and Karpel 2004).

Simplified Jaffe–McGlamery model. As $I_d(x) \gg I_f(x)$, the forward-scattering term $I_f(x)$ can be negligible. Further assuming a homogeneous medium with a constant $\beta(\lambda)$, and approximating the backscattering phase function S_b as isotropic (thus treated as constant), leads to a simpler integral form:

$$\begin{aligned} I_b(x) &= \int_0^{d(x)} L_b S_b e^{-\beta(\lambda)s} ds \\ &= L(\lambda) S_b \int_0^{d(x)} e^{-\beta(\lambda)s} ds \\ &= L(\lambda) S_b \frac{1 - e^{-\beta(\lambda)d(x)}}{\beta(\lambda)} \\ &= A(\lambda) [1 - T(x)], \end{aligned} \quad (6)$$

where $A(\lambda) = \frac{L(\lambda) S_b}{\beta(\lambda)}$ is the spatially invariant ambient light. Thus, a widely used *simplified Jaffe–McGlamery IFM* for the observed intensity I becomes:

$$I(x) = J(x)T(x) + A(\lambda) [1 - T(x)]. \quad (7)$$

This simplified IFM captures the essential interplay between direct attenuation and backscatter while omitting the more complex forward-scattering integral. Despite its approximations, it remains effective for many underwater imaging tasks, especially when water clarity is moderate and the scene is relatively close to the camera.

Atmospheric Scattering Model (ASM). Considering water-induced degradation is similar to haze in aerial images, several works (Chiang and Chen 2012; Peng and Cosman 2017; Peng et al. 2018; Li et al. 2021a; Schechner and Karpel 2004; Berman et al. 2017; Carlevaris-Bianco et al. 2010; Drews Jr et al. 2013; Lu et al. 2015) also treat underwater image formation as an extension of the atmospheric scattering model (ASM) (Narasimhan and Nayar 2002, 2003; Tan 2008; Fattal 2008; Narasimhan and Nayar 2000). The simplified mathematical representation is given by:

$$I(x) = J(x)T(x) + A[1 - T(x)]. \quad (8)$$

Compared to the simplified Jaffe–McGlamery model in Eq. (7), the ASM assumes that the ambient illumination remains constant across the spectrum. The function of ASM is to remove the veiling effect, similar to dehazing, but it cannot correct color cast issues. In shallow water regions, we can assume that the attenuation rate of all wavelengths is consistent. Therefore, ASM can achieve a similar effect to Eq. (7) in shallow underwater scenes (1–5 m). However, in underwater scenes beyond 5 m, ASM-based images tend to exhibit a noticeable green or blue color cast.

RGB Channel-Based ASM. For practical applications, this model is often expressed in terms of the RGB channels (Fattal 2008; Tarel and Hautiere 2009; Peng et al. 2018):

$$\begin{aligned} I^c(x) &= J^c T^c(x) + A^c [1 - T^c(x)], \\ T^c(x) &= e^{-\beta^c d(x)}, \quad \text{for } c \in \{R, G, B\}. \end{aligned} \quad (9)$$

Here, the coefficients $\beta_R \gg \beta_G > \beta_B$ indicate that red light is absorbed more rapidly than green and blue light, leading to the characteristic blue-green appearance of underwater images. By applying a simple transformation to Eq. (9), we can calculate the scene radiance $J^c(x)$ by:

$$J^c(x) = \frac{I^c(x) - A^c}{T^c(x)} + A^c \quad (10)$$

The RGB ASM can be considered an intermediate-complexity physical model between the simplified Jaffe–McGlamery model (Equation 7) and ASM (Equation 8). It reduces the dependence on the wavelength λ by leveraging RGB channels while addressing the color distortion issue that ASM fails to handle.

Revised underwater image formation model. A key refinement, particularly relevant for modern restoration and neural rendering, is the revised underwater image formation model of [Akkaynak and Treibitz \(2018\)](#). Instead of sharing one attenuation term between object radiance decay and backscatter accumulation, the revised model separates the direct-signal attenuation and the backscatter growth:

$$\begin{aligned} I^c(x) &= D^c(x) + B^c(x), \\ D^c(x) &= J^c(x) e^{-\beta_D^c d(x)}, \\ B^c(x) &= B_\infty^c \left(1 - e^{-\beta_B^c d(x)}\right), \end{aligned} \tag{11}$$

where β_D^c and β_B^c denote distinct wideband coefficients for direct transmission and backscatter, respectively. Compared with the simplified models in Eqs. (7)–(9), this formulation avoids conflating two different physical processes and therefore provides a more suitable starting point for SeaThru-style restoration, depth-aware color recovery, and underwater NeRF/3DGS pipelines that explicitly model the medium.

In the next section, we build on these physical insights to survey underwater image enhancement methods, including purely physics-based restoration, histogram-based techniques, and Retinex-based corrections. Comprehending these foundational methods provides a basis for later analysis of more advanced, learning-centric pipelines.

3 Underwater Visual Enhancement

Improving underwater imagery involves numerous challenges, including color inconsistencies due to selective wavelength absorption, light scattering from suspended particulates, and viewpoint-dependent refraction effects. This section provides a detailed literature survey of the existing approaches, spanning from traditional statistical-based to data-driven deep-learning methods. While some algorithms rely on simplified assumptions (*e.g.*, uniform attenuation), others incorporate domain knowledge or advanced neural architectures to handle in-situ complexities. We group the methods according to their underlying strategies and highlight open research problems for future development.

As summarized later in Table 3, underwater enhancement has evolved from prior- and IFM-driven restoration toward increasingly data-driven paradigms that must simultaneously handle visual quality, domain shift, and downstream task utility.

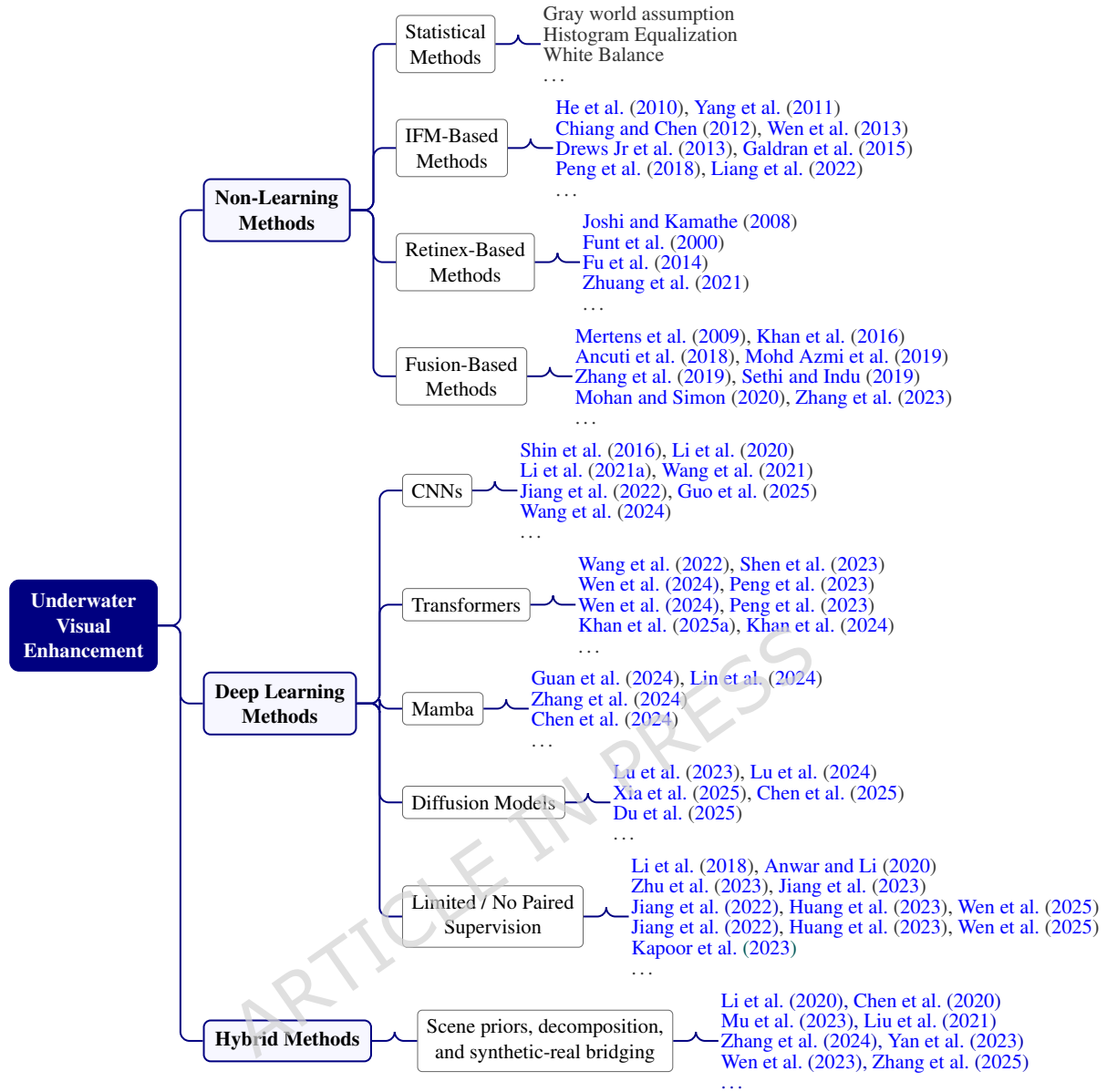


Fig. 7 Taxonomy of selected underwater image and video enhancement works across non-learning and learning paradigms. Table 3 provides a complementary chronological view of the rise of deep-learning methods in UIE.

3.1 Conventional Methods

3.1.1 Statistical Approaches

Histogram Equalization. The simplest approaches to enhancing underwater imagery involve histogram stretching, similar to contrast enhancement for images in a clear medium. Figure 8 presents the enhanced underwater images using various histogram equalization techniques, including histogram equalization (HE), adaptive histogram equalization (AHE) (Pizer et al. 1987), and contrast-limited adaptive histogram equalization (CLAHE) (Pizer et al. 1990). These methods enhance contrast in underwater imagery to varying degrees, with AHE and CLAHE providing more localized adjustments. CLAHE is a typical baseline in this category.

Approaches for underwater scenes are usually slightly more advanced, incorporating channel compensation, as different wavelengths of light affect the image appearance differently. Many works rely on global or local

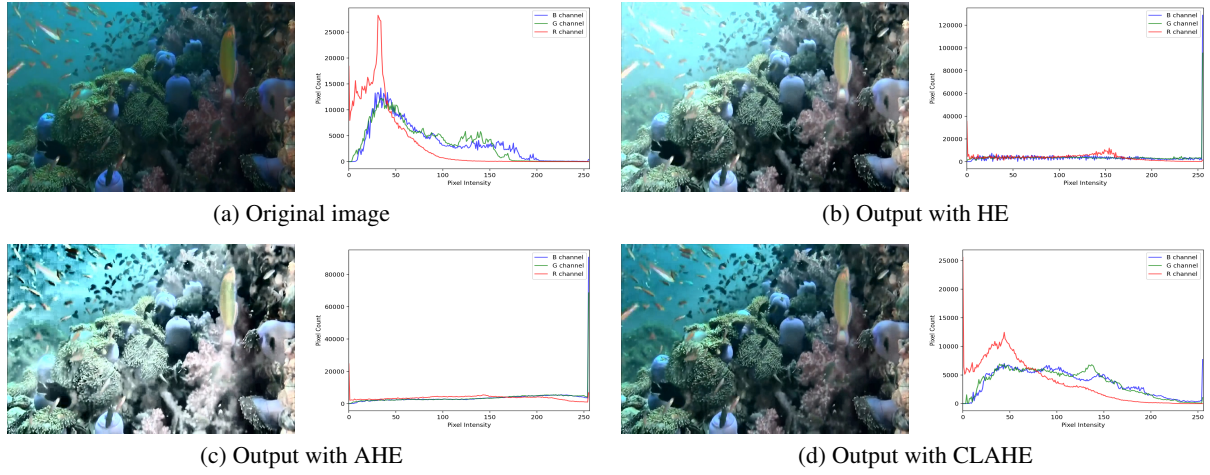


Fig. 8 Comparison of different histogram equalization techniques applied to the underwater image

histogram adjustments to address color bias. For instance, [Zhou et al. \(2023\)](#) propose sub-histogram equalization across multiple intervals, while [Zhang et al. \(2024\)](#) incorporate pixel-level gradient constraints for channel-specific stretching. Although computationally light, purely histogram-driven approaches often lack the spatial adaptivity to handle backscatter or patchwise variations in clarity.

3.1.2 IFM-based Methods

Traditional prior-based underwater image enhancement methods often adapt single-image RGB dehazing schemes from atmospheric context to underwater conditions by modifying them for wavelength-dependent attenuation. In the single RGB underwater image enhancement methods based on RGB, the recovered image can be obtained using Eq. (10), where the transmission map $T(x)$ and ambient light A^c are unknown variables. Therefore, we can use some prior assumptions to estimate the two variables.

For many IFM-based methods, the estimation pipeline can be summarized compactly as

$$\hat{A}^c = \Phi_A(I), \quad \hat{T}^c = \Phi_T(I), \quad \hat{J}^c(x) = \frac{I^c(x) - \hat{A}^c}{\max(\hat{T}^c(x), \epsilon)} + \hat{A}^c, \quad (12)$$

where Φ_A and Φ_T denote the background-light and transmission estimators induced by a specific prior. Most classical variants therefore differ less in the reconstruction formula itself than in how they parameterize $\hat{A}^c$ and $\hat{T}^c$.

Transmission Map Estimation with Dark Channel Prior (DCP). Originally introduced for atmospheric haze removal ([He et al. 2010](#)), DCP estimates the transmission map $T(x)$ and was later applied to underwater imagery by adjusting β^c in Eq. (9) to account for color-selective absorption ([Peng et al. 2018](#)). DCP exploits the observation that, in most local patches $\Omega(x)$, at least one color channel tends to have near-zero intensity in clear scenes. Formally, the dark channel is:

$$J_{\text{dark}}^{RGB}(x) = \min_{c \in \{R, G, B\}} \left[\min_{y \in \Omega(x)} (J^c(y)) \right] \approx 0. \quad (13)$$

Through DCP (Equation 13) and the RGB ASM (Equation 9), the transmission map can be estimated simply by:

$$\tilde{T}(x) = 1 - \min_c \left[\min_{y \in \Omega(x)} \left(\frac{I^c(y)}{A^c} \right) \right], \quad (14)$$

where A^c is a constant value selected from one of the farthest and haziest pixels in the input image. After the transmission map is estimated, the recovered image can be calculated using Eq. (10). DCP, as a simple prior assumption, performs well in atmospheric haze environments. However, due to its overly simplistic assumption, it

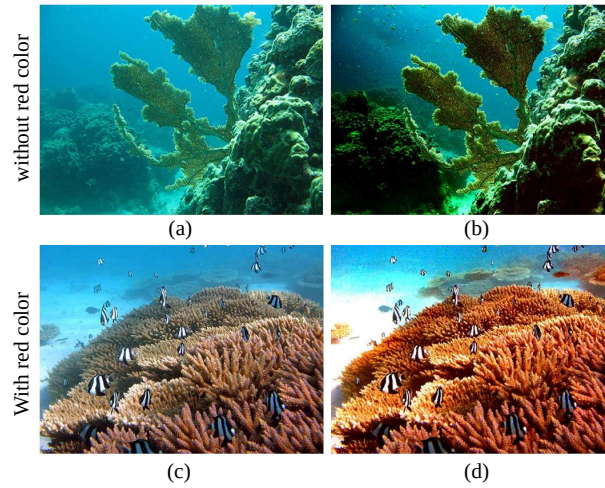


Fig. 9 Enhanced underwater images using MIP (Carlevaris-Bianco et al. 2010). Images (a) and (c) are the input images with and without red color, respectively, while images (b) and (d) are their corresponding enhanced results. MIP fails in images with blue-green dominant light, leading to issues such as loss of details

cannot be directly applied to other scenarios, such as sandstorms and underwater turbid images. Below, we outline some notable issues with DCP: 1) A key issue with DCP's transmission map estimation (14) is its assumption of uniform transmission across color channels (*i.e.*, $T^c(x) := \tilde{T}(x)$ for $c \in \{R, G, B\}$) (Galdran et al. 2015). This renders DCP ineffective in addressing wavelength-dependent color casts, a limitation that has emerged as a key research focus in subsequent studies based on DCP. 2) Meanwhile, DCP tends to overestimate the transmission map in certain regions, leading to color distortions and halo artifacts around edges, according to Huang et al. (2014, 2015). 3) In underwater photography with artificial light sources, the light intensity decreases with distance, which is the opposite of the assumption about ambient light A in DCP, according to Peng et al. (2018) and Peng and Cosman (2017). A follow-up study (Chao and Wang 2010) directly applies DCP without modifications to underwater image processing, but the resulting visual quality shows limited improvement. Nevertheless, the authors highlight that the normalized image (I^c/A) can mitigate the impact of wavelength-dependent color absorption in underwater images.

DCP variants for Underwater. Inspired by DCP, Carlevaris-Bianco et al. (2010) proposed the maximum intensity prior (MIP) to estimate a coarse depth estimation by leveraging the difference between the red channel and the blue-green channels, $\max_{x \in \Omega} I^R(x) - \max_{x \in \Omega, c \in \{B, G\}} I^c(x)$. However, since it relies entirely on the existence of red light, this algorithm is not applicable in deep-sea environments where red light is absent, as shown in Figure 9. As red light attenuates much faster than blue and green light, the smallest value among the RGB channels is always in the red channel in deep water scenes. Consequently, the DCP in RGB channels, termed DCP_{RGB} , becomes merely a zero map, which leads to an erroneous transmission map and results in poor restoration, as shown in Figure 11. To address this problem, research works, such as those proposed by Wen et al. (2013), Drews Jr et al. (2013) and Emberton et al. (2015), calculate the dark channel based only on the blue and green channels, termed DCP_{GB} . For instance, Drews Jr et al. (2013) and Drews et al. (2016) propose Underwater Dark Channel Prior (UDCP) by focusing on the blue and green channels, typically dominant underwater. Later, Liang et al. (2022) proposed a generalized method of Underwater Dark Channel Prior (GUDCP), estimating image transmission from multiple spectral profiles of different water types, enhancing its robustness across varied underwater conditions.

Galdran et al. (2015) proposed the Red Channel Prior, which mitigates the erroneous transmission estimation caused by the small values in the red channel by inverting the values of the red channel. Additionally, it computes separate transmission maps for the three color channels: $T^R(x)$, $T^G(x)$, and $T^B(x)$. Chiang and Chen (2012) proposed a hybrid method combining wavelength compensation (to address color distortion from depth-dependent absorption) and dehazing (to reduce scattering effects) by modeling how longer wavelengths (*e.g.*, red) attenuate more rapidly than shorter ones (*e.g.*, blue/green). Their semi-inverse approach estimates an approximate α_c for each channel using: $\beta_c = (a_c + b_c) \beta(d(\mathbf{x}))$, where a_c and b_c capture absorption and scattering effects for color channel c , and $\beta(d)$ modulates these based on depth or water type. The estimated transmission map (TM) based on DCP exhibits block-like artifacts, resulting in a halo effect and blurred edges, even when soft matting (Levin et al. 2008) is applied to mitigate this issue. To eliminate the halo effect and preserve boundaries, Yang et al. (2011)

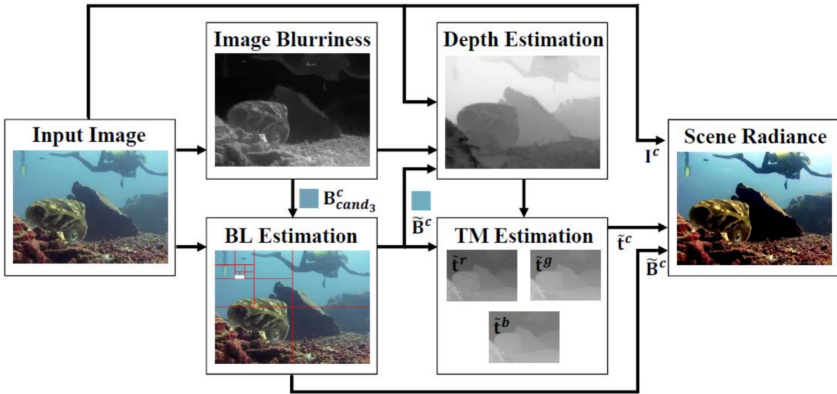


Fig. 10 Illustration of an integrated framework, combining light absorption modeling and image blurriness estimation for UVE (Peng and Cosman 2017)

and Gibson et al. (2012) proposed the Median DCP. In short, this method replaces $\min_{y \in \Omega(x)}$ in Eq. (14) with $\text{med}_{y \in \Omega(x)}$, where med represents a median filter.

In deep underwater scenes where sunlight is weak or absent, artificial light becomes the dominant illuminant. Under such conditions, points closer to the camera appear brighter, while those further away in the background appear darker. This illumination pattern directly contradicts the assumption of DCP. This means simply relying on color information for transmission estimation is not enough. Therefore, [Peng et al. \(2015\)](#) proposed a method to estimate the transmission map and scene depth based on the level of blurriness in the scene, considering that objects further away from the camera exhibit a blurrier appearance due to the scattering effect. This approach effectively restores underwater images that deviate from the assumptions of DCP- or MIP-based methods, as it does not rely on color channel information for underwater scene depth estimation. In their sequel ([Peng and Cosman 2017](#)), the authors further integrated light absorption and image blurriness information to estimate the transition map, the flowchart of which is presented in [Figure 10](#).

However, previous DCP-based or MIP-based methods often fail to estimate the background light A in complex underwater conditions. Song et al. (2020) argue that the poor performance of previous DCP-based methods stems from the overly aggressive assumption that $J_{\text{dark}}^{RGB} = 0$. To improve the accuracy of background light and transmission map estimation, the authors conducted a statistical analysis on 500 high-quality underwater images (i.e., images with minimal distortion) and found that the actual value is closer to $J_{\text{dark}} = 0.1$. Based on this finding, they proposed the New Underwater Dark Channel Prior (NUDCP), which adopts a less aggressive darkness assumption and leverages high-quality underwater images to mitigate artificial lighting distortions. Additionally, their two-step enhancement approach (Restoration + Color Correction) further improves contrast, visibility, and color fidelity. However, the real underwater dataset used in their study, limited to 500 images, may not generalize well to all underwater conditions.

To facilitate a clearer comparison and understanding of various IFM-based UVE methods, [Table 2](#) summarizes the formulas used for background light and transmission map estimation. Additionally, [Figure 11](#) presents qualitative comparisons of different MIP- and DCP-based approaches. As illustrated in the figure, these IFM-based methods struggle to effectively correct color distortion when not supplemented with additional color correction techniques such as histogram equalization or white balance.

So far, IFM-based methods, including those not previously mentioned (Liu and Chau 2016; Liang et al. 2022; Peng et al. 2018; Li et al. 2016; Wang et al. 2018), have struggled to directly correct color distortion in images captured in blue-green seawater. In most cases, additional color correction is needed as a post-processing step to mitigate this issue. This limitation has led to the emergence of fusion- and Retinex-based methods, which offer improved color correction capabilities compared to purely IFM-based approaches. Overall, these single-image priors prioritize efficiency, requiring limited computation beyond local patch statistics. Although assumptions like horizontally homogeneous water or minimal forward scattering can be restrictive, the resulting simplicity still offers a favorable blend of speed and effectiveness in moderately challenging conditions.

Table 2 Formulas for Estimation of background light (BL) and transmission map (TM) in underwater visual enhancement methods

Method	BL Estimation (A or A^c)	TM Estimation ($\tilde{T}$ or $\tilde{T}^c$)
Chao and Wang (2010)	$I^c(\arg \max_x p(x))$	$\tilde{T}(x) = 1 - \min_c \left(\min_{y \in \Omega(x)} \frac{I^c(y)}{A^c} \right)$
Carlevaris-Bianco et al. (2010)	$I^c(\arg \min_x \tilde{T}(x))$	$\tilde{T}(x) = D_{\text{MIP}}(x) + (1 - \max_x D_{\text{MIP}}(x))$
(Yang et al. 2011)	$I^c(\arg \max_{x \in p_{0.1\%}} (\sum_c I^c(x)))$	$\tilde{T}(x) = 1 - \min_c \left(\text{med}_{y \in \Omega(x)} \frac{I^c(y)}{A^c} \right)$
(Chiang and Chen 2012)	$I^c(\arg \max_x I_{dar_k}^c(x))$	$\tilde{T}^R(x) = 1 - \min_k \left(\min_{y \in \Omega(x)} \frac{I^c(y)}{A^c} \right),$ $\tilde{T}^c = (\tilde{T}^R)^{\frac{\beta^c}{\beta R}}$
(Wen et al. 2013)	$I^c \left(\arg \min_x (I_{dar_k}^R(x) - \max_{c'} (I_{dar_k}^{c'}(x))) \right)$	$\tilde{T}^c(x) = 1 - \min_{c'} \left(\min_{y \in \Omega(x)} \frac{I^{c'}(y)}{A^{c'}} \right),$ $\tilde{T}^R = (\tau \max_y I^R(y)), \tau = \frac{\text{avg}_x(\tilde{T}^c(x))}{\text{avg}_x(\max_{y \in \Omega(x)} I^R(y))}$
(Drews Jr et al. 2013)	$I^c(\arg \max_x p(x))$	$\tilde{T}(x) = 1 - \min_k \left(\min_{y \in \Omega(x)} \frac{I^R(y)}{A^R} \right)$
(Galdran et al. 2015)	$I^c(\arg \min_{x \in p_{10\%}} I^R(x))$	$\tilde{T}^c(x) = 1 - \min \left(\frac{\min_{y \in \Omega(x)} (1 - I^R(y))}{1 - A^R}, \frac{\min_{y \in \Omega(x)} I^G(y)}{A^G} \right),$ $\frac{\min_{y \in \Omega(x)} I^B(y)}{A^B}$
(Zhao et al. 2015)	$I^c(\arg \max_{x \in p_{0.1\%, c'}} I^R(x) - I^{c'}(x))$	$\tilde{T}^R(x) = 1 - \min_c \left(\min_{y \in \Omega(x)} \frac{I^c(y)}{A^c} \right),$ $\tilde{T}^c = (\tilde{T}^R)^{\frac{\beta^c}{\beta R}}$
(Peng et al. 2015)	$\frac{1}{ p_{0.1\%} } \sum_{x \in p_{0.1\%}} I^c(x)$	$\tilde{T}(x) = F_s(P_{blr}(x))^{\dagger}$

p denotes the gray-level pixel intensity of observed image I .

$c \in \{R, G, B\}$ and $c' \in \{G, B\}$

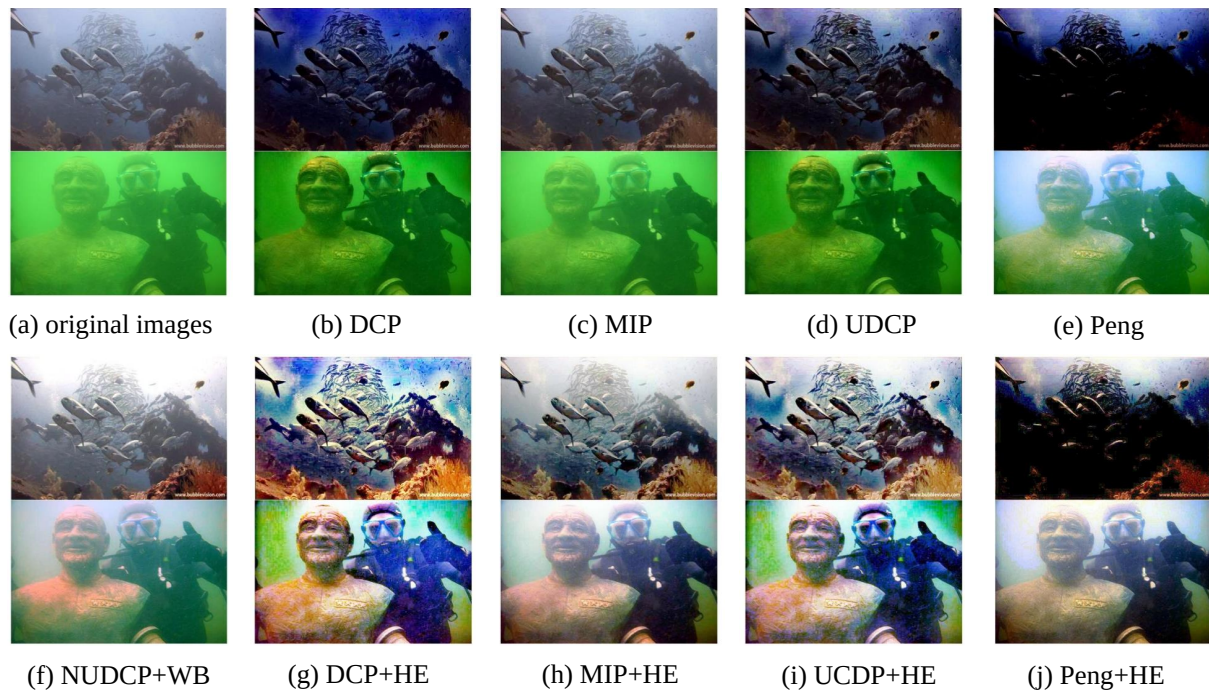


Fig. 11 Comparative results of various underwater image enhancement methods. (a) Original images, (b) DCP (He et al. 2010), (c) MIP (Carlevaris-Bianco et al. 2010), (d) UDCP (Drews Jr et al. 2013), (e) Peng (Peng and Cosman 2017), (f) NUDCP + WB (white balance) (Song et al. 2020), (g) DCP + HE (histogram equalization), (h) MIP + HE, (i) UDCP + HE, and (j) Peng + HE

3.1.3 Retinex-based Methods

Beyond the dark-channel variants, Retinex-based approaches decompose images into reflectance and illumination components, thereby tackling non-uniform lighting. For example, Hou et al. (2020) show that pre-estimating a color restoration map significantly helps with strong color cast removal. More recent approaches adopt multi-color space decompositions (e.g., HSV or Lab domains) to handle severe color shift, as in the UIEC²-Net (Wang et al. 2021) or the WCID approach (Chen et al. 2021), both enhancing results for a variety of underwater conditions. Despite such progress, prior-based methods sometimes produce oversaturated reds or overcorrected backgrounds in complex scenes.

3.1.4 Fusion-based Methods

Fusion-based methods aim to integrate the strengths of multiple enhancement techniques to overcome the limitations of individual approaches when restoring degraded underwater images. In these methods, separate processes such as color correction, contrast enhancement, and dehazing are first applied to generate different "views" of the input image. Then, using various fusion strategies, these complementary results are combined into a single enhanced image with improved visibility, natural colors, and better contrast.

In earlier work, Mertens et al. (2009) introduced a straightforward exposure fusion technique to enhance underexposed images by compensating for insufficient illumination. Their method utilizes a Laplacian pyramid-based multi-scale fusion approach to blend multiple images with different exposure levels, producing a well-exposed final image. Later, Ancuti and Ancuti (2013) extended this technique to the dehazing task by first generating two enhanced versions of the input image, one through white balance correction and the other via contrast enhancement. These two images were then fused multiple times using weight maps to produce a haze-free output. In work (Ancuti et al. 2012, 2018), a similar approach was adopted and applied to underwater image and video enhancement tasks. Furthermore, the authors introduced a temporal consistency filtering method to reduce noise in videos while preserving edge details. Instead of using multi-scale Laplacian fusion, Vasamsetti et al.

(2017) adopted discrete wavelet transform (DWT) to decompose images into low-frequency and high-frequency components. The low-frequency component represents global brightness and color information, while the high-frequency component captures edges and texture details. Additionally, contrast enhancement was achieved using Euler-Lagrange Variational Optimization, effectively preventing excessive sharpening.

Due to the lack of local contrast and color adjustment in white balance, Garg et al. (2018) applied CLAHE to enhance local contrast and used percentile stretching on pixel intensities to restore attenuated colors. Additionally, the authors blended the enhanced results from the RGB color space with those from the HSV color space to further improve color restoration. Ghani and Isa (2014); Abdul Ghani and Mat Isa (2015) used a similar process but additionally employed Rayleigh stretching to preserve edge details. In their subsequent work (Abdul Ghani and Mat Isa 2017), the authors iteratively adjusted the histogram of the V (value) channel toward a Rayleigh distribution to enhance brightness and contrast while preserving color relationships in the HSV color space.

3.2 Data-Driven Approaches

3.2.1 CNN-Based Methods

With the surge of deep learning, Convolutional Neural Networks (CNNs) have shown promise in underwater image restoration. Particularly when ground-truth or realistic synthetic data are available, supervised learning approaches are employed. Typical pipelines predict a per-pixel transmission map or color correction from raw underwater inputs in an end-to-end manner. Examples include: Li et al. (2020) constructed a dataset of aligned underwater/tank images and used a CNN to predict color-corrected outputs via channel-wise attenuations. Jiang et al. (2020) developed a multi-branch network with specialized modules for color balancing and detail preservation, supervised by a curated dataset of synthetic pairs. Recently, Rao et al. (2024) proposed an end-to-end framework, integrating a color compensation module with an enhancement module. The first module extracts features from brightness and colors separately and merges them back using Probabilistic Volume Aggregation with simple MLP layers. Ju et al. (2025) exploit a synthetic dataset to remove marine snow in the multi-scale Fourier domain using a simple U-Net architecture.

While these data-driven methods excel in typical underwater scenes, collecting fully paired ground truth is notoriously difficult. Hence, many rely on synthetic data generation pipelines (e.g., Blender, Unreal Engine) or weakly supervised setups. Li et al. (2018) introduced WaterGAN to synthesize underwater images for training restoration networks, enabling robust real-world inference. Similarly, Hou et al. (2020) compile extensive real datasets (U48, EUVP, UFO-120) with approximate reference targets, fostering deeper CNN training.

In many scenarios, image sequences are captured under extreme low-light conditions with high ISO noise. Fu et al. (2022) incorporate homology-based constraints for denoising and color balancing, whereas Liu et al. (2019) adopt a deep residual framework to handle both shot noise and scattering artifacts. Recent transformer-based architectures (Peng et al. 2023) further push performance, especially in extremely low signal-to-noise ratio conditions, by modeling global contexts across entire frames. Still, noise statistics vary drastically between different water types, raising open questions about generalization across diverse diving environments.

3.2.2 Transformer-Based Methods

Transformer-based UIE has evolved into a major branch of learning-based enhancement because self-attention is well suited to the spatially non-uniform degradations of underwater imagery. In contrast to purely convolutional models, hierarchical or windowed transformers can aggregate long-range contextual cues that stabilise global colour correction while still preserving local structures through multi-scale feature fusion. This is particularly valuable underwater, where attenuation, backscatter, turbidity, and illumination often vary across the same frame rather than acting as a single global degradation. Building on the broader success of vision transformers and Swin-style hierarchical attention (Liu et al. 2021), recent UIE methods increasingly use transformer blocks to jointly model local texture recovery and scene-level colour consistency.

One prominent line follows U-shape or hierarchical encoder–decoder designs that inject attention into otherwise U-Net-like enhancement pipelines. AutoEnhancer (Tang et al. 2022) explores this direction through transformer-on-U-Net architecture search, while U-shape Transformer (Peng et al. 2023) provides a representative end-to-end design for multi-scale feature aggregation and robust colour correction. UDAformer (Shen et al. 2023) strengthens this family with dual attention that combines channel self-attention and shifted-window pixel self-attention, and the adaptive group attention-based multiscale cascade transformer (Huang et al. 2022) further emphasises coordinated local–global restoration across scales. Related efforts such as underwater image enhancement using a pre-trained transformer (Boudiaf et al. 2022) and DAUT (Badran and Torki 2023) show how transfer from stronger transformer priors or depth-aware conditioning can further improve recovery in challenging scenes. Taken together, these methods establish the main transformer recipe in UIE: preserve the strong local-detail pathways of encoder–decoder restoration while using attention to better address non-uniform colour casts and contrast loss.

More recent work moves beyond generic hierarchical attention toward underwater-specific transformer modules. WaterFormer (Wen et al. 2024) is a representative example that combines global–local transformer reasoning with an explicit environment adaptor, allowing enhancement behaviour to be conditioned on prevailing water characteristics rather than assuming a single mapping for all domains. This design is especially relevant for robustness across different water types and turbidity levels. In parallel, UIE-Convformer (Wang et al. 2024) blends convolutional inductive biases with transformer reasoning, X-CAUNet (Pramanick et al. 2024) explicitly models cross-colour channel interactions, Phaseformer (Khan et al. 2025b) introduces phase-based attention to better recover structures under severe degradation, and the Globally Deformable Information Selection Transformer (Zhuang et al. 2024) highlights flexible feature selection under spatially heterogeneous underwater distortions. These architectures indicate a clear shift from simply importing generic vision transformers toward designing transformer blocks that reflect underwater colour coupling, multi-scale scattering effects, and environment-dependent degradation.

The transformer family is also expanding toward more task-aware and physics-aware variants. Reinforced Swin-Convts Transformer (Ren et al. 2022) couples enhancement with super-resolution for degraded sensing imagery, TAFormer (Li et al. 2025) injects transmission-aware priors into transformer restoration, Histoformer (Peng et al. 2025) uses histogram-guided modelling to correct global tone statistics efficiently, and UIE-SFIFormer (Zhou et al. 2025) combines physical guidance with spatial–frequency interaction. These extensions suggest that transformer-based UIE is no longer a single architectural template, but a spectrum of models that can incorporate transmission cues, colour-statistics priors, or physical guidance. At the same time, transformer-based enhancement remains computationally demanding and typically more data-hungry than lightweight CNN baselines. For practical deployment on low-power underwater platforms such as ROVs and AUVs, an important open issue is how to retain cross-domain robustness and stable restoration quality without incurring prohibitive memory, latency, or temporal-consistency costs.

From a physics-aware viewpoint, the attraction of transformers is not only larger receptive fields. Global attention helps relate distant regions that share the same water mass, illumination trend, or attenuation pattern, making it easier to correct spatially varying colour loss and backscatter than with purely local convolution alone. This is particularly useful when artificial lighting or turbidity affects only part of the frame. The trade-off is that more expressive global modelling can also alter low-level gradients and radiometric ordering more aggressively, so reconstruction-oriented use still benefits from structure-aware losses and cross-view consistency constraints.

From an engineering perspective, this usually places transformer-based UIE on the offline or high-end near-online side of the deployment spectrum. Reported latency, FLOPs, and memory vary too widely across datasets and hardware to support a fair numeric ranking, but the practical trend is that full transformer restoration is rarely the safest choice for tightly power-bounded SLAM or navigation front-ends.

3.2.3 Mamba-based Methods

Recently, state space models, known as ‘Mamba,’ have emerged as a linear alternative to Transformers (Gu and Dao 2024). Their relevance to underwater enhancement is not merely architectural novelty: they offer a practical way to capture long-range dependencies caused by spatially varying attenuation and illumination while avoiding the quadratic memory growth of full attention. This efficiency is potentially valuable for field systems that must process high-resolution frames under limited onboard memory. The Mamba blocks assembled in a UNet-like

architecture were first introduced by Ruan et al. (2024). MUIR (Chen et al. 2024) incorporates depth estimation into the framework. UWMamba (Guan et al. 2024) combines visual state space to capture long-range global features and convolution to capture local, detailed features. PixMamba (Lin et al. 2024) improves the clarity and sharpness of fine details by merging two branches: pixel-level and patch-level, both based on Mamba architectures. BEM (Huang et al. 2026) models the one-to-many mapping relations between input and targets by integrating vision Mamba into Bayesian neural networks. Relative to heavier transformers, these models offer a more attractive compromise between context modelling and efficiency, although most current implementations still target high-quality offline enhancement rather than tightly power-bounded AUV or ROV deployment.

Among global-context UIE families, compact Mamba variants are therefore one of the more plausible candidates for near-online robotic preprocessing. Even so, the current literature still reports them mainly in high-quality offline settings, and clearer platform-level reporting will be needed before practitioners can judge whether a given model is realistic for embedded underwater deployment.

3.2.4 Diffusion Model-based Methods

Diffusion models are among the most effective generative AI techniques, having demonstrated their ability to create realistic, high-resolution images and videos (Anantrasirchai et al. 2025). These techniques have also gained traction in underwater enhancement. The Denoising Diffusion Probabilistic Model (DDPM) was utilized in (Lu et al. 2023, 2024), which required training on paired datasets. This method has been adapted to a patch-based approach (Xia et al. 2025) to better capture local information and achieve higher resolution. Alternatively, UIEDP (Du et al. 2025) utilizes a pre-trained diffusion model to circumvent the scarcity of paired underwater datasets. This method integrates natural image priors where clean in-air images provide balanced color distributions and rich structural details, which can be transferred into underwater restoration priors. Similarly, BDMUIE (Chen et al. 2025) uses prior information from clean air images in Bayesian diffusion models, effectively merging top-down (prior) and bottom-up (data-driven) information distributions.

These models are attractive when degradation is severe because strong generative priors can recover missing contrast and plausible colour statistics. However, this same flexibility creates a tension with geometry-sensitive downstream use. Iterative denoising remains substantially heavier than one-pass CNN, Mamba, or GAN inference, and independently restored frames can drift in colour or local detail in ways that hurt key-point repeatability, pose estimation, or multi-view consistency. In practice, diffusion-based UIE is therefore better suited to offline post-processing, archival restoration, or human-facing visualisation than to real-time navigation, SLAM, or calibration-critical SfM stages unless additional structure, temporal, or cross-view constraints are imposed.

For the same reason, diffusion is currently better understood as an offline restoration or visualization tool than as a front-end for real-time robotic reconstruction. Without strong temporal or geometric constraints, per-frame diffusion enhancement can improve appearance while still changing the very cues that a SLAM, SfM, NeRF, or 3DGS pipeline depends on for stable optimization.

3.2.5 Learning Under Limited or No Paired Supervision

Paired clean references are particularly difficult to obtain underwater because the same scene is rarely captured with identical geometry, illumination, turbidity, and depth before and after degradation. Consequently, a large part of learning-based UIE has shifted toward settings with limited supervision, including unpaired translation, weak supervision, semi-supervised learning, and domain adaptation. We group these methods together because they address a shared problem: how to learn a robust restoration mapping when synthetic supervision is incomplete and real deployment conditions differ from the training distribution.

One major route relies on adversarial learning and cycle-consistent translation. WaterGAN (Li et al. 2018) synthesizes underwater imagery from in-air data to create realistic training pairs, while Fabbri et al. (2018) use GAN-based restoration to improve perceptual quality under real degradation. Domain-adversarial and unpaired translation frameworks such as Uplavikar et al. (2019) further reduce the gap between synthetic and real underwater domains, and cycle-consistent variants such as U-CycleGAN (Anwar and Li 2020), the multi-scale formulation of Desai et al. (2021), and the model-driven UW-CycleGAN (Yan et al. 2023) show how adversarial mapping can be

constrained to preserve colour trends and structure without requiring exact pixel-aligned references. Their main limitation is that global appearance transfer can still hallucinate textures or over-correct colours when the source and target domains only partially overlap.

Another line of work seeks to stabilise weakly supervised learning through representation regularisation rather than relying only on adversarial discrimination. Twin adversarial contrastive learning (Liu et al. 2022) uses contrastive consistency to retain structure during enhancement, while the perception-driven framework without paired supervision of Jiang et al. (2023) explicitly optimizes perceptual objectives when reference images are unavailable. Content-style disentanglement methods such as Zhu et al. (2023) and double-order contrastive formulations such as Yin et al. (2024) further separate degradations from scene content, aiming to reduce hallucination and make no-paired-supervision training less brittle. Compared with plain GAN translation, these models generally place more emphasis on preserving semantic layout and suppressing over-enhancement.

A third branch explicitly targets semi-supervised and domain-adaptive transfer. The two-step framework of Jiang et al. (2022) decomposes degradation adaptation and enhancement refinement, while Chen and Pei (2022) and Bing et al. (2023) adapt features across domains through content-style separation or in-air-to-underwater transfer. Semi-UIR (Huang et al. 2023) introduces contrastive semi-supervised learning with a reliable bank to stabilise optimisation on real data, and Wen et al. (2025) further combine synthetic supervision with cross-domain feature alignment for real-world UIE. These methods are especially important underwater because synthetic data remain useful for controllable supervision, yet model performance often collapses unless the training objective also adapts to the statistics of real water types and turbidity conditions.

Domain-adversarial learning provides a particularly direct solution path for this transfer problem. In addition to early domain-adversarial formulations such as Uplavikar et al. (2019), Kapoor et al. (2023) show that aligning latent features across source and target domains during enhancement training can improve real-world robustness without assuming abundant paired underwater annotations.

Taken together, these methods show that underwater UIE increasingly depends not only on the backbone architecture, but also on how supervision is constructed when exact references are absent. Some recent approaches go one step further by coupling limited-supervision learning with explicit physical or model-based components; we discuss those physics-guided designs separately in subsection 3.3 to avoid conflating supervision strategy with physical/model coupling.

Table 3 complements the taxonomy in Figure 7 by summarising the data-driven UIE methods discussed in Sec. 3.2 in chronological order, making the progression from early GAN/CNN models to transformer-, Mamba-, diffusion-, and domain-adaptive methods more explicit.

Table 3: Chronological summary of data-driven underwater image enhancement methods, highlighting the shift from early deep-learning models toward more recent transformer-, Mamba-, diffusion-, and domain-adaptive approaches.

Method	Family / Setting	Contribution	Remarks
Li et al. (2018)	GAN / synthetic paired	Synthesizes realistic underwater imagery from in-air data to create restoration training pairs for subsequent enhancement models.	Data generation; colour correction
Fabbri et al. (2018)	GAN / perceptual enhancement	Uses generative adversarial restoration to improve the perceptual quality of real degraded underwater images.	Perceptual restoration
Liu et al. (2019)	CNN / residual	Applies a deep residual framework to suppress noise and scattering artefacts while recovering colour and contrast.	Denoising + enhancement
Uplavikar et al. (2019)	GAN / domain-adversarial	Learns domain-invariant enhancement across diverse water types through domain-adversarial training.	Cross-domain robustness
Li et al. (2020)	CNN / supervised	Introduces the UIEB benchmark and a benchmark-driven supervised enhancement framework built from candidate restorations and learned fusion.	Benchmark-driven UIE
Jiang et al. (2020)	CNN / supervised	Develops a multi-branch network for coordinated colour balancing and detail preservation using curated paired supervision.	Colour-detail balance
Anwar and Li (2020)	Cycle-consistent unpaired	Represents the paired-free cycle-consistent restoration line that constrains translation without exact pixel-aligned references.	Unpaired restoration

Continued on next page

Table 3 – Continued from previous page

Method	Family / Setting	Contribution	Remarks
Desai et al. (2021)	GAN / generation	Generates realistic underwater imagery to narrow the gap between synthetic training data and real degradations.	Data generation; domain bridging
Fu et al. (2022)	CNN / unsupervised	Uses homology-based constraints to regularize denoising and colour balancing without paired supervision.	Severe turbidity; overlaps with hybrid discussion
Tang et al. (2022)	Transformer / NAS	Searches transformer-on-U-Net enhancement architectures automatically to improve restoration design under underwater degradation.	Architecture search
Huang et al. (2022)	Transformer / multi-scale	Uses adaptive group attention and a multiscale cascade design for coordinated local-global enhancement.	Multi-scale attention
Boudiaf et al. (2022)	Transformer / pre-trained	Transfers a pre-trained transformer prior into underwater enhancement to improve restoration under difficult conditions.	Transfer learning
Ren et al. (2022)	Transformer / task-aware	Couples enhancement with super-resolution in a reinforced Swin-Convts transformer for degraded sensing imagery.	Joint enhancement + SR
Liu et al. (2022)	Contrastive / weak supervision	Introduces twin adversarial contrastive learning to preserve structure during paired-free enhancement.	Contrastive consistency
Jiang et al. (2022)	Semi-supervised / domain-adaptive	Decomposes degradation adaptation and enhancement refinement into a two-step synthetic-to-real transfer framework.	Two-step transfer
Chen and Pei (2022)	Domain adaptation	Separates content and style to adapt enhancement models across domains with differing underwater appearance statistics.	Content-style separation
Peng et al. (2023)	Transformer / U-shape	Provides an end-to-end U-shape transformer with strong multi-scale aggregation and robust colour correction.	Overall enhancement
Shen et al. (2023)	Transformer / dual-attention	Combines channel self-attention and shifted-window pixel attention to improve underwater restoration.	Dual attention
Badran and Turki (2023)	Transformer / depth-aware	Injects depth-aware conditioning into a U-shape transformer to better handle scene-dependent degradation.	Depth-aware conditioning
Lu et al. (2023)	Diffusion / paired	Adapts DDPM to underwater restoration with paired supervision for high-fidelity generative enhancement.	Paired-data dependent
Yan et al. (2023)	GAN / model-driven	Constrains CycleGAN-style restoration with model-driven colour and structure cues to reduce artefacts.	Unpaired restoration
Jiang et al. (2023)	Weak supervision / perceptual	Optimizes perceptual objectives for underwater enhancement without requiring paired references.	Perception-driven
Zhu et al. (2023)	Disentanglement / unpaired	Separates scene content and degradation style to stabilise unsupervised underwater enhancement.	Content-style disentanglement
Bing et al. (2023)	Domain adaptation / transfer	Transfers in-air priors to underwater enhancement through deep domain adaptation.	In-air to underwater
Huang et al. (2023)	Semi-supervised	Uses a reliable-bank contrastive strategy to stabilise semi-supervised underwater restoration on real data.	Reliable bank
Kapoor et al. (2023)	Domain-adversarial	Aligns latent source and target features during training to improve real-world UIE robustness.	Cross-domain robustness
Rao et al. (2024)	CNN / end-to-end	Integrates colour compensation with enhancement through coupled feature extraction and aggregation.	Generalised colour compensation
Wen et al. (2024)	Transformer / environment-adaptive	Couples global-local transformer reasoning with an environment adaptor to condition restoration on water characteristics.	Water-type adaptation
Wang et al. (2024)	Transformer / convformer	Blends convolutional inductive bias with transformer reasoning for balanced local and global restoration.	Conv-transformer hybrid
Pramanick et al. (2024)	Transformer / channel-aware	Explicitly models cross-colour channel interactions for underwater enhancement.	Colour coupling
Zhuang et al. (2024)	Transformer / deformable selection	Uses globally deformable information selection to adapt feature usage under spatially heterogeneous distortions.	Flexible feature selection
Chen et al. (2024)	Mamba / depth-aware	Extends Mamba-based restoration with integrated depth estimation to improve underwater enhancement.	Depth-aware Mamba
Guan et al. (2024)	Mamba / visual state space	Combines state-space global modelling with local convolutions for underwater enhancement.	Long-range + local detail
Lin et al. (2024)	Mamba / dual-branch	Uses pixel-level and patch-level Mamba branches to improve clarity and fine-detail recovery.	Sharpness/detail recovery
Lu et al. (2024)	Diffusion / efficient	Speeds up DDPM-based underwater enhancement toward more practical real-time deployment.	Real-time oriented
Yin et al. (2024)	Contrastive / disentanglement	Uses double-order contrastive disentanglement to improve unpaired underwater enhancement.	Stronger regularisation

Continued on next page

Table 3 – Continued from previous page

Method	Family / Setting	Contribution	Remarks
Ju et al. (2025)	CNN / synthetic marine snow	Removes marine snow in the multi-scale Fourier domain using synthetic supervision and a lightweight U-Net back-bone.	Marine snow removal
Khan et al. (2025b)	Transformer phase-aware	/ Introduces phase-based attention to recover structures under severe underwater degradation.	Structure recovery
Li et al. (2025)	Transformer physics-aware	/ Injects transmission-aware priors into transformer-based restoration.	Transmission-aware
Peng et al. (2025)	Transformer histogram-guided	/ Uses histogram modelling to improve efficient global tone correction in underwater images.	Tone statistics
Zhou et al. (2025)	Transformer spatial-frequency	/ Combines physical guidance with spatial-frequency interaction for underwater enhancement.	Physics-aware transformer
Xia et al. (2025)	Diffusion / patch-based	Uses patch-based diffusion to capture local details more accurately at higher resolution.	High-resolution local detail
Du et al. (2025)	Diffusion / prior-guided	Transfers diffusion priors from clean natural images to reduce dependence on paired underwater training data.	Diffusion prior
Chen et al. (2025)	Diffusion Bayesian	/ Combines air-image priors with Bayesian diffusion to merge top-down and bottom-up restoration cues.	Top-down + bottom-up
Wen et al. (2025)	Semi-supervised domain-adaptive	/ Combines synthetic supervision with cross-domain feature alignment for real-world underwater enhancement.	Real-world transfer
Huang et al. (2026)	Mamba / Bayesian	Models one-to-many enhancement mappings with Bayesian neural networks and Vision Mamba.	Uncertainty-aware

3.3 Hybrid Approaches

Hybrid approaches in underwater UIE explicitly learned restoration with underwater imaging knowledge, i.e., physics-guided UIE. Whereas Sec. 3.2.5 focuses on how enhancement models learn when paired references are scarce, this subsection focuses on what is coupled into the model: underwater IFMs, transmission or background-light estimation, scene priors, decomposition constraints, and controllable synthetic-real training mechanisms. The common goal is to move beyond a purely black-box mapping and instead embed medium-aware structure into the restoration process.

One major line couples networks with scene priors or physically motivated restoration terms. Underwater scene prior inspired enhancement (Li et al. 2020) uses scene priors to guide restoration across images and videos, while perceptual underwater image enhancement with deep learning and physical priors (Chen et al. 2020) combines learned enhancement with physically informed constraints on color and structure. Related designs such as the generalized physical-knowledge-guided dynamic model (Mu et al. 2023) and IPMGAN (Liu et al. 2021) further show how physical priors can be injected into the objective, network dynamics, or adversarial training process. Across these methods, the hybrid element lies in reintroducing underwater imaging knowledge into optimization rather than relying solely on a data-driven image-to-image transform.

A second line emphasizes decomposition and explicitly constrained architectures. Instead of predicting an enhanced image directly, these methods decompose the problem into physically meaningful components such as transmission, background light, or scene radiance, then optimize them jointly. The homology-driven framework of Fu et al. (2022) illustrates how domain knowledge can regularize unsupervised restoration under severe turbidity, while Mamba-UIE (Zhang et al. 2024) uses a physically constrained multi-branch formulation to estimate medium-related variables and scene content together. Hybrur (Yan et al. 2023) likewise combines physical modeling with neural restoration in an unsupervised setting, and wavelet-guided designs such as Zhang et al. (2025) show that hybrid coupling can also occur through reflectance or frequency-domain priors. These methods make clear that hybrid UIE is not only about mixing data sources; it also includes model-level and objective-level coupling.

A third line focuses on synthetic-real bridging under physical guidance. Here the aim is to exploit controllable degradations from simulation or physical modeling while still anchoring training to real underwater imagery. SyreaNet (Wen et al. 2023) is a representative example of this direction rather than its sole defining method: it uses physically controllable synthetic degradations together with real images to reduce the gap between modeled water conditions and deployment scenes. In this sense, synthetic-real bridging complements the limited/no-paired-supervision strategies in Sec. 3.2.5, but we discuss it here because the defining feature is not merely the supervision regime; it is the explicit physical or model-coupled design that governs how supervision is constructed. Across these

Table 4 Summary of publicly available underwater image enhancement and restoration datasets with their key properties. “# Images” gives the approximate number of samples. “Real/Synth.” specifies whether data are captured in real water or generated synthetically via simulation/rendering. “Paired?” indicates whether ground-truth/reference images are provided

Dataset	Year	# Images	Real/Synth.	Paired?	Download
UIEB (Li et al. 2019)	2019	890	Real	Approx. paired	Link
EUVP (Islam et al. 2020a)	2020	4,414 real + 6,850 syn.	Both	Paired (syn.) / Unpaired (real)	Link
U45 (Li et al. 2019)	2019	45	Real	Unpaired	Link
UFO-120 (Islam et al. 2020)	2020	1,550	Real	Unpaired	Link
RUIE (Liu et al. 2020)	2020	4230	Real	Unpaired	Link
WaterGAN (Li et al. 2018)	2017	10K+ (synthetic)	Synthetic	Paired (rendered pairs)	Link
Sea-Thru (Akkaynak and Treibitz 2019a)	2019	1,114	Real	Partial (depth-registered)	Link
SQUID (Berman et al. 2020)	2020	Various	Real	Unpaired	Link
OceanDark (Marques and Albu 2020)	2020	183	Real	Unpaired	Link
LNRUD (Ye et al. 2022)	2022	50,000	Synthetic	Paired	Link
LSUI (Peng et al. 2023)	2023	4,279	Real	Paired	Link
UVEB (Xie et al. 2024)	2024	453,000	Real	Paired	Link

strands, physics-guided and model-coupled UIE offers a promising route to better cross-domain stability (Li et al. 2020; Hou et al. 2020), although its effectiveness still depends on how well the chosen priors and decomposition assumptions match real underwater conditions.

3.4 Evaluation and Benchmark

A key challenge in underwater enhancement is the scarcity of reliable ground-truth pairs. Several datasets attempt to mitigate this issue, including UIEB (Li et al. 2020), EUVP (Islam et al. 2020a), UVEB (Xie et al. 2024), and LSUI (Peng et al. 2023). However, in UIEB the pairs are not fully registered, which introduces misalignment between degraded inputs and reference images, thereby limiting the reliability of supervised learning.

The EUVP dataset (Islam et al. 2020a) provides over 12K paired and 8K unpaired samples, where the paired references are generated using CycleGAN. U45 (Hou et al. 2020) serves as a test set of 45 real-world underwater images degraded by color casts, low contrast, and haze-like effects. OceanDark (Marques and Albu 2020) is composed of 183 underwater images of 1280 x 720 pixels captured by video cameras located in profound depths using artificial lighting. LNRUD (Ye et al. 2022) contains 50000 clean images and 50000 corresponding underwater images synthesized from 5000 real underwater scene images. A comprehensive dataset summary is provided in Table 4.

For benchmarking and training design, it is useful to distinguish three dataset regimes. First, *fully synthetic* resources such as WaterGAN and LNRUD offer controllable degradations and clean supervision, but inevitably simplify real underwater variability. Second, *pseudo-paired or model-based generated* resources use physics-inspired generation or restoration targets to construct approximate pairs; representative examples include the scene-prior-driven framework of Li et al. (2020) and the synthetic-real bridging strategies discussed in Section 3.3. Third, *real unpaired or approximately paired* collections such as UIEB, EUVP, Sea-Thru, LSUI, and UVEB better reflect deployment conditions, but often trade away exact correspondence or strict radiometric ground truth. This distinction matters because reported gains are often tied as much to the supervision regime as to the model architecture itself. While these resources have significantly advanced the field, the lack of standardized evaluation protocols and the inherent variability of underwater conditions continue to impede fair comparison across methods. For a comprehensive benchmark of deep learning-based approaches, we refer readers to Cong et al. (2021).

For readers interested in task-driven evaluation beyond enhancement quality, Fish4Knowledge (Spampinato et al. 2014) and the Underwater Change Detection Dataset (Radolko et al. 2016) provide useful complementary resources for fish monitoring, behaviour analysis, change detection, and moving-object segmentation. These datasets are not enhancement benchmarks in the strict sense, but they help illustrate whether an enhancement method preserves motion cues, target boundaries, and background stability well enough for adjacent video-analysis settings.

3.4.1 Evaluation Metrics

The assessment of image enhancement quality is non-trivial, particularly in underwater scenarios where distortions are complex and subjective human perception often diverges from pixel-wise fidelity measures. In this section, we review both general-purpose image quality metrics and underwater-specific evaluation criteria.

- **MSE and PSNR.** The most widely used distortion-based criteria are Mean Squared Error (MSE) and its logarithmic counterpart, Peak Signal-to-Noise Ratio (PSNR). MSE computes the average squared difference between a reference image x and a test image y , expressed as

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2, \quad (15)$$

where N denotes the number of pixels. PSNR is subsequently derived as

$$\text{PSNR} = 10 \log_{10} \frac{L^2}{\text{MSE}}, \quad (16)$$

with L representing the dynamic range of pixel intensities (*e.g.*, 255 for 8-bit images). Despite their mathematical simplicity, clear interpretation, and popularity in optimization contexts, these measures exhibit weak correlation with human visual perception (Wang and Bovik 2009). They treat all pixel errors equally and ignore spatial structures, making them inadequate for perceptual quality assessment.

- **SSIM.** Wang and Bovik (2002) introduced the Structural Similarity Index (SSIM), which evaluates luminance, contrast, and structural similarity between local patches. Formally, SSIM is defined as

$$\begin{aligned} \text{SSIM}(x, y) &= l(x, y) \cdot c(x, y) \cdot s(x, y) \\ &= \left(\frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \right) \left(\frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \right) \left(\frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \right), \end{aligned} \quad (17)$$

where μ , σ , and σ_{xy} denote local means, standard deviations, and cross-covariance, respectively. Constants C_1 , C_2 , and C_3 stabilize the computation. SSIM has been demonstrated to align more closely with the human visual system by emphasizing structural integrity rather than pixel accuracy.

- **PCQI.** The Patch-based Contrast Quality Index (PCQI) (Wang et al. 2015) extends perceptual assessment by explicitly modeling three independent components within local patches: mean intensity, contrast change, and structural distortion. It is formulated as

$$\text{PCQI} = q_i(x, y) \cdot q_c(x, y) \cdot q_s(x, y), \quad (18)$$

where q_i , q_c , and q_s quantify luminance, contrast, and structure fidelity, respectively. Compared with global metrics, PCQI better reflects local contrast variations but incurs higher computational complexity.

- **UCIQE.** For underwater imagery, Yang and Sowmya (2015) proposed the Underwater Color Image Quality Evaluation (UCIQE) index, which operates in the perceptually uniform CIELab color space. UCIQE linearly combines chroma dispersion, luminance contrast, and average saturation as

$$\text{UCIQE} = C_1\sigma_c + C_2\text{con}_l + C_3\mu_s, \quad (19)$$

where σ_c is the standard deviation of chroma, con_l the luminance contrast, and μ_s the mean saturation. UCIQE is widely used because it can be computed without reference images and captures several degradations that are

common underwater; however, like other no-reference metrics, it reflects only a limited proxy of perceptual quality and should not be interpreted as a substitute for reference-based fidelity measures when reliable ground truth is available. Its components can also reward over-saturated or contrast-stretched outputs, so a high UCIQE score does not necessarily imply physically plausible colour recovery or stable multi-view geometry.

- **UIQM.** Another underwater-specific measure, the Underwater Image Quality Measure (UIQM) (Panetta et al. 2015), explicitly incorporates aspects of the human visual system (HVS) without requiring ground-truth images. UIQM integrates three components: the Underwater Image Colorfulness Measure (UICM), the Underwater Image Sharpness Measure (UISM), and the Underwater Image Contrast Measure (UIConM):

$$\text{UIQM} = c_1 \cdot \text{UICM} + c_2 \cdot \text{UISM} + c_3 \cdot \text{UIConM}, \quad (20)$$

where the coefficients c_1 , c_2 , and c_3 can be adjusted depending on whether color fidelity, sharpness, or contrast is prioritized. UIQM is practical when no reference image exists, but it can still reward over-sharpening or colour shifts that are undesirable for restoration fidelity or downstream vision tasks. This is especially important for robotics and reconstruction, where inflated sharpness or colourfulness scores can coincide with degraded feature repeatability, altered radiometry, or unstable correspondences.

- **UIF.** More recently, Zhou et al. (2023) proposed the Underwater Image Fidelity (UIF) index, which aims to measure color consistency in underwater scenes by analysing the distribution of pixel intensities in multiple sub-interval histograms. Specifically, UIF divides the chroma channel into K intervals and computes the fidelity by evaluating deviations between the reference-free enhanced image and an idealized uniform distribution. Formally, UIF can be expressed as

$$\text{UIF} = \sum_{k=1}^K w_k \cdot \left(1 - \frac{|h_k - \bar{h}_k|}{\bar{h}_k + \epsilon} \right), \quad (21)$$

where h_k is the observed histogram value in the k^{th} sub-interval, $\bar{h}_k$ is the expected reference value under an idealized distribution, w_k is a weighting factor reflecting the perceptual importance of the interval, and ϵ is a small constant to prevent division by zero.

By quantifying chroma distribution consistency, UIF provides a more fine-grained evaluation of underwater color distortions than global statistics such as UCIQE or UIQM. Nevertheless, the correlation with subjective human perception is still imperfect, highlighting the need for task-driven or perceptual-learning-based evaluation frameworks in underwater image enhancement.

- **Learned underwater IQA and ranking.** Recent work has started to replace hand-crafted no-reference criteria with learned ranking models. Underwater Ranker (Guo et al. 2023), for example, learns pairwise preferences over underwater image quality and can be used both to compare outputs and to guide enhancement. This direction is important because subjective preference, restoration fidelity, and downstream task utility are not equivalent objectives: an image preferred by a human observer may still alter geometric cues, detection boundaries, or colour constancy in ways that hurt analysis. Learned ranking therefore complements, rather than replaces, full-reference metrics, classical no-reference metrics, and task-driven evaluation.
- **Task-driven geometry and robotics metrics.** For robotics and 3D reconstruction, enhancement quality should also be judged by whether the restored images preserve the cues needed by downstream geometry pipelines. Relevant indicators include feature repeatability and inlier matching rates, track length, reprojection or pose-estimation error, and downstream SLAM robustness or drift. These measures are often more meaningful than UIQM or UCIQE when the target application is SfM, visual SLAM, COLMAP, or neural rendering, because they test whether enhancement helps or harms correspondence stability and geometric consistency rather than only whether the image appears vivid (Malyugina et al. 2025; Vrochidis et al. 2025a).

Table 5 Summary of commonly used evaluation metrics for image enhancement

Metric	Reference Required	Perceptual-inspired	Underwater-specific
MSE / PSNR	✓	✗	✗
SSIM	✓	✓	✗
PCQI	✓	✓	✗
UCIQE	✗	✓	✓
UIQM	✗	✓	✓
UIF	✗	✓	✓
Underwater Ranker	✗	✓	✓

For joint enhancement–reconstruction pipelines, geometry-aware reporting should go further still. Useful targets include Chamfer Distance or Cloud-to-Cloud (C2C) distance for reconstructed geometry, Absolute Trajectory Error (ATE) or Relative Pose Error (RPE) for camera-motion quality, and track-stability or reprojection statistics for correspondence reliability. The current literature rarely reports these together with classical UIE metrics, which makes cross-domain assessment difficult and should itself be regarded as a major open challenge.

In summary, different evaluation metrics serve complementary purposes depending on the application. Full-reference measures such as PSNR, SSIM, and PCQI remain more reliable whenever trustworthy aligned references exist, because they directly quantify fidelity to a target image. No-reference measures such as UCIQE, UIQM, UIF, or learned underwater rankers remain useful in real underwater deployments where references are unavailable. However, neither family is sufficient on its own: reference metrics can miss human preference and task relevance, while no-reference metrics can reward visually pleasing but geometrically or semantically harmful alterations. [Table 5](#) therefore should be read as a toolbox rather than a leaderboard, and later benchmarking results should be interpreted together with downstream utility and cross-domain robustness.

For reconstruction-oriented settings, that toolbox should explicitly include task-driven reporting. Feature-matching success, pose accuracy, correspondence stability across views, and SLAM robustness can reveal failure modes that no-reference image-quality scores miss entirely, especially when colour vividness is achieved at the expense of geometric consistency.

3.4.2 Performance Evaluation of various Enhancement methods

We summarize the performance of recent enhancement methods across several benchmark datasets and evaluation metrics. As shown in [Table 6](#), the models are grouped into deterministic and generative categories, and their quantitative results are reported on the UIEB-R90, UIEB-C60, U45, and UCCS test sets using PSNR, SSIM, UIQM, and UCIQE. All values are taken from the original publications rather than reproduced here under one unified experimental protocol.

With the rapid progress of generative modeling in recent years and its broad application in image-related tasks, generative approaches have increasingly shown advantages over deterministic models. In particular, generative methods tend to achieve stronger perceptual quality, as reflected by consistently competitive or superior performance on perceptual metrics such as UIQM and UCIQE across multiple datasets. This trend is especially evident on challenging benchmarks (e.g., UIEB-C60 and U45), where diffusion- and uncertainty-aware models demonstrate a clear ability to enhance global appearance and color fidelity.

At the same time, these metric trends should be interpreted cautiously. Strong UIQM or UCIQE scores do not by themselves imply better reference fidelity, temporal stability, or better suitability for downstream analysis, which is why the quantitative comparisons in [Table 6](#) are best read together with the task-oriented discussion in [Section 3.5](#).

In contrast, deterministic models generally perform well on distortion-based metrics such as PSNR and SSIM, especially on relatively constrained datasets. However, their performance on perceptual metrics is often less consistent, suggesting limitations in modeling the inherent ambiguity of underwater image degradation. Overall, these results highlight the growing importance of generative modeling for underwater image enhancement, particularly in scenarios where perceptual quality and visual realism are critical.

Table 6 Quantitative comparisons on the UIEB-R90, UIEB-C60, U45, and UCCS datasets in terms of PSNR, SSIM, UIQM, and UCIQE. For each column, the best, second-best, and third-best results across all methods are highlighted with **red**, **orange**, and **yellow** backgrounds, respectively. A dash ‘-’ indicates that the corresponding method did not report this metric in the original publication.

Method	UIEB-R90			UIEB-C60			U45			UCCS		
	PSNR $\uparrow$	SSIM $\uparrow$	UIQM $\uparrow$	UIQM $\uparrow$	UCIQE $\uparrow$	UCIQE $\uparrow$	UIQM $\uparrow$	UCIQE $\uparrow$	UCIQE $\uparrow$	UIQM $\uparrow$	UCIQE $\uparrow$	UCIQE $\uparrow$
Deterministic Models												
WaterNet (Li et al. 2019)	21.04	0.860	2.399	2.399	0.591	–	–	–	–	2.275	0.556	0.556
Ucolor (Li et al. 2021b)	20.13	0.877	2.482	2.482	0.553	3.148	3.148	0.586	0.586	3.019	0.550	0.550
FiveA+ (Jiang et al. 2023)	23.06	0.911	–	–	–	–	–	–	–	–	–	–
HCLR-Net (Zhou et al. 2024)	24.99	0.925	2.695	2.695	0.587	3.103	3.103	0.610	0.610	3.045	0.579	0.579
Restormer (Zamir et al. 2022)	23.82	0.903	2.688	2.688	0.572	3.097	3.097	0.600	0.600	2.981	0.542	0.542
CECF (Cong et al. 2024)	21.82	0.894	–	–	–	–	–	–	–	–	–	–
U-Shape (Peng et al. 2023)	20.39	0.803	2.730	2.730	0.560	3.151	3.151	0.592	0.592	–	–	–
PixMamba (Lin et al. 2024)	23.59	0.921	2.868	2.868	0.586	–	–	–	–	3.053	0.561	0.561
WaterMamba (Guan et al. 2024)	24.72	0.931	2.835	2.835	0.582	–	–	–	–	3.057	0.555	0.555
X-CAUNET (Pramanick et al. 2024)	22.30	0.908	2.683	2.683	0.564	–	–	–	–	2.922	0.541	0.541
Convformer (Wang et al. 2024)	23.13	0.904	2.684	2.684	0.572	–	–	–	–	2.946	0.555	0.555
WF12-Net (Zhao et al. 2024)	23.86	0.873	–	–	–	3.181	3.181	0.619	0.619	–	–	–
Phaseformer (Khan et al. 2025b)	25.98	0.928	–	–	–	4.491	4.491	–	–	–	–	–
GS-Transformer (Zhuang et al. 2024)	24.42	0.861	–	–	–	–	–	–	–	–	–	–
Generative Models												
UIE-DM (Tang et al. 2023)	23.03	0.910	–	–	–	–	–	–	–	–	–	–
FUnIEGAN (Islam et al. 2020b)	19.12	0.832	2.867	2.867	0.556	2.495	2.495	0.545	0.545	3.095	0.529	0.529
PUGAN (Cong et al. 2023)	22.65	0.902	2.652	2.652	0.566	–	–	–	–	2.977	0.536	0.536
PUIE-MP (Fu et al. 2022)	21.05	0.854	2.524	2.524	0.561	3.169	3.169	0.569	0.569	2.758	0.489	0.489
Semi-UJR (Huang et al. 2023)	22.79	0.909	2.667	2.667	0.574	3.185	3.185	0.606	0.606	3.079	0.554	0.554
DCGF (Zhang et al. 2024)	–	–	–	–	–	–	–	–	–	1.377	0.609	0.609
DiffWater (Guan et al. 2023)	20.97	0.895	4.655	4.655	0.433	4.730	4.730	0.462	0.462	–	–	–
BEM (Huang et al. 2026)	25.62	0.940	2.931	2.931	0.567	3.406	3.406	0.620	0.620	3.224	0.561	0.561

3.5 Discussion and Open Challenges

This subsection focuses on bottlenecks that remain specific to underwater image enhancement. Shared issues such as benchmark design, supervision scarcity across videos and multi-view data, and broader evaluation strategy are synthesized later in [subsection 6.1](#) and [subsection 6.3](#).

Generalization Across Water Types and Depth Ranges. Many learning-based models are still tuned to a narrow distribution of water optics or illumination conditions, partly because supervised training often optimizes a one-to-one mapping tied to a specific dataset. However, recent methods now treat this issue as a concrete methodological direction rather than only an open challenge. As discussed in [subsubsection 3.2.5](#), learning under limited or no paired supervision uses synthetic-to-real transfer, reliable-bank or pseudo-label regularization, consistency constraints, and feature alignment to reduce performance drops across water types and turbidity levels. Even so, robust generalization under severe shifts in depth, particulate concentration, and illumination remains difficult, especially when appearance correction must transfer across water types without disturbing structure that later processing depends on.

The mismatch becomes especially severe when models trained on synthetic or controlled-tank data are deployed in field conditions with strong turbidity, spatially non-uniform caustics, drifting marine snow, or unstable artificial lighting. These effects are hard to approximate with idealized water models, yet they strongly influence whether restored images remain useful for later multi-view matching and reconstruction.

Low-Light, Turbulence, and Temporal Stability. Severe noise, forward scattering, and non-stationary particulate clutter remain central UIE-specific bottlenecks, especially in deep or murky water. Methods that improve single-frame contrast may still introduce flicker, unstable colour correction, or detail hallucination when applied to video. Lightweight restoration that jointly improves low-light visibility, suppresses turbulence-related artefacts, and preserves frame-to-frame consistency is therefore still needed for autonomous underwater vehicles (AUVs), ROVs, and other resource-constrained deployments.

Task-Oriented Utility of Enhancement. Enhancement utility depends on the intended task, not only on visual appeal. In operational settings such as underwater surveillance, inspection, teleoperation, and photogrammetry, low contrast, colour cast, and backscatter can obscure small targets, destabilise background models, and weaken situational awareness or measurement reliability ([Rout et al. 2024](#); [Garcia-Garcia et al. 2020](#)). Related application studies likewise show that image-analysis pipelines benefit when enhancement or pre-processing recovers discriminative structure rather than merely producing vivid colours ([Helan et al. 2006](#); [Bazeille et al. 2006](#)), while more recent reviews of underwater object detection underline the importance of dataset bias, degraded visibility, and task-oriented robustness in practical machine-processing settings ([Fu et al. 2023](#)). Motion-analysis methods are especially sensitive to temporal consistency, boundary preservation, and stable background modelling under flicker and dynamic underwater clutter ([Nissar et al. 2026](#); [Kapoor et al. 2024, 2025](#)), and tracking or long-duration surveillance makes similar demands on local visibility and temporal stability ([Zhang and Palaoag 2024](#); [Humbert et al. 2023](#)).

For this review, the main implication is that enhancement methods intended for machine processing should preserve structure, radiometric consistency, and temporal stability more carefully than methods aimed mainly at perceptual vividness. These properties are also the ones most likely to benefit reconstruction-oriented pipelines, where reliable correspondences, stable geometry-relevant cues, and frame-to-frame consistency matter directly. Broader questions of dataset coverage and evaluation for such machine-oriented settings are revisited in [subsection 6.1](#) and [subsection 6.3](#). Readers interested in wider downstream underwater surveillance or motion-analysis settings may also refer to the surveys of [Rout et al. \(2024\)](#) and [Garcia-Garcia et al. \(2020\)](#).

Perceptual Quality, Geometric Consistency, and Deployment Constraints. For reconstruction-oriented pipelines, better-looking enhancement is not automatically better geometry. Mild physical-model correction or conservative structure-preserving restoration often perturbs local gradients and cross-view radiometry less than aggressive black-box generation, whereas GAN- and diffusion-style enhancement can improve visibility while also hallucinating textures, shifting colours, or changing contrast in ways that destabilise feature matching, bundle adjustment, or neural-rendering supervision ([Malyugina et al. 2025](#); [Vrochidis et al. 2025a](#)). Lightweight CNN and Mamba-style restorers often lie between these extremes: they can improve contrast and denoising with less stylistic drift, but they still require structure-aware losses if the output is later consumed by SfM, SLAM,

Table 7 Geometry-consistency and deployment implications of major underwater enhancement families.

Family	Typical effect on geometry / key-points	Typical compute / memory	Deployment suitability	Typical best-use scenario
IFM-based / mild physics-guided correction	Usually preserves edges and radiometric ordering best when correction is conservative; lowest risk for calibration-sensitive stages	Low to moderate	Front-end or offline	Preprocessing for SfM, SLAM, pose estimation, metric reconstruction
Lightweight CNN / residual CNN	Often improves contrast and denoising with limited hallucination; geometry impact usually moderate and controllable	Low to moderate	Front-end / near-online feasible for compact models	Navigation support, inspection, operator assistance, robust preprocessing
Transformer-based UIE	Strong global colour correction for spatially varying attenuation, but can alter local gradients if trained aggressively	High memory and latency	Mostly offline; near-online only on high-end GPUs	High-quality offline enhancement across variable water conditions
Mamba-based UIE	Better efficiency than full attention with useful long-range context; geometry preservation depends on how strongly appearance is remapped	Moderate	Near-online for compact variants; otherwise offline or mixed	A compromise between context modelling and deployment cost
GAN / unpaired translation	Most prone to style drift, hallucinated texture, or cross-view colour inconsistency despite strong perceptual gains	Moderate inference; training is heavy	Mostly offline or operator-facing unless explicitly lightweight	Visual inspection, perceptual enhancement, domain translation
Diffusion-based UIE	Can recover severe degradations, but iterative inference and generative priors make geometry-relevant cues less predictable if unconstrained	High to very high	Offline only in current practice	Challenging restoration, archival enhancement, human-facing outputs

COLMAP, or pose estimation. In calibration-critical stages, raw images or only lightly corrected images combined with explicit physical or refractive camera models may therefore be preferable to strong learned enhancement; stronger restoration is often safer later in offline dense reconstruction, novel-view rendering, or operator-facing visualisation (Wright et al. 2020; Sedlazeck et al. 2009; Vrochidis et al. 2025b).

This is a genuine negative-transfer risk rather than a minor caveat. A method may detect more corner-like responses or produce visually sharper frames while still reducing the number of repeatable, view-consistent correspondences that survive geometric verification across time and viewpoint changes. In other words, single-view perceptual gain and multi-view geometric reliability are related but not interchangeable objectives.

Exact FLOPs, latency, and memory are reported inconsistently across underwater enhancement papers and hardware settings. Table 7 therefore summarizes typical operational tendencies rather than a strict cross-paper benchmark.

4 3D Reconstruction for Underwater Scenes

Underwater 3D reconstruction is challenged by scattering, absorption, refraction at the housing interface, low-contrast imagery, illumination inconsistency, and platform motion. Together, these factors make feature extraction, correspondence estimation, and geometry recovery substantially less stable than in clear-medium scenes. Practical systems must also balance reconstruction fidelity against computational limits in applications such as inspection, navigation, and field mapping.

These challenges matter across marine science, archaeology, inspection, and seafloor mapping, and they have shaped a clear methodological progression. Underwater 3D reconstruction has moved from calibration-heavy photogrammetry and SLAM toward learning-based underwater MVS, NeRF, 3D Gaussian Splatting, and physics-guided hybrid formulations that model the water medium more explicitly.

Within this progression, visual enhancement remains relevant because reconstruction quality depends not only on geometry estimation but also on how underwater image degradation is handled. In this review, underwater physics provides the forward model, UIE provides image-domain priors, and reconstruction methods either rely on these assumptions implicitly or encode them explicitly.

This section therefore first covers photogrammetry and underwater MVS within the photogrammetry part (subsection 4.1, subsubsection 4.1.3), then discusses NeRF (subsection 4.2), 3D Gaussian Splatting (subsection 4.3), and finally the main limitations and open challenges for robust underwater scene reconstruction (subsection 4.6).

4.1 Photogrammetry

Fundamental Principles. Photogrammetry reconstructs 3D structures from overlapping 2D images by identifying correspondences across viewpoints and solving for camera poses and scene geometry. Key modules include: 1) feature extraction and matching, which identify local features (corners, edges, or learned descriptors) robust to changes in viewpoint or illumination; 2) camera pose estimation via SfM, which incrementally or globally determines camera extrinsic/intrinsic parameters such that reprojected correspondences align in 3D; 3) dense reconstruction with MVS, which estimates detailed geometry by matching pixels across multiple images, typically using techniques such as patch-based stereo or plane sweeping; and 4) meshing and texturing, which convert the resulting point cloud into a polygon mesh and map images onto the surface to retain realism.

Underwater photogrammetry modifies these steps to handle color distortion, low contrast, and refraction effects. For instance, robust matching frequently requires color normalization or contrast enhancement in a preprocessing pipeline.

4.1.1 Photogrammetry Approaches for Underwater Scenes.

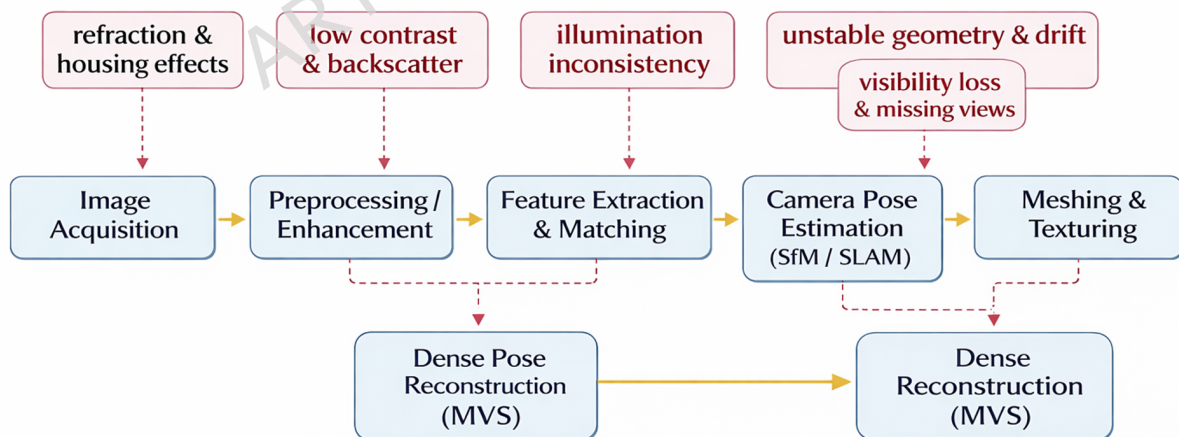


Fig. 12 Conceptual flow of underwater photogrammetry. Compared with clear-medium pipelines, underwater processing must explicitly handle refraction, low contrast, backscatter, and illumination inconsistency across acquisition, matching, pose estimation, and dense reconstruction.

A large body of literature (Prado et al. 2020; Zhang et al. 2022a) has tackled photogrammetry in the subaqueous setting. For instance, Prado et al. (2020) developed a specialized pipeline for capturing circalittoral rocky shelves, demonstrating that terrain classification can enhance region-based matching accuracy. Similarly, Wright et al. (2020) assessed the accuracy of SfM pipelines for archaeology-related tasks, such as mapping submerged shipwrecks. Their findings indicated that SfM often outperformed Real-Time Kinematic (RTK) surveying for localized applications. In

another study, [Nocerino et al. \(2020\)](#) incorporated Multi-View Stereo (MVS) to refine the reconstructed geometry following the initial SfM step, effectively mitigating the bowling effect, where large continuous surfaces appear erroneously curved.

Common issues in underwater SfM revolve around stable pose estimation when the scene lacks strong textures or includes extensive repeated patterns (e.g., sandy seabed). Bowling or doming arises from marginal pose constraints ([Wright et al. 2020](#)), where small camera rotations or inaccurate correspondences accumulate into large global errors. Although advanced bundle adjustment approaches can alleviate some drift, the presence of refraction or local lighting variations often complicates the inherent assumption of a single pinhole projection model. To address these problems, specialized calibration techniques have been introduced. Researchers sometimes use multi-camera rigs with known baselines, allowing for direct stereo matching that is more robust to color anomalies. Others implement *flat port* or *dome port* calibrations, explicitly modeling the glass boundary through which images are captured. The distinction matters: flat ports generally introduce stronger refraction and violate central pinhole assumptions more severely, while hemispherical dome ports often yield better photogrammetric behavior in practice and can approach a single-viewpoint model only when the lens is well centered and aligned with the dome geometry ([Menna et al. 2017](#); [She et al. 2022](#)). Even so, both configurations can bias metric geometry if refractive interfaces are ignored, which is why refractive calibration or ray-based modeling remains important for accurate underwater photogrammetry ([Sedlazeck et al. 2009](#)).

Bowling Effects and Remedies. As noted, large uniform terrains, such as expansive sandy seafloors or regions covered with short vegetation, can hinder robust alignment in SfM due to a lack of high-frequency features. This often leads to degenerate configurations, resulting in reconstructions with exaggerated curvature or bowl-shaped deformations ([Wright et al. 2020](#); [Samboko et al. 2022](#)). To mitigate these issues, several strategies have been proposed. One common approach is to introduce artificial markers with known geometry or color-coded patterns, which serve as reliable anchor points for SfM ([Wright et al. 2020](#); [Wittmann et al. 2024](#)). Divers or automated systems can place these markers strategically to enhance feature matching. Another effective method involves leveraging trajectory constraints. When an autonomous underwater vehicle (AUV) or remotely operated vehicle (ROV) logs inertial or acoustic data, these measurements can be integrated into the reconstruction pipeline to minimize drift and improve overall stability. Additionally, mesh regularization techniques ([Aubram 2013](#)) are often employed as a post-processing step following standard multi-view stereo (MVS). By enforcing surface smoothness or incorporating planar constraints, these methods help reduce spurious curvature in the final model. In cases where partial depth measurements are available, incorporating data from sonar or short-range LiDAR can further stabilize the photogrammetry pipeline ([Istenič et al. 2019](#)). These local depth cues provide additional constraints on scale and geometry, effectively reducing global distortions in the reconstructed scene.

4.1.2 Real-Time Visual SLAM

Where offline SfM reconstructs a scene after collecting images, visual SLAM attempts to solve camera localization (odometry) and mapping on the fly. Underwater robots can deploy SLAM for navigation, obstacle avoidance, or on-the-spot mapping. According to [Storlazzi et al. \(2016\)](#), visual SLAM can exceed the resolution of large-scale LiDAR or side-scan sonar data, which typically provide coarser point clouds at a broader scale. This advantage is crucial for tasks like surveying coral polyps, delicate rock formations, or subtle archaeological relics. However, achieving real-time performance requires carefully chosen features or deep learning-based front ends robust to turbidity and color distortions. Some pipelines also incorporate acoustic or inertial measurements for multi-sensor fusion, offsetting the difficulties introduced by water's optical properties.

4.1.3 End-to-End Underwater MVS

Recent work has begun to explore end-to-end underwater MVS as an alternative to purely sequential SfM-plus-MVS pipelines. The main idea is not merely to enhance frames before reconstruction, but to make underwater-aware correspondence learning and dense geometry inference part of the reconstruction model itself.

In particular, [Yang et al. \(2025\)](#) propose an end-to-end underwater multi-view stereo framework for dense scene reconstruction that learns geometry directly from underwater image sets while explicitly accounting for

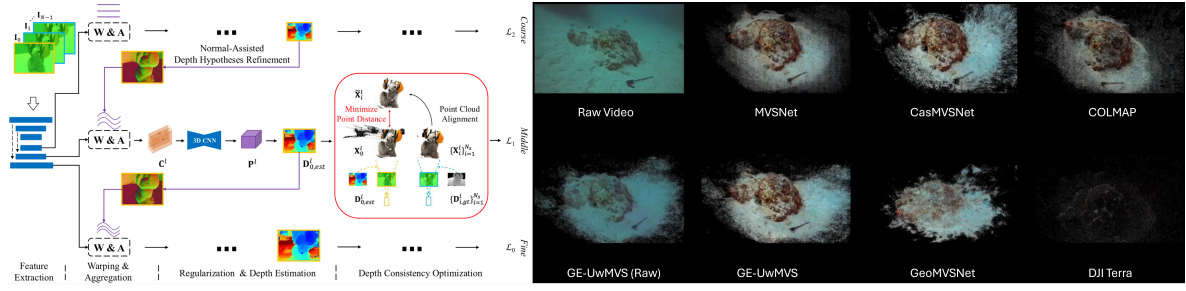


Fig. 13 (a) End-to-end underwater MVS framework, GE-UwMVS (Yang et al. 2025). (b) Dense reconstruction comparison against representative MVS baselines, including COLMAP (Schonberger and Frahm 2016), MVSNet (Yao et al. 2018), CasMVSNet (Gu et al. 2020), GeoMVSNet (Zhang et al. 2023), and the commercial software DJI Terra (DJI 2026).

appearance degradations caused by the medium. The importance of this work is less that it establishes a mature family of methods, and more that it shows underwater MVS can be formulated as a learned dense reconstruction problem rather than as a fixed photogrammetric pipeline with enhancement only attached upstream. In this sense, it is an early but important attempt to incorporate underwater appearance distortion into the dense matching and depth inference stages themselves. Overall, end-to-end underwater MVS remains a small but important direction for closing the loop from visual enhancement to dense geometry recovery under real underwater conditions.

4.2 Neural Radiance Fields (NeRF)

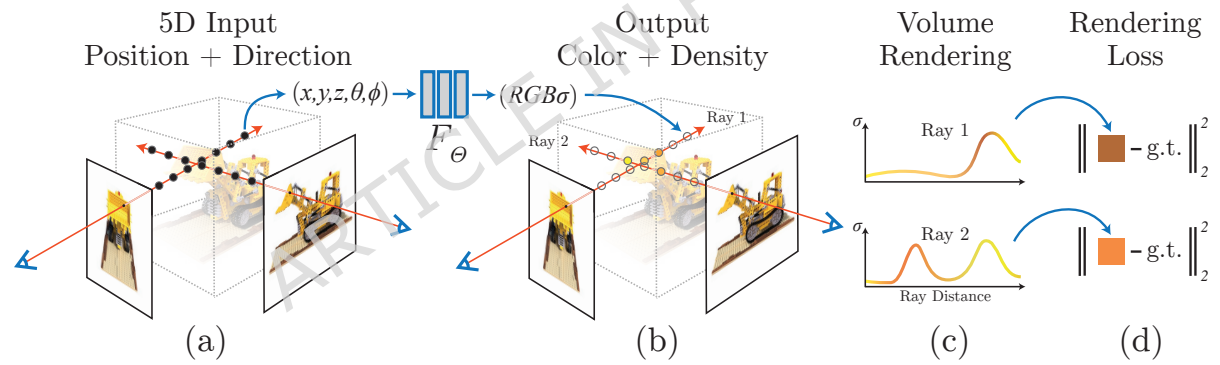


Fig. 14 Illustration of NeRF and its differentiable rendering process. It involves sampling 5D coordinates (position and direction) along camera rays (a), using an MLP to produce color and density (b), and rendering these into an image (c). The differentiable function allows optimization by minimizing differences between rendered and actual images (d) (Mildenhall et al. 2020)

Neural Radiance Fields (NeRF), introduced by Mildenhall et al. (2020), have brought a paradigm shift to 3D reconstruction and novel view synthesis. Unlike traditional representations such as voxel grids or explicit point clouds, NeRF encodes scene appearance and geometry through a Multi-Layer Perceptron (MLP). Given a 3D point $\mathbf{x}$ and a viewing direction $\mathbf{d}$, NeRF predicts color $c(\mathbf{x}, \mathbf{d})$ and volume density $\sigma(\mathbf{x})$, allowing a continuous representation of the underlying scene.

4.2.1 Principles and Volume Rendering

The main idea behind NeRF is grounded in the volume rendering equation (Kajiya 1986). A ray is parameterized as:

$$\mathbf{r}(t) = \mathbf{o} + t\mathbf{d},$$

Table 8 High-level comparison of NeRF features

Aspect	NeRF Advantages	NeRF Disadvantages
Representation	Implicit, continuous 3D model	Complex neural inference
View Synthesis	Photorealistic novel views	High computational cost
Geometry	Learned implicitly	No direct mesh extraction
Generalization	Uses raw image supervision	Often data-hungry

where $\mathbf{o}$ is the camera center and $\mathbf{d}$ is the unit viewing direction. The radiance along the ray is accumulated as:

$$C(\mathbf{r}) = \int_{t_n}^{t_f} T(t) \sigma(\mathbf{r}(t)) c(\mathbf{r}(t), \mathbf{d}) dt,$$

where

$$T(t) = \exp\left(-\int_{t_n}^t \sigma(\mathbf{r}(s)) ds\right)$$

represents the transmittance from t_n to t . NeRF approximates this integral by sampling discrete points along the ray and summing their contributions.

To train NeRF, one minimizes the discrepancy between synthesized pixels and real observed pixels in the input images. A common objective is the mean squared error (MSE) between the rendered color C_i and the corresponding ground truth $\hat{C}_i$:

$$L = \sum_i \|C_i - \hat{C}_i\|^2.$$

Through gradient-based optimization, the MLP learns both the geometry (σ) and appearance (c).

Comparison with Traditional 3D Reconstruction. Conventional 3D reconstruction methods, such as Structure-from-Motion (SfM) and Multi-View Stereo (MVS), rely on geometric feature matching and explicit estimation of depth maps or point clouds (Snavely et al. 2006; Seitz et al. 2006). While they can yield accurate geometry with sufficient texture and baseline, they typically require clear, feature-rich images and separate steps for dense reconstruction. Photometric Stereo (Woodham 1980) is adept at capturing surface normals under controlled lighting, but it lacks flexibility for unstructured, real-world scenes.

By contrast, NeRF directly learns an implicit volumetric function from raw images, often resulting in superior novel view synthesis. However, it can be computationally demanding, and extracting explicit geometry (e.g., meshes) is less straightforward. NeRF also tends to require many images around the subject for optimal training, though newer variants reduce that data requirement.

4.2.2 NeRF Variants

Rather than exhaustively surveying the very large clear-medium NeRF literature, we retain here only the generic variants that best contextualize the later underwater discussion. The most relevant background directions concern efficiency, robustness to unconstrained capture, and dynamic/video modeling. The next subsection then focuses on the genuinely underwater methodological divergence.

Efficient NeRF representations. One major line of work reduces NeRF’s heavy training and rendering cost through more compact scene encodings. Instant-NGP (Müller et al. 2022a) uses multi-resolution hash grids to cut training time from hours to seconds, Plenoxels (Fridovich-Keil et al. 2022) replace the large MLP with directly optimized sparse voxels and spherical harmonics, and TensoRF (Chen et al. 2022) factorizes the radiance field into low-rank tensor components for compact storage and faster optimization. These methods are representative because they define the efficiency baseline against which later underwater NeRF systems must still be judged: once medium-aware rendering is added, training cost and memory usage remain a practical bottleneck.

Unconstrained and pose-robust NeRF. Another relevant direction relaxes NeRF’s assumption of uniformly captured, well-posed image sets. NeRF-W (Martin-Brualla et al. 2021) introduces appearance latents and transient components to absorb lighting variation and occluders, while BARF (Lin and et al. 2021) and UP-NeRF (Kim et al. 2023) explicitly tackle uncertain or missing camera poses during optimization. These variants matter for underwater sensing because real deployments often combine viewpoint drift, radiometric inconsistency, and transient distractors rather than the carefully calibrated capture conditions assumed by the original NeRF.

Dynamic / video NeRF. Dynamic NeRF variants extend the radiance field to time-varying scenes through deformation fields or temporal consistency models. Representative examples include NeRFies (Park et al. 2021) and D-NeRF (Pumarola et al. 2021) for deformable scenes, together with Video-NeRF (Xian et al. 2021) and Neural Radiance Flow (Du et al. 2021) for temporally coherent 4D rendering. This direction is especially relevant underwater because moving organisms, marine snow, flicker, and platform motion all challenge the static-scene assumption; indeed, Gough et al. (2025) note that generic dynamic NeRF methods remain insufficient for high-frequency underwater disturbances without more explicit disturbance-aware modeling.

Taken together, these generic NeRF variants mainly address efficiency, robustness, or temporal modeling in clear-medium settings. Underwater NeRF methods inherit these concerns, but they must additionally account for medium-aware rendering, radiometric degradation, and tighter coupling between restoration and reconstruction.

4.2.3 Underwater NeRF Applications

We organize existing underwater NeRF methods by their modeling choices. Compared with clear-medium NeRF variants, the underwater extensions usually modify the rendering equation, introduce explicit medium parameters, or couple restoration and reconstruction within the same optimization. To make this design space more explicit, we first summarize the main underwater NeRF families through compact methodological equations and then map representative methods onto these routes.

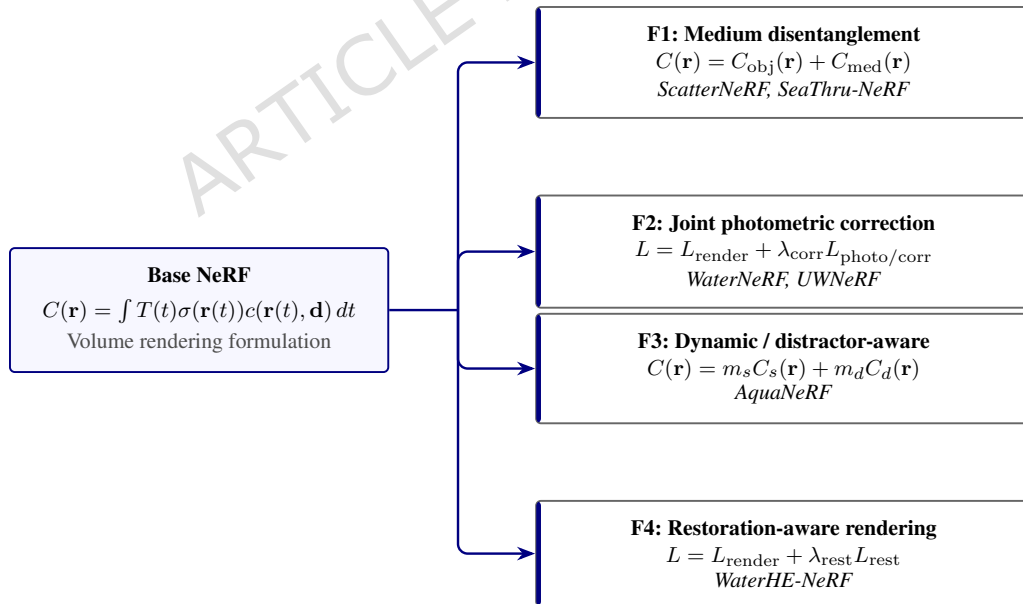


Fig. 15 Methodological design space of underwater NeRF variants. The main families diverge through medium disentanglement, joint photometric correction, dynamic distractor handling, and restoration-aware rendering.

Medium disentanglement and revised-IFM-guided rendering.

$$C(\mathbf{r}) = C_{\text{obj}}(\mathbf{r}) + C_{\text{med}}(\mathbf{r}),$$

where $C_{\text{obj}}(\mathbf{r})$ captures scene radiance transported along the ray and $C_{\text{med}}(\mathbf{r})$ captures medium-dependent terms such as backscatter, attenuation, or water colour. ScatterNeRF (Ramazzina et al. 2023) demonstrates the basic idea of disentangling scattering effects from scene content in participating media, which carries over naturally to underwater settings. SeaThru-NeRF (Levy et al. 2023) is the first dedicated underwater NeRF to build this idea on the revised underwater image formation model (Akkaynak and Treibitz 2018), described in Equation 11, introducing per-ray backscatter, attenuation-density, and medium-colour branches so that object radiance and water-medium appearance are modeled separately. SP-SeaNeRF (Chen et al. 2024) extends this line by adding learnable illumination embeddings and an explicit degradation simulation step, improving sharpness under non-uniform lighting. The main strength of this family is that it grounds novel-view synthesis in a physically interpretable forward model; its main limitation is the added optimization complexity and the need to separate medium effects from sparse observations.

Photometric correction coupled with reconstruction.

$$L = L_{\text{render}} + \lambda_{\text{corr}} L_{\text{photo/corr}},$$

where the optimization is driven jointly by radiance-field reconstruction and a correction term that stabilizes colour or illumination across views. A second family therefore couples radiometric correction with geometry estimation rather than treating enhancement as an external pre-processing stage. WaterNeRF (Sethuraman et al. 2023) combines underwater light transport modeling with optimal-transport-based colour correction to stabilize view-to-view appearance. UWNeRF (Tang et al. 2024) integrates photometric correction directly into the NeRF rendering process, allowing the network to infer clearer 3D point appearance while preserving cues that may be discarded by independent enhancement. This direction is important because it explicitly links UIE-style colour recovery to multi-view consistency, instead of assuming that the two can be optimized separately.

Dynamic scenes, floaters, and distractors.

$$C(\mathbf{r}) = m_s C_s(\mathbf{r}) + m_d C_d(\mathbf{r}), \quad m_s + m_d = 1,$$

where the rendering is decomposed into static and disturbance-related contributions, or equivalently biased toward a dominant surface to suppress floating clutter. Dynamic underwater disturbances remain a central difficulty because classical NeRF assumes a static scene. UWNeRF (Tang et al. 2024) differentiates between static and dynamic components using motion masks as a secondary mechanism, while AquaNeRF (Gough et al. 2025) reduces the impact of floaters and moving objects by enforcing a single dominant surface along each ray and maintaining medium transmittance with a Gaussian weighting scheme. These strategies reduce artifact accumulation around fish or suspended particles, but their effectiveness still depends on the quality of masks or ray-level visibility assumptions.

Enhancement-aware and restoration-aware rendering.

$$L = L_{\text{render}} + \lambda_{\text{rest}} L_{\text{rest}},$$

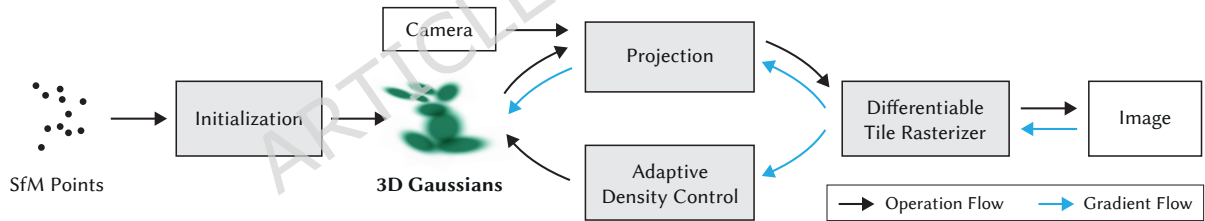
where restoration-aware terms or auxiliary fields are written directly into the radiance-field optimization rather than applied as a separate post-processing stage. Some methods therefore introduce restoration cues more explicitly into the radiance-field parameterization. WaterHE-NeRF (Zhou et al. 2025), for example, augments NeRF with a water-ray matching field derived from Retinex-style reasoning, aiming to recover colour and illumination while reconstructing geometry. Relative to standard NeRF variants, these models are closer to joint enhancement-reconstruction systems, but they also inherit the risk that aggressive restoration assumptions may distort the very correspondences needed for geometry estimation.

Table 9 maps representative methods to four shorthand families: F1 = medium disentanglement / revised-IFM-guided rendering, F2 = photometric correction coupled with reconstruction, F3 = dynamic / distractor-aware rendering, and F4 = restoration-aware rendering.

Table 9 Summary of representative underwater NeRF methods.

Method	Primary Core family	underwater modeling idea	Relation physics / IFM	to Relation enhancement	to Dynamic / dis- tractor handling	Strength	Limitation
ScatterNeRF (Ramazzina et al. 2023)	F1	Separates scene content from scattering medium in participating media	Medium-aware rendering, but not underwater-specific IFM	Indirect; improves visibility through disentanglement	Not explicit	General scattering disentanglement	Limited underwater specificity
SeaThru-NeRF (Levy et al. 2023)	F1	Per-ray backscatter, attenuation, and medium colour branches	Explicitly based on revised underwater IFM	Jointly restores colour while reconstructing	Static-scene assumption	Physically interpretable underwater rendering	Higher model complexity and data sensitivity
SP-SeaNeRF (Chen et al. 2024)	F1	SeaThru-NeRF plus illumination embeddings and degradation simulation	Physics-guided with learnable lighting factors	Joint enhancement and reconstruction	Limited dynamic handling	Better sharpness under non-uniform illumination	More parameters and training overhead
UWNeRF (Tang et al. 2024)	F2	Integrates photometric correction into NeRF geometry learning	Physics-aware radiometric rendering	Strong coupling of correction and reconstruction	Motion masks for dynamic regions	Preserves geometry cues better than separate pre-enhancement	Depends on mask quality and robust pose estimates
WaterHE-NeRF (Zhou et al. 2025)	F4	Water-ray matching field with Retinex-style colour correction	Implicit physical prior plus restoration field	Explicit restoration-aware rendering	Not a primary focus	Connects colour recovery to view synthesis	Restoration assumptions can bias geometry
AquaNeRF (Gough et al. 2025)	F3	Single-surface-per-ray rendering to suppress floaters	Medium transmittance modeled along ray	Indirect enhancement through cleaner rendering	Explicit floater suppression	More robust static-object reconstruction in clutter	May oversimplify complex multi-layer scenes
WaterNeRF (Sethuraman et al. 2023)	F2	Couples underwater transport modeling with colour-consistency correction	Physics-aware transport plus colour stabilization	Jointly optimizes appearance correction and reconstruction	Not a primary focus	Better view-to-view photometric stability	Additional coupled losses and optimization burden

4.3 3D Gaussian Splatting

**Fig. 16** Overview of the 3D Gaussian Splatting pipeline. Starting from a sparse SfM point cloud, an initial set of 3D Gaussians is constructed. These Gaussians are iteratively refined through projection, adaptive density control, and differentiable tile-based rasterization, where gradients are propagated to update their parameters. This optimization procedure enables efficient training while maintaining high fidelity, and the final representation supports real-time rendering and interactive navigation across diverse scenes (Kerbl et al. 2023a)

3D Gaussian Splatting (3DGS) (Kerbl et al. 2023a) is an explicit scene representation in which a 3D scene is modeled by Gaussian primitives rather than an implicit volume. Each Gaussian primitive $\mathcal{G}_i$ is parameterized by a center position μ_i , a covariance matrix Σ_i , an opacity α_i , and a view-dependent colour c_i often decoded from spherical harmonics. For differentiable optimization, the covariance is typically factorized into rotation and scaling components:

$$\Sigma_i = R_i S_i S_i^T R_i^T. \quad (22)$$

For rendering, the 3D Gaussian is projected to the image plane through a viewing transformation W . Under a local affine approximation, its 2D covariance can be written as

$$\Sigma'_i = J W \Sigma_i W^T J^T, \quad (23)$$

where J is the Jacobian of the projection around the current viewing configuration. The contribution of each projected Gaussian is then accumulated by alpha compositing:

$$C = \sum_i c_i \alpha'_i \prod_{j=1}^{i-1} (1 - \alpha'_j), \quad \alpha'_i = \alpha_i \exp\left(-\frac{1}{2}(x - \mu'_i)^T (\Sigma'_i)^{-1} (x - \mu'_i)\right). \quad (24)$$

Here, μ'_i denotes the projected 2D mean of the Gaussian and α'_i is its effective blending weight at pixel x . In practice, optimization alternates between updating Gaussian parameters and adapting the Gaussian density, which enables fast convergence and real-time rendering but can still require careful control to avoid redundant primitives or loss of fine detail.

3DGS offers several advantages:

- **Speed of Convergence:** The explicit nature of 3DGS enables faster training and more accurate estimation of scene geometry and color with fewer samples compared to methods such as NeRF.
- **Real-Time Rendering Potential:** Once trained, 3DGS supports real-time rendering by directly rasterizing Gaussian points using GPU-based alpha blending.
- **Adaptability:** Leveraging spherical harmonics allows 3DGS to capture appearance variations under complex lighting conditions. However, additional modeling is needed to handle phenomena such as reflection or refraction.

4.3.1 3DGS Variants

Since the introduction of 3DGS (Kerbl et al. 2023a), several enhancements have been proposed to extend its capabilities. Some techniques that could potentially be used for underwater scenes are described here:

Anti-Aliasing. Mip-Splatting (Yu et al. 2024) addresses aliasing artifacts that emerge when the sampling rate changes—a shortcoming primarily due to the vanilla 3DGS’s lack of 3D frequency constraints and its reliance on a 2D dilation filter. By integrating 3D smoothing with 2D mip filters, this method enables aliasing-free rendering. Similarly, Yan et al. (2024) report that aliasing becomes pronounced in low-resolution renderings when the pixel size drops below the Nyquist threshold. They propose a multi-scale rendering strategy using levels-of-detail (LOD) and mipmap techniques that synthesize larger Gaussians for low-resolution outputs by aggregating smaller Gaussians from high-resolution inputs.

Deblurring. Chen and Liu (2024) observe that novel view synthesis from the original 3DGS degrades significantly with blurred input images, which could possibly be one of the challenges for underwater 3D reconstruction.. To mitigate this, they introduce a physically based model that approximates the camera trajectory by incorporating pseudo-camera poses along the path and blending the corresponding images to simulate motion blur. Additionally, Lee et al. (2024) propose an alternative variant that specifically addresses defocusing blur, further enhancing performance in challenging imaging conditions .

4D Rendering. . The original 3DGS is designed for static scenes, limiting its ability to capture dynamic inputs. To overcome this, 4DGS (Wu et al. 2024) employs an encoder-decoder deformation network (Yang et al. 2024) that uses both timestamps and Gaussian center coordinates to predict deformed positions and covariance matrices. In a related approach, Lin et al. (2024) introduces the Gaussian-flow method, which uses a dual-domain deformation model to estimate deformed attributes for efficient 4D scene rendering. However, while effective for modeling slow motion, these techniques struggle with fast dynamics and unpredictable trajectories.

Density Control. 3DGS improves the representation of the Gaussian point cloud through an adaptive density control strategy. Abnormal average 2D gradients upon projection reveal regions of under-reconstruction, prompting the subdivision or duplication of Gaussians based on size. Nevertheless, Zhang et al. (2024) note that this approach can introduce needle-like and blurred artifacts in sparse initial point density areas. They propose using the total coverage pixel count across different views as the weighting metric—rather than solely the number of viewpoints—to achieve more detailed reconstructions in regions with repetitive textures. These clear-medium variants mainly improve splatting as a generic rendering and optimization framework. The more distinct underwater divergence begins in the next subsection, where the literature is better organized by medium modeling, structural priors, disturbance handling, and restoration coupling.

4.3.2 Underwater 3DGS Applications

Despite advances that have enabled 3DGS to generate more accurate and real-time reconstructions, it still faces challenges in underwater environments. The inherent design of the original 3DGS focuses on representing geometric features and does not account for the scattering characteristics of the medium. Furthermore, underwater images are affected by complex optical phenomena—such as light absorption, backscattering, and motion-induced blur—that further complicate reconstruction.

Yang et al. (2024) compare vanilla NeRF and vanilla 3DGS on underwater captures and report a pattern that broadly motivates later underwater-specific splatting methods: 3DGS is attractive for fast and sharp rendering in static scenes, but it needs additional modeling to cope with underwater physics, sparse baselines, and dynamic distractors. We therefore organize underwater 3DGS variants by design choice rather than chronology. Similar to the underwater NeRF overview above, this subsection first summarizes the main underwater splatting families through compact methodological equations and then maps representative methods onto these routes.

Compared with NeRF, this limitation is also representational. NeRF’s volumetric rendering can more naturally absorb distributed attenuation and scattering along a ray, whereas standard 3DGS represents the scene with explicit Gaussian particles whose colour and opacity can easily entangle true geometry with water-column effects. In turbid water, backscatter and marine snow may therefore be baked into floating splats or unstable opacity fields unless the renderer explicitly separates medium and object transport. This is one reason why underwater 3DGS papers increasingly add medium-aware transmittance, distractor masks, temporal degradation models, or restoration-coupled optimisation rather than relying on clear-medium splatting alone.

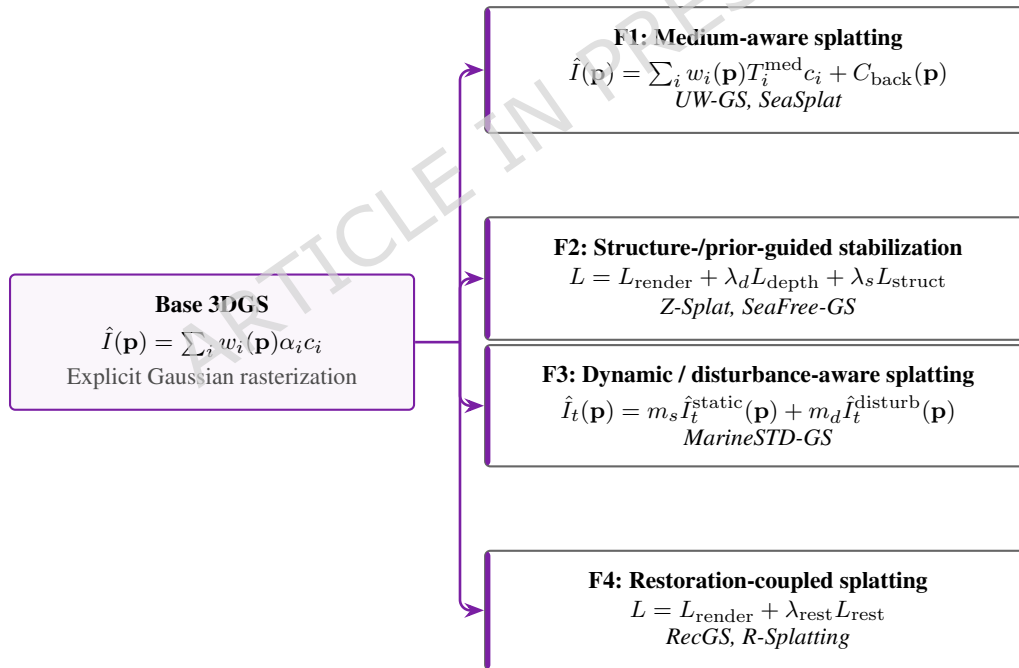


Fig. 17 Methodological design space of underwater 3D Gaussian Splatting variants. The main families diverge through medium-aware splatting, structure- and prior-guided stabilization, dynamic disturbance modeling, and restoration-coupled optimization.

Medium-aware splatting and medium-object transport.

$$\hat{I}(\mathbf{p}) = \sum_i w_i(\mathbf{p}) T_i^{\text{med}} c_i + C_{\text{back}}(\mathbf{p}),$$

where medium-aware transmittance and backscatter terms are added on top of Gaussian colour and opacity so that the renderer separates object appearance from water-column effects. This is the main direction in UW-GS (Wang

et al. 2025), SeaSplat (Yang et al. 2025), and WaterSplatting (Li et al. 2025), all of which explicitly model attenuation, backscatter, or object-medium transport within splatting. Gaussian Splashing (Muallem et al. 2024) and Aquatic-GS (Liu et al. 2024) follow the same broad idea by treating the water medium as part of the rendering problem rather than as noise to be ignored. Relative to standard 3DGS variants, this family is the closest analogue to revised-IFM-aware underwater NeRFs, but it also increases optimization complexity and parameter coupling.

Structure- and prior-guided stabilization.

$$L = L_{\text{render}} + \lambda_d L_{\text{depth}} + \lambda_s L_{\text{struct}},$$

where depth, edge, semantic, or smoothness priors are injected into splatting to stabilize geometry under degraded appearance. Z-Splat (Qu et al. 2024) improves limited-baseline underwater reconstruction by fusing sonar information with RGB splats, while SeaFree-GS (Liu et al. 2025), RestorGS (Qiao et al. 2025), RUSplatting (Jiang et al. 2025a), SWAGSplatting (Jiang et al. 2025b), and 3D-UIR (Yuan et al. 2025) further exploit pseudo-depth, edge-aware losses, semantic cues, or smoothness constraints. UW-GS also uses normalized pseudo-depth from Depth Anything (Yang et al. 2024) as a secondary mechanism. Compared with generic 3DGS regularization, these underwater variants use priors not merely for denoising but to stabilize geometry when attenuation and colour shifts weaken correspondence reliability.

Dynamic disturbances, marine snow, and temporal degradations.

$$\hat{I}_t(\mathbf{p}) = m_s \hat{I}_t^{\text{static}}(\mathbf{p}) + m_d \hat{I}_t^{\text{disturb}}(\mathbf{p}), \quad m_s + m_d = 1,$$

where the observation at time t is decomposed into static structure and disturbance-related terms, or equivalently filtered by masks and temporal degradation models. Dynamic disturbances arise not only from moving fish or suspended particles, but also from underwater-specific temporal effects such as caustics and flickering. MarineSTD-GS (Liu et al. 2025) explicitly models temporal degradations in the underwater image formation process, while UW-GS uses motion masks to reduce marine distractors as a secondary component. These methods differ from clear-medium dynamic 3DGS by treating marine clutter and temporal lighting instability as part of the observation model itself.

This also raises a coupling question for integrated pipelines: front-end marine-snow removal or dehazing can help dynamic reconstruction only if it remains temporally stable. If enhancement is applied independently to each frame and changes particle appearance, motion boundaries, or flicker statistics inconsistently, it can disrupt motion masks, erase useful temporal cues, or confuse the dynamic components of a 3DGS or NeRF backend instead of supporting them.

Restoration-coupled splatting.

$$L = L_{\text{render}} + \lambda_{\text{rest}} L_{\text{rest}},$$

where restoration cues are written directly into splatting optimization rather than applied as a separate pre-processing stage. RecGS (Zhang et al. 2024) improves perceptual consistency by suppressing caustics with low-pass filtering and recurrent training, and R-Splatting (Huang et al. 2025) fuses multiple restoration outputs into a single 3DGS model to address cross-view illumination shifts. AtlantisGS (Yi et al. 2025) can be read as a more aggressive quality-oriented continuation of this route, pushing rendering quality further through stronger Gaussian optimization. These methods highlight a central point of this review: underwater 3D reconstruction increasingly depends on how enhancement cues are integrated, not simply on how well a vanilla renderer fits degraded images.

The main engineering trade-off relative to underwater NeRF is therefore clear: 3DGS offers faster training and much stronger interactive rendering once the representation is optimized, but NeRF-style volumetric models remain conceptually better matched to participating media and ray-wise scattering. In practice, underwater 3DGS compensates by introducing explicit medium terms, priors, or distractor suppression, whereas underwater NeRF variants can often encode those effects more directly in volumetric rendering at the cost of heavier optimisation.

Table 10 maps representative methods to four shorthand families: F1 = medium-aware splatting, F2 = structure-/prior-guided stabilization, F3 = dynamic / disturbance-aware splatting, and F4 = restoration-coupled splatting.

Table 10 Summary of representative underwater 3D Gaussian Splatting methods.

Method	Primary family	Core underwater modeling idea	Relation to physics / IFM	Relation to enhancement	to Dynamic / dis-tractor handling	Strength	Limitation
Z-Splat (Qu et al. 2024)	F2	Extends splats along depth and fuses sonar with RGB under limited baselines	No explicit water-medium model	Indirect; focuses on geometry support	Not a primary focus	Helps missing-cone and sparse-baseline cases	Limited colour/medium fidelity
UW-GS (Wang et al. 2025)	F1	Color-appearance model plus physics-guided density control	Explicit scattering-aware modeling	Produces cleaner rendered views through joint modeling	Motion masks plus pseudo-depth as secondary mechanisms	Strong joint gains in rendering and geometry	Depends on pseudo-depth and mask quality
SeaSplat (Yang et al. 2025)	F1	Physically grounded underwater IFM inside splatting	Explicit medium-aware rendering	Generates enhanced renderings during reconstruction	Static-scene assumption	Fast rendering with improved visibility	Limited handling of moving objects
WaterSplat (Li et al. 2025)	F1	Separate transmittance for objects and surrounding medium	Strong physics-guided transmittance modeling	Joint restoration and rendering	Limited explicit dynamics	Competitive quality with real-time rendering	Added model complexity
RecGS (Zhang et al. 2024)	F4	Recurrent training with caustic suppression	Weak physical prior	Restoration-coupled via caustic removal	Indirect temporal consistency only	Better perceptual stability	Does not explicitly model water medium
R-Splatting (Huang et al. 2025)	F4	Fuses multiple restoration outputs into one splat model	Uses restoration cues more than explicit IFM	Strong restoration-coupled reconstruction	Handles illumination variation across views	Improves geometric fidelity under varying lighting	Sensitive to upstream restoration quality
MarineSTD-GS (Liu et al. 2025)	F3	Integrates temporal degradations such as caustics and flicker	Physics-guided temporal degradation model	Indirect enhancement through temporal correction	Explicit spatiotemporal degradation modeling	Better robustness to dynamic illumination	Larger model and training cost

4.4 Performance Evaluation of Underwater NeRF/3DGS Models

This subsection summarizes reported performance and visual comparisons of representative underwater NeRF- and 3DGS-based reconstruction models. Rather than treating these results as a unified reproduced benchmark, we use them to compare how existing models couple rendering, restoration, medium modeling, and dynamic-scene handling under underwater degradation.

Table 11 compiles representative reported results for NeRF- and 3DGS-based methods on the SeaThru-NeRF dataset. All values are taken from the original publications rather than reproduced here under one unified experimental protocol, so the table should be read as a literature-level synthesis rather than as a strict fair-play leaderboard. It is most useful for illustrating trends: early physics-guided models such as UW-GS already outperform vanilla 3DGS by a clear margin, and later methods such as AtlantisGS (Yi et al. 2025) report further gains in rendering quality and efficiency. The specific ranking should therefore be interpreted together with each method’s assumptions about dynamics, medium modeling, priors, preprocessing, and training setup.

Representative underwater scene reconstruction models increasingly integrate visual recovery with medium-aware rendering, rather than treating enhancement as a separate post-processing step. Figure 18 presents the rendered images and the estimated clean images produced by UW-3DGS (Wang et al. 2025), where the water-medium component is removed from the rendered result.

Figure 19 provides an illustrative sparse-view rendering comparison on the UWNeRF dataset among SeaThru-NeRF, UWNeRF, vanilla 3DGS, WaterSplatting, and AtlantisGS. The main review-level conclusion is not that one method is universally “best”, but that underwater reconstruction models benefit when medium effects, sparse-view constraints, and restoration cues are handled within the rendering process. In this comparison, AtlantisGS recovers sharper local structure and more coherent appearance under semi-transparent underwater media, whereas clear-medium or less medium-aware baselines tend to blur fine details or retain colour and opacity inconsistencies.

It is also important to note that most current benchmarking still reports rendering quality or image-domain fidelity rather than true joint geometry-and-trajectory outcomes. Metrics such as Chamfer distance, Cloud-to-Cloud

Table 11 Quantitative comparison of NeRF- and Splatting-based methods on the SeaThru-NeRF dataset. All values are literature-reported results from the original publications, included here for indicative comparison rather than as a unified reproduced benchmark.

Scene Method	Curacao			Panama			IUI-Reasea			Japanese-Redsea			Average		
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
<i>NeRF-based methods</i>															
MIP-360 (Barron et al. 2022)	28.23	0.683	0.571	18.32	0.556	0.595	19.62	0.624	0.492	19.55	0.510	0.520	21.93	0.593	0.545
Instant-NGP (Müller et al. 2022b)	27.91	0.707	0.385	23.46	0.636	0.426	20.63	0.558	0.603	23.23	0.655	0.357	23.81	0.639	0.443
UWNeRF (Tang et al. 2024)	30.03	0.828	0.238	23.75	0.687	0.263	25.81	0.853	0.183	22.70	0.624	0.348	25.57	0.748	0.258
SeaThru-NeRF (Levy et al. 2023)	29.92	0.856	0.298	26.90	0.789	0.330	25.89	0.754	0.353	21.75	0.735	0.337	26.11	0.784	0.330
ZipNeRF (Barron et al. 2023)	29.93	0.938	0.124	32.34	0.956	0.064	29.35	0.899	0.106	23.45	0.883	0.136	28.77	0.919	0.108
<i>Splatting-based methods</i>															
3DGS (Kerbl et al. 2023b)	30.97	0.936	0.193	30.80	0.919	0.196	23.13	0.877	0.240	21.97	0.867	0.202	26.72	0.900	0.208
WildGaussian (Kulhanek et al. 2025)	29.52	0.880	0.313	24.94	0.756	0.395	28.34	0.870	0.177	22.08	0.839	0.312	26.22	0.836	0.299
SeaSplat (Yang et al. 2025)	30.30	0.900	0.190	28.76	0.900	0.150	26.67	0.870	0.210	22.70	0.870	0.180	27.61	0.885	0.183
WA-GS (Fan et al. 2025)	28.29	0.900	0.158	30.07	0.938	0.084	30.43	0.891	0.186	23.17	0.864	0.153	27.99	0.898	0.145
3D-UIR (Yuan et al. 2025)	30.98	0.907	0.187	29.82	0.900	0.181	28.30	0.841	0.252	23.37	0.857	0.187	28.12	0.876	0.202
UW-GS (Wang et al. 2025)	31.77	0.943	0.144	31.79	0.936	0.116	28.65	0.933	0.125	23.05	0.860	0.190	28.82	0.918	0.144
RUSplatting (Jiang et al. 2025a)	30.96	0.932	0.161	31.87	0.934	0.140	29.80	0.929	0.185	24.54	0.871	0.181	29.29	0.917	0.167
RestorGS (Qiao et al. 2025)	31.95	0.944	0.055	30.79	0.932	0.046	29.97	0.952	0.028	24.05	0.882	0.071	29.19	0.928	0.050
WaterSplatting (Li et al. 2025)	32.67	0.957	0.110	31.49	0.948	0.075	29.39	0.910	0.180	25.20	0.904	0.113	29.19	0.930	0.120
SWAGSplatting (Jiang et al. 2025b)	31.84	0.941	0.161	31.82	0.939	0.139	30.37	0.936	0.180	24.52	0.889	0.177	29.64	0.926	0.164
R-Splatting (Huang et al. 2025)	32.98	0.956	0.163	32.52	0.930	0.107	30.15	0.947	0.105	24.03	0.868	0.211	29.92	0.925	0.147
AtlantisGS (Yi et al. 2025)	33.27	0.949	0.112	32.35	0.953	0.076	31.42	0.935	0.201	26.59	0.916	0.108	30.91	0.938	0.124

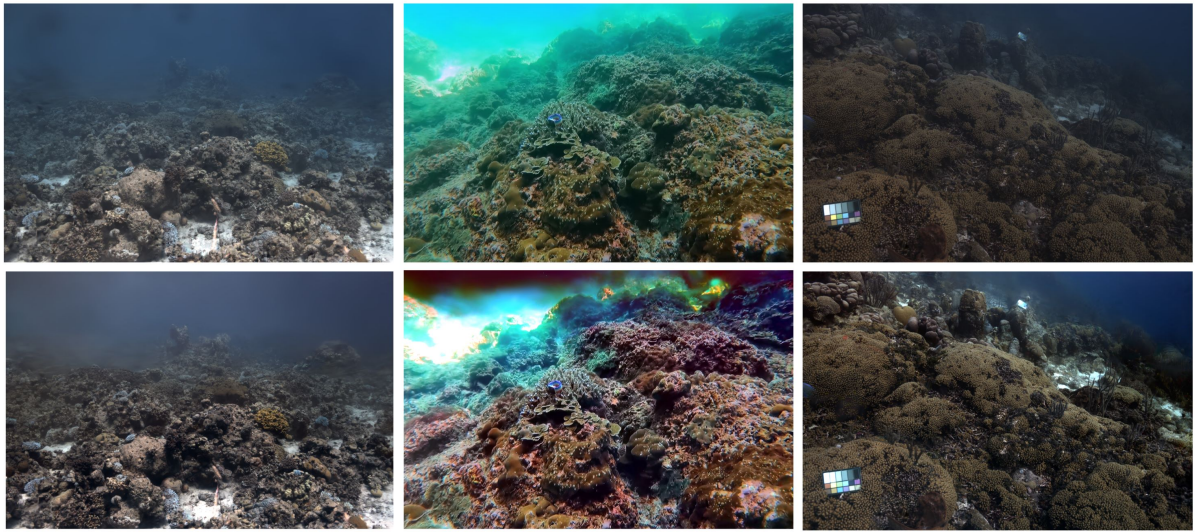


Fig. 18 Visualization of rendered images (top) and the estimated clean images (bottom) using UW-3DGS (Wang et al. 2025)

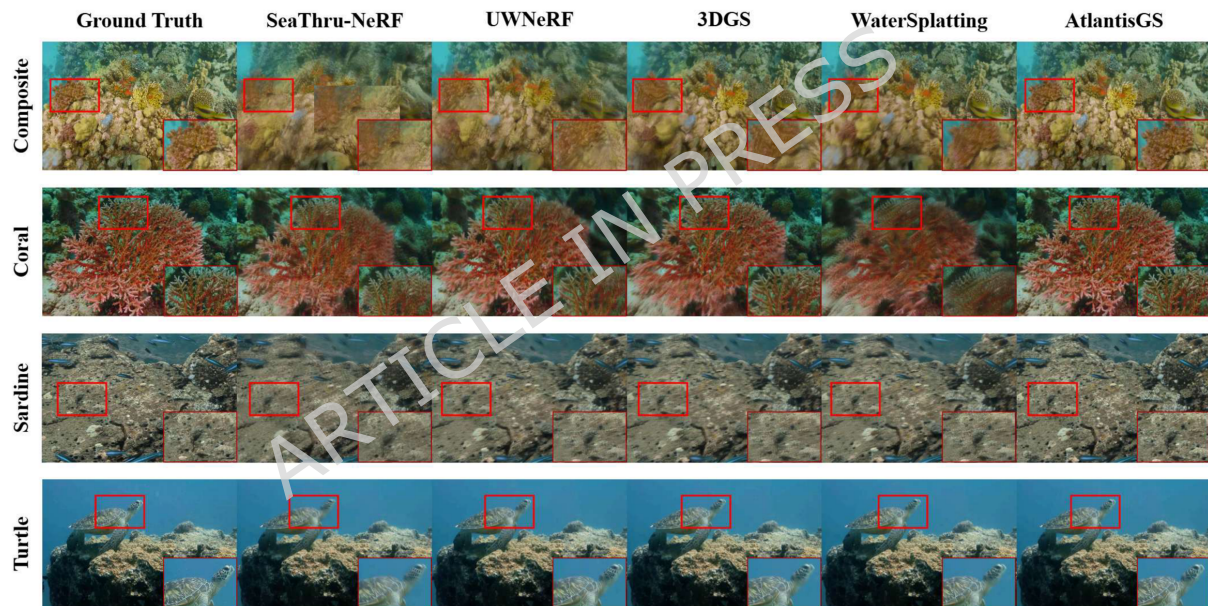


Fig. 19 Underwater scene rendering comparison on the UWNeRF dataset, reproduced from Fig. 5 of AtlantisGS (Yi et al. 2025). From left to right: ground truth, SeaThru-NeRF, UWNeRF, 3DGS, WaterSplatting, and AtlantisGS. The comparison highlights that traditional 3DGS and NeRF methods with proposal sampling struggle with semi-transparent underwater media.

distance, ATE, or RPE are still rarely reported together with enhancement-oriented measures, which makes fair cross-domain assessment of integrated pipelines an important open challenge.

Even as NeRF/3DGS-based approaches promise photorealistic reconstructions, they remain computationally demanding, especially for large-scale underwater surveys. Processing thousands of images from extensive sites, such as reefs spanning hundreds of meters, is far beyond the typical usage scenario in small object scans or single-room reconstructions. Practical considerations include: *Data Collection Overlap*: Achieving consistent coverage in a turbid environment can be challenging. If certain areas are poorly visible or have drastically different color casts, the optimization might fail. *Long Training Times*: While faster NeRF variants exist, training can still take hours or days for high-resolution scenes. 3DGS training time is shorter, but generating the initial point cloud,

generally through COLMAP, still takes hours. Real-time or near-real-time feedback is typically unattainable in standard setups. *Refraction and Partial Occlusions*: The integrated assumption that each pixel ray corresponds to a linear path in Euclidean space is flawed with thick camera housings or highly refractive ports. Additional geometry modeling is required to handle these complexities.

Despite these hurdles, NeRFs/3DGSs and their successors represent a key direction for future underwater 3D modeling, given their ability to produce physically consistent volumetric reconstructions that merge geometry with realistic appearance.

For dynamic scenes, however, the interaction with front-end enhancement remains delicate. Dynamic NeRF and 3DGS backends need temporally stable cues to separate moving objects, marine snow, and illumination fluctuations; front-end enhancement helps only when it is disturbance-aware and temporally consistent. Independent per-frame beautification can instead erase motion cues or hallucinate temporally inconsistent structure, pushing enhancement and dynamic reconstruction to work at cross-purposes rather than synergistically.

4.5 Hybrid and Multi-Sensor Systems

While the preceding subsections focus primarily on optical data and neural or geometric reconstructions, there is a growing trend toward fusing optical imagery with other sensor modalities. Sonar or acoustic cameras can provide robust wide-area scans even in turbid conditions, albeit at reduced resolution. Short-range lasers or structured light can yield accurate depth near the camera. By merging these complementary datasets, reconstructions can become both more extensive (covering large areas in rough resolution) and more detailed (where optical data is available, it refines geometry and color).

Examples of multi-sensor approaches might incorporate:

- **Acoustic Bathymetry + Photogrammetry**: Large-scale mapping using side-scan or multibeam sonar plus local photogrammetry for high-detail in key areas (Łacka and Łubczonek 2024).
- **Forward-Looking Sonar + Visual SLAM**: Robotics platforms that rely on sonar for broad obstacle detection, combined with visual-inertial SLAM for fine-scale corridor or hull inspection (Rahman et al. 2019; Cheng et al. 2022; Zhang et al. 2024).
- **RGB-D Fusion**: Underwater variants of Kinect-like sensors or scanning lasers to collect partial depth maps (Anwer et al. 2017; Lu et al. 2017), integrated into a more general volumetric or point-based pipeline.

Though each sensor type introduces unique calibration complexities (particularly with refraction), the synergy can mitigate the classic photogrammetry pitfalls (e.g., lack of features, severe scattering).

4.6 Discussion and Open Challenges

This subsection focuses on challenges that remain specific to underwater reconstruction itself. Shared issues such as supervision scarcity, benchmark coverage, and cross-domain evaluation are synthesized later in subsection 6.1 and subsection 6.3; here the emphasis is on geometry-specific failure modes and deployment bottlenecks.

Refraction Modeling. Explicitly handling the multi-layer interfaces in camera housings or free-floating cameras is essential for accurate geometry. Classical SfM/MVS pipelines, COLMAP-style calibration, vanilla NeRF, vanilla 3DGS, and most learned MVS methods still inherit pinhole-like or central-camera assumptions by default. In underwater deployments, that approximation is most fragile for flat-port housings, while dome-port systems can sometimes be closer to a single-viewpoint model if alignment is good. Methods that incorporate refractive calibration or ray-tracing through flat ports, curved domes, or multiple media transitions can reduce systematic distortions, but require more complex solvers (Sedlazeck et al. 2009; Diamanti and Øyvind Ødegård 2024). Underwater neural renderers such as SeaThru-NeRF, WaterHE-NeRF, SeaSplat, and WaterSplatting partly move in this direction by coupling rendering with medium-aware optics (Levy et al. 2023; Zhou et al. 2025; Yang et al. 2025; Li et al. 2025), yet a robust and widely adoptable solution for metric field use is still missing.

Domain Adaptation and Robust Learning. For reconstruction, robust learning matters because appearance shifts must be handled without breaking multi-view consistency. Domain adaptation, synthetic-to-real simulation, and geometry-aware self-supervision are therefore useful when they stabilize correspondence and depth across viewpoints rather than merely improving isolated views. Broader shared issues of data scarcity, supervision, and evaluation are discussed later in [subsection 6.1](#) and [subsection 6.3](#).

Dynamic Scene Reconstruction. Many oceanic scenes contain dynamic elements, from small fish to shifting vegetation. While some tasks focus on static structures (like coral reefs, ship hulls, or archaeological remains), others may explicitly aim to capture dynamic processes—e.g., ecological interactions or pipeline flow. Generalizing approaches such as 4D Gaussian Splatting or dynamic NeRF to handle partial occlusions, swirling particulates, or flickering illumination remains challenging.

Efficient Rendering and Interactivity. NeRF training times can be lengthy, while classical SfM can be slow or memory-intensive for large sites. In operational contexts—e.g., an ROV exploring a deep shipwreck—an interactive interface could provide immediate feedback on coverage or areas needing more data. Achieving near real-time or incremental updates for underwater 3D scenes requires optimizing every stage of the pipeline: from feature matching and bundle adjustment to volumetric or splat-based rendering. This also means distinguishing genuinely deployable online tools from offline post-processing pipelines: many high-performing enhancement, NeRF, and 3DGS systems remain better suited to post-dive reconstruction than to frame-by-frame robotic navigation.

Enhanced Physical Modeling. Both classical photogrammetry and neural approaches can benefit from better integration of the physical laws governing underwater light transport. For instance, embedding the Jaffe-McGlamery model ([McGlamery 1980](#)) or the revised model ([Akkaynak and Treibitz 2018](#)) into cost functions might more accurately tease apart geometry from color attenuation. Similarly, scattering phenomena could be parameterized for each camera viewpoint based on distance or angle, significantly boosting reconstruction fidelity in murky or partially lit conditions.

Towards Autonomous Large-Scale Survey. A key ambition is to perform wide-area mapping of underwater environments autonomously at high resolution. Potentially, a swarm of AUVs or ROVs could coordinate to gather photometric data from multiple vantage points, merging the partial reconstructions. This could yield a holistic map of reefs or canyons spanning kilometers. Achieving stable stitching of partial reconstructions from many vantage points or vehicles, each with its own dynamic lighting, remains a complex challenge, especially if there is little ground truth or fixed reference.

To conclude the methodological review before benchmarking, [Table 12](#) summarizes the main strengths, limitations, and best-use scenarios of the principal UIE and 3D reconstruction families, together with their typical impact on downstream geometry.

5 Pipeline-Level Evaluation

This section summarizes public dataset context, evaluation criteria, and pipeline-level case studies for underwater 3D reconstruction. It is intended as a review-oriented evaluation discussion rather than a claim of a new standalone benchmark contribution. We focus on two practical reconstruction settings: (1) direct reconstruction without enhancement and (2) a two-stage pipeline that first enhances underwater images and then reconstructs geometry. Model-level reported results for existing underwater NeRF/3DGS methods are discussed earlier in [subsection 4.4](#); here the comparisons are included to show how different image-processing choices affect geometry, correspondence quality, and visible failure modes.

We rely on public datasets wherever possible. Publicly accessible underwater 3D scene datasets remain scarce, as summarized in [Table 13](#); many contain only a small number of scenes or images per scene, which in turn limits how broadly any comparison can be generalized. The most commonly used evaluation criteria for NeRF-based and 3DGS-based methods are based on the visual quality of image reconstruction (also see [Section 3.4.1](#)). Rendering quality is computed using full-reference metrics such as PSNR, SSIM, and LPIPS ([Zhang et al. 2018](#)). Other metrics for 3D modeling are also used, including feature matching success rate, Absolute Trajectory Error (such as that used by [Malyugina et al. \(2025\)](#)), Relative Pose Error (RPE), Chamfer Distance (CD) measuring the dissimilarity between two finite point sets, Cloud-to-Cloud (C2C) distance, F-score combining precision and recall

Table 12 Summary of major underwater enhancement and reconstruction method families.

Family	Core idea	Strengths	Limitations	Best-use scenario	Impact on 3D reconstruction
Classical UIE	Histogram, Retinex, DCP/ASM-style priors, and fusion heuristics	Fast, interpretable, light-weight, often easy to deploy	Limited robustness, brittle assumptions, framewise inconsistency	On-board preview, moderate distortions, shallow-water correction	Can improve feature visibility, but may also amplify colour bias or mismatch views
Learning-based UIE	CNN, GAN, transformer, Mamba, and diffusion restoration	Strong visual quality and data-driven adaptation	Data hungry, risk of hallucination or domain overfitting	Complex appearance distortions with sufficient training data	Helpful when structural fidelity is preserved; harmful if geometry cues are altered
Physics-guided UIE	Combines deep models with IFM or revised-IFM constraints	Better physical plausibility and cross-water robustness	Requires medium assumptions or auxiliary priors	Real scenes with limited labels and strong water-medium effects	Often safer for downstream geometry because enhancement remains physically constrained
Photogrammetry	SfM/MVS with calibration, matching, bundle adjustment, and meshing	Mature, interpretable geometry, low-cost capture	Sensitive to low contrast, refraction, and repetitive texture	Static scenes with sufficient overlap and calibration	Strong geometry when correspondences are reliable; deteriorates rapidly under degraded imagery
NeRF	Implicit volumetric scene representation with differentiable rendering	High-quality novel views and joint geometry/appearance learning	Slow training, high compute, difficult explicit geometry extraction	Medium-sized scenes prioritizing view synthesis and appearance fidelity	Can absorb enhancement and medium modeling into one optimizer, but requires stable supervision
3DGS	Explicit Gaussian representation with fast rasterization	Real-time rendering, rapid convergence, strong visual detail	Needs robust initialization and additional modeling for underwater physics	Interactive rendering and faster reconstruction loops	Benefits strongly from medium-aware splatting and restoration-aware priors
Integrated enhancement + reconstruction	Joint or tightly coupled restoration and geometry estimation	Better correspondence preservation than naive pre-enhancement and more faithful rendering than degraded-only training	Higher modeling complexity and stronger data/optimization requirements	Challenging scenes where appearance degradation directly harms geometry	Most promising route for closing the UIE-to-3D loop under real underwater conditions

of reconstructed points, and dynamic point tracking metrics, including occlusion accuracy (OA) measuring the binary accuracy of occlusion predictions, and average Jaccard (AJ) evaluating tracking and occlusion prediction jointly (Doersch et al. 2023).

Table 13 Summary of publicly available underwater 3D scene datasets and their key characteristics “Total # Images” refers to the total number of raw images in each dataset.

Dataset	Year	# Scenes	Total # images	Add. Info.	Resolution	Download
UWBund (Skinner et al. 2017)	2016	1	32	None	1K	Link
SeaThru (Levy et al. 2023)	2023	4	88	Pose	1K	Link
NUSR (Tang et al. 2024)	2024	4	82	Motion mask, Pose	1K – 2K	Link
BVI-Coral (Anantrasirichai 2024)	2024	23	>2000	None	1K	Link
S-UW (Wang et al. 2025)	2024	4	96	Pose	1K	Link
Submerged3D (Jiang et al. 2025a)	2025	4	80	Pose	1K	Link

5.1 Pipeline-Level Reconstruction Case Studies

As an illustrative baseline setting, we consider reconstruction pipelines that do not explicitly correct underwater degradations before geometry estimation. The purpose is not to claim a new baseline benchmark, but to make the effect of later enhancement-aware pipelines easier to interpret.

Photogrammetry. We employ COLMAP (Schonberger and Frahm 2016) to reconstruct an underwater scene. The input images, along with their corresponding depth and normal maps, are presented in Figure 20. Snapshots of the point clouds and the reconstructed 3D meshes are shown in Figure 21.

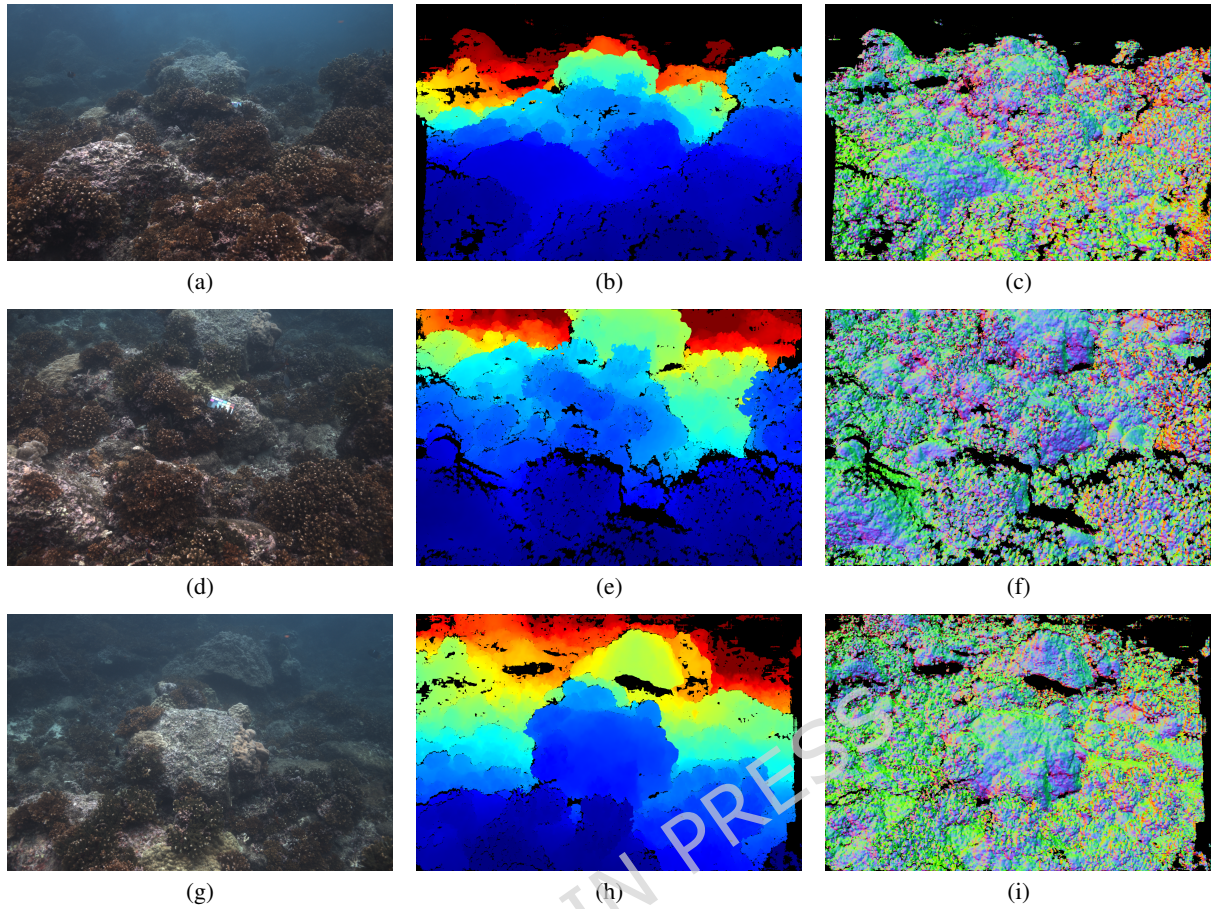


Fig. 20 Visualization of input images (a) and their corresponding depth map (b) and normal map estimated with SfM (c). The input images are from SeaThru (Levy et al. 2023)

We observe that when the input images lack sufficient coverage across diverse camera viewpoints, the reconstructed scene exhibits significant missing regions. This issue is particularly evident in Poisson surface reconstruction, where the absence of viewpoints from different angles leads to incomplete and fragmented surfaces. The Delaunay-based reconstruction, while more structurally connected, also suffers from irregularities due to the limited input perspectives. These results highlight the importance of capturing a well-distributed set of images to ensure a more complete and accurate 3D reconstruction.

NeRF. Instant-NGP (Müller et al. 2022a) is a NeRF variant optimized for real-time rendering. As shown in Figure 24(a), when applied directly to degraded underwater imagery, it outperforms traditional photogrammetry by enabling high-quality, view-dependent novel view synthesis with smooth interpolation and fine detail preservation. However, noticeable floater artifacts appear around scene boundaries, typically caused by inaccuracies in the estimated depth and density fields, particularly in regions where input observations are sparse or inconsistent. Moreover, due to the turbid water medium and light absorption, the reconstructed scene suffers from color cast, diminished visibility of distant objects, and reduced overall brightness.

Enhancement + 3D Reconstruction. To mitigate the issues arising from quality degradation in underwater imagery, Malyugina et al. (2025) first apply image enhancement models to improve the quality of raw underwater inputs—such as removing marine snow—and subsequently use the enhanced images to reconstruct dense 3D scenes. The simplified diagram of this two-stage pipeline is shown in Figure 22, which is although not end-to-end, but can effectively improve the visibility on the reconstruction.

Furthermore, Malyugina et al. (2025) observed that different enhancement methods can strongly influence both the number and the spatial distribution of features extracted during 3D reconstruction. Ideally, feature points should be concentrated around stable object regions, as illustrated in Figure 23 (right). More detected features,

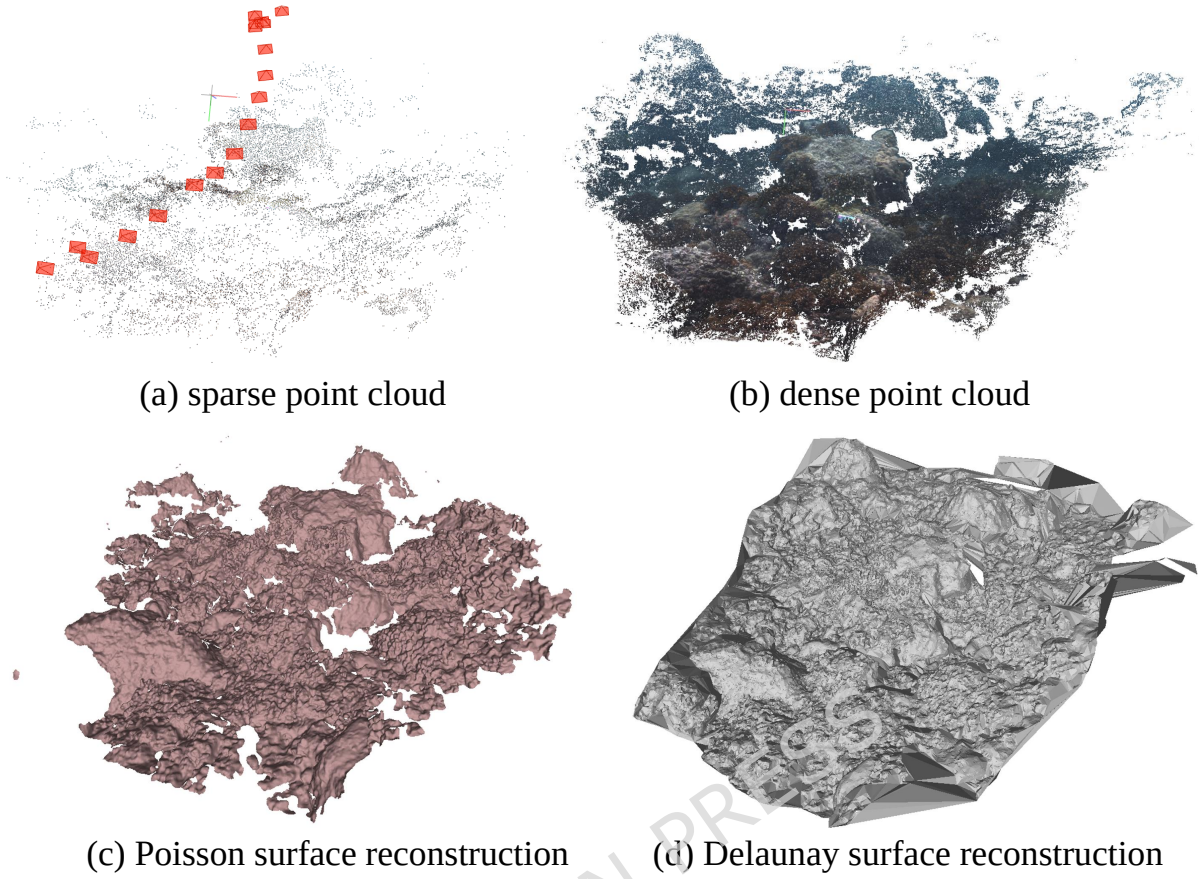


Fig. 21 Snapshots of the sparse point cloud (a), dense point cloud (b), Poisson surface reconstruction (c) and Delaunay surface reconstruction (d) of an underwater scene

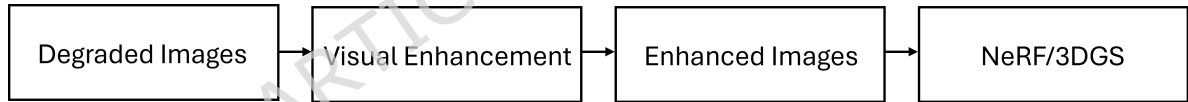


Fig. 22 Illustration of two-stage Enhancement + 3D reconstruction pipeline.

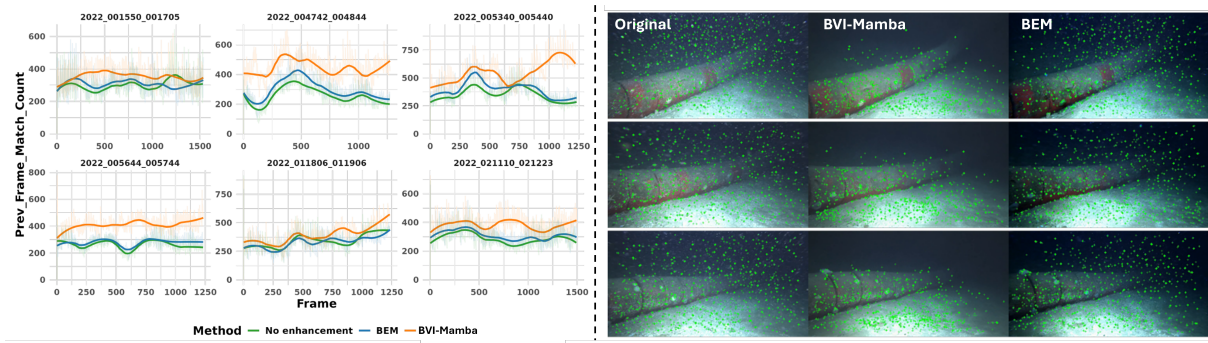


Fig. 23 Comparison of feature matching in 3D reconstruction using raw images versus enhanced images produced by BVI-Mamba (Huang et al. 2025) and BEM (Huang et al. 2026). (left) Number of frame-to-frame feature matches (higher values indicate more informative features for SLAM); curves are smoothed for clarity. (right) Example frames with detected feature points overlaid in green. The images are from Malyugina et al. (2025).

however, do not automatically imply more geometrically reliable correspondences: strong enhancement can also

create unstable gradients, hallucinated texture, or colour inconsistencies across views that fail later geometric verification. In practice, mild structure-preserving enhancement can improve matchability in low-contrast regions, whereas aggressive per-frame enhancement may hurt pose estimation or bundle adjustment even if the visual result appears sharper. This observation is consistent with recent bridge studies that explicitly examine how enhancement and visualisation choices affect 3D outcomes. [Vrochidis et al. \(2025a\)](#) report that underwater image enhancement choices measurably influence model quality, while [Vrochidis et al. \(2025b\)](#) show that even post-reconstruction colour-map selection can change the interpretability of 3D structures. Together, these studies reinforce that enhancement and visualisation are not merely cosmetic add-ons: they influence correspondence quality, surface readability, and ultimately the usefulness of the reconstructed model.

By comparing [Figure 24](#) (a) with (b), we observe that the two-stage reconstruction pipeline can effectively reduce color cast and haze effects compared to pure 3D reconstruction approaches, and in this example there are fewer floater artifacts. However, because the enhancement method is applied to individual images, slight colour differences can still lead to image misalignment or local inaccuracies. This is why some reconstruction pipelines still prefer raw or only lightly corrected images for camera pose estimation, while reserving stronger enhancement for later dense reconstruction or visualisation stages.

Overall, the visual-enhancement-plus-3D-reconstruction paradigm is a viable solution that can address several challenges caused by image degradation. However, it introduces additional development and deployment overhead, and its benefit depends on whether enhancement preserves geometry-relevant cues instead of only producing perceptually stronger images. The generalization capability of enhancement models is often limited by the relatively small size and domain specificity of available training datasets, which means that their effectiveness may vary across different scenarios. This typically requires extra manual effort to select or train enhancement models tailored to each environment. Looking ahead, this technical pipeline would benefit from addressing two key challenges: 1) developing end-to-end frameworks that jointly perform visual enhancement and 3D reconstruction, and 2) improving the adaptability of enhancement models, e.g., through scene-aware mechanisms or test-time optimization, to enhance generalization and avoid the need for multiple task- or scene-specific enhancement models.

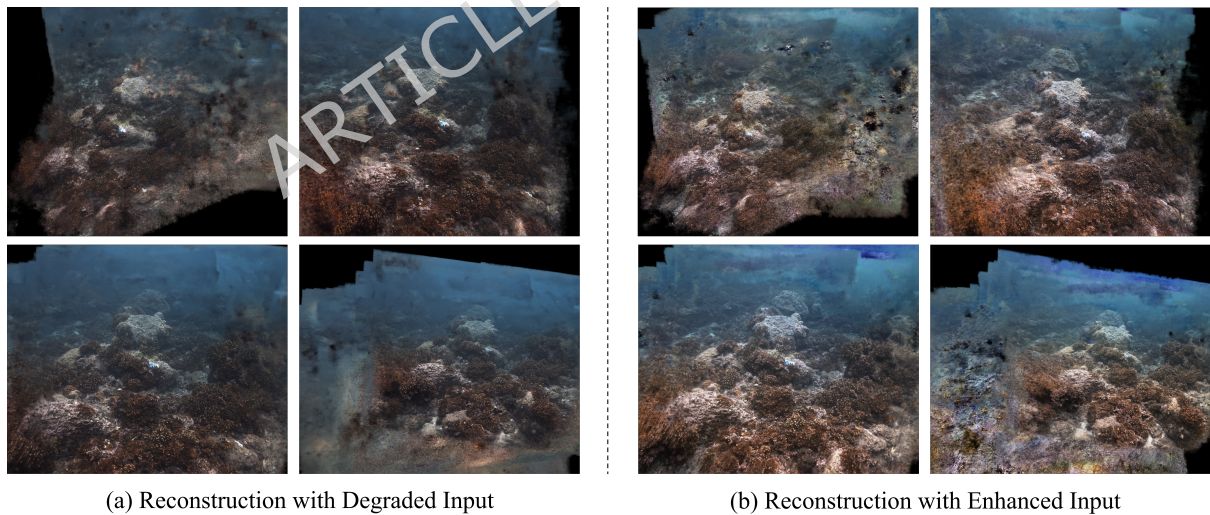


Fig. 24 Comparison of reconstructed underwater scenes produced by InstantNGP ([Müller et al. 2022a](#)) without applying any visual enhancement techniques (a) and with enhancement incorporated into the pipeline (b).

The practical implication is that low-contrast scenes often benefit most from mild, structure-preserving correction, while pose-estimation-critical stages may still prefer raw or lightly corrected imagery together with explicit physical or refractive modeling. Stronger enhancement is usually safer later in offline dense reconstruction, novel-view rendering, or operator-facing visualization, once camera geometry has already been stabilized.

6 Overall challenges and future work

6.1 Cross-cutting Data, Supervision, and Deployment Challenges

Many remaining obstacles are shared by enhancement and reconstruction rather than belonging to one stage alone. The most persistent problem is still the lack of reliable supervision: paired clean underwater targets are scarce, synthetic data reproduce only part of real water behaviour, and domain shift across turbidity, illumination, depth, and site conditions remains severe. Unlike some other restoration settings, underwater data vary strongly with local water properties and dynamic particulate content, which is why models trained in one environment often degrade when transferred to another (Lin et al. 2024; Li et al. 2018; Xiao et al. 2023; Akkaynak and Treibitz 2019b; Yi et al. 2025; Wang et al. 2025).

These shared supervision issues become even more difficult when enhancement, video analysis, and 3D reconstruction must operate together. Multi-view reconstruction requires appearance correction that remains stable across viewpoints, while long-duration video demands temporal consistency rather than isolated frame improvement. Recent methods such as Malyugina et al. (2025); Liu et al. (2025); Huang et al. (2025) show that progress increasingly depends on coupling restoration targets with scene dynamics, viewpoint consistency, and field robustness. At the deployment level, long videos, synchronised multi-camera capture, and high-resolution survey data are still expensive to acquire, which helps explain why UIE video methods remain relatively limited (Li and Anantrasirichai 2025) and why underwater 3D reconstruction is still dominated by static-scene settings despite encouraging progress on dynamic reconstruction in clear media (Yang et al. 2024, 2025).

Reducing the sim-to-real gap will require more than simply enlarging synthetic datasets. Promising strategies include broader randomization of water parameters during simulation, synthetic-to-real curricula, pseudo-paired supervision, and more realistic generative data creation for turbidity, lighting flicker, and marine-snow patterns. WaterGAN-style simulation remains influential for this reason (Li et al. 2018), but future datasets will likely need richer environmental randomization and more realistic dynamic clutter if models are to generalize reliably in field deployment.

The central difficulty is that controlled tank or synthetic captures still under-represent the field conditions that most disrupt deployment, including severe turbidity, non-uniform caustics, drifting particulate clouds, and camera-light interactions that vary along a survey trajectory. Bridging this gap is therefore as much about environmental realism and temporal variability as about the number of training samples.

6.2 Foundation models for underwater imagery

Beyond these shared bottlenecks, foundation models are best viewed as a forward-looking tool direction rather than another repetition of the challenge discussion above.

Foundation models (FMs) are large-scale models trained on diverse datasets, typically through self-supervised learning, and can be adapted or fine-tuned for a wide range of downstream tasks. Their development has been driven by rapid advances in AI-oriented computational power and they appear to be particularly well-suited for domains rich in data but lacking ground-truth annotations, including underwater applications. However, to date, MarineInst (Zheng et al. 2024) is the only foundation model developed specifically for marine applications. It's built upon Segment Anything Model (SAM) (Kirillov et al. 2023), trained on MarineInst20M, a large-scale dataset constructed from three main sources: (1) existing public marine and underwater datasets, (2) manually collected images from public and private datasets as well as YouTube videos, and (3) publicly available Internet images. MarineInst provides text-image matching (instance captioning) and instance segmentation capabilities. MarineInst shows strong potential as a foundation model. Its downstream tasks, however, perform well mainly on high-level computer vision applications, such as scene understanding and segmentation. Its usability for low-level tasks like enhancement remains uncertain, particularly for 3D reconstruction, as MarineInst was not trained to understand geometry.

Some underwater image enhancement and 3D reconstruction methods leverage foundation models (FMs) trained on natural images and text. These approaches use the output features and prior knowledge from such

models, based on the assumption that FMs trained on large datasets have learned rich patterns and representations of visual and textual information, which can potentially generalize to underwater imagery. For example, Wang et al. (2025) employ SAM to separate foreground and background regions and apply color correction separately. DreamSea (Zhang et al. 2025) exploits Depth Anything v2 (Yang et al. 2024) and DINOv2 (Oquab et al. 2024) to extract depth and other feature information for 3D Gaussian Splatting (3DGS). Similarly, SWAGSplatting uses BLIPo3 (Chen et al. 2025) to identify and capture regions of interest. Many approaches also utilize CLIP (Contrastive Language-Image Pre-training) (Radford et al. 2021) to generate captions for downstream tasks such as object detection and scene understanding. This approach could potentially be extended to enhancement and 3D reconstruction, similar to those developed for clear-medium scenes (Zhou et al. 2024; Wei et al. 2025).

6.3 Perspectives on Datasets, Modeling Tools, and Evaluation

Future progress will depend not only on better architectures but also on broader task coverage in data collection. Current benchmarks remain dominated by still images or short clips, whereas practical deployments in surveillance, fish monitoring, intervention support, and long-baseline reconstruction require longer videos, richer dynamic content, paired and unpaired video benchmarks, denser task annotations, and more comprehensive coverage across harbours, reefs, lakes, and deep-water domains. Bridging enhancement, video analysis, and 3D reconstruction will therefore require benchmark design that couples restoration targets with motion, semantics, and geometry.

In terms of modelling, graph-based learning is a promising complement to convolutional, transformer, Mamba, and diffusion backbones when labels are sparse or relational structure matters. Recent surveys by Ponzi and Napoli (2025) and Sadasivan et al. (2025) highlight that graph neural networks are well suited to modelling interactions, non-Euclidean structure, and long-range dependencies with modest annotation budgets. For underwater vision, such priors could be valuable for fish schools, moving-object associations across frames, multi-view correspondence graphs, and cross-sensor fusion, complementing recent graph-based underwater motion analysis methods (Kapoor et al. 2024, 2025).

Evaluation protocols also require broader reporting. A single ranking score is often insufficient because a method may improve one challenge while degrading another; recent discussions by Piérard and Van Droogenbroeck (2025a,b) likewise caution against relying on one summary statistic alone. Future benchmarks should therefore report restoration quality, temporal consistency, downstream utility, and cross-domain robustness side by side, rather than collapsing all performance into one leaderboard.

For integrated enhancement-and-reconstruction settings, this broader reporting should also include geometry- and trajectory-aware criteria such as Chamfer or Cloud-to-Cloud distance, ATE/RPE, and correspondence-stability indicators wherever possible. Without such multi-domain evaluation, it remains too easy for a method to look strong on image metrics while still degrading mapping or localization reliability.

For engineering deployment, the same principle applies to efficiency reporting: latency, memory footprint, and platform suitability should be documented whenever a method is proposed for robotic operation. A model that performs well in offline post-processing may still be unusable for navigation or online inspection if it requires seconds per frame or high-end GPUs.

6.4 Ethical Issues and Bias

As underwater image enhancement and 3D reconstruction technologies become increasingly sophisticated, they raise important ethical considerations that demand careful attention from researchers, developers, and end users. A fundamental challenge lies in distinguishing legitimate enhancement from misleading manipulation. Aggressive enhancement or reconstruction methods may introduce artifacts, alter colors unnaturally, or generate content that misrepresents actual underwater scenes, risks that are particularly acute when techniques involve generative AI, such as GANs and diffusion models. In scientific applications including marine biology surveys, archaeological documentation, and environmental monitoring, such distortions can compromise research integrity and lead to incorrect conclusions about ecosystem health, species behavior, or site conditions.

These authenticity concerns are compounded by systematic biases in the datasets used to train enhancement algorithms. Most existing methods rely on image formation models trained on limited datasets that fail to represent the full diversity of underwater environments. Current benchmark datasets are predominantly captured in specific geographic regions, water types, and depth ranges, meaning that models trained on coral reef imagery from tropical waters often perform poorly in murky harbor environments, kelp forests, or high-latitude conditions. This geographic and environmental bias can produce enhancements that work well for some scenarios while generating unrealistic or misleading results for others.

Beyond technical limitations, important questions surrounding data ownership, consent, and privacy arise when imagery is collected in protected habitats, cultural heritage sites, or industrial contexts. High-resolution reconstructions of archaeological shipwrecks could inadvertently facilitate looting, while detailed seafloor mapping might reveal sensitive information about underwater infrastructure or ecologically vulnerable areas. Similarly, images captured in marine protected areas or private facilities raise questions about appropriate data sharing and access control.

To address these interconnected challenges, future work should prioritize ethical dataset governance and transparent methodology. This includes detailed documentation of data provenance, balanced representation of environmental diversity across training datasets, standardized disclosure of model limitations and enhancement boundaries, and clear protocols for handling sensitive imagery from protected or culturally significant sites. By proactively addressing these ethical dimensions, the underwater imaging community can ensure that these powerful technologies serve scientific understanding and environmental stewardship while minimizing potential harms.

7 Conclusions

Underwater imaging is essential for scientific exploration, industrial applications, and environmental conservation, covering a broad range of fields including marine biology, archaeology, geological surveying, and infrastructure inspection. It helps manage resources, monitor ecosystems, and assess the condition of subsea infrastructure like pipelines and offshore platforms. Advancements in technology allow for high-quality images crucial for studying marine life, creating detailed 3D models of submerged archaeological sites, and ensuring the safe operation of industrial facilities under challenging visibility conditions. These efforts are crucial in tracking environmental changes and supporting resource exploration by providing precise mappings of the seafloor, thus minimizing risks and operational costs.

The review begins by addressing the unique challenges of underwater environments and outlines its scope, including discussions on image enhancement and 3D reconstruction pathways. We describe the physics of underwater light propagation and image formation, setting the stage for an exploration of various visual enhancement methods, both traditional and data-driven, and their applicability to underwater scenes. The review further elaborates on different 3D reconstruction techniques tailored for underwater use, including photogrammetry, NeRF, and 3D Gaussian Splatting, discussing their motivations, methodologies, and specific challenges. It concludes with a benchmarking discussion on these methods, emphasizing the need for enhancement integration to achieve accurate underwater 3D reconstructions.

Future research will likely delve deeper into physically correct light-transport modeling, large-scale real-time systems, domain adaptation to address data scarcity, and multi-sensor integration, ultimately broadening underwater exploration, scientific study, and industrial deployment.

References

- Ancuti, C.O., Ancuti, C.: Single Image Dehazing by Multi-Scale Fusion. *IEEE Transactions on Image Processing* **22**(8), 3271–3282 (2013) <https://doi.org/10.1109/TIP.2013.2262284>
- Anwer, A., Azhar Ali, S.S., Khan, A., Mériaudeau, F.: Underwater 3-D Scene Reconstruction Using Kinect v2 Based on Physical Models for Refraction and Time of Flight Correction. *IEEE Access* **5**, 15960–15970 (2017) <https://doi.org/10.1109/ACCESS.2017.2733003>
- Ancuti, C.O., Ancuti, C., De Vleeschouwer, C., Bekaert, P.: Color Balance and Fusion for Underwater Image Enhancement. *IEEE Transactions on Image Processing* **27**(1), 379–393 (2018) <https://doi.org/10.1109/TIP.2017.2759252>
- Ancuti, C., Ancuti, C.O., Haber, T., Bekaert, P.: Enhancing underwater images and videos by fusion. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 81–88 (2012). <https://doi.org/10.1109/CVPR.2012.6247661>
- Abdul Ghani, A.S., Mat Isa, N.A.: Enhancement of low quality underwater image through integrated global and local contrast correction. *Applied Soft Computing* **37**, 332–344 (2015) <https://doi.org/10.1016/j.asoc.2015.08.033>
- Abdul Ghani, A.S., Mat Isa, N.A.: Automatic system for improving underwater image contrast and color through recursive adaptive histogram modification. *Computers and Electronics in Agriculture* **141**, 181–195 (2017) <https://doi.org/10.1016/j.compag.2017.07.021>
- Anwar, S., Li, C.: Diving deeper into underwater image enhancement: A survey. *Signal Processing: Image Communication* **89**, 115978 (2020)
- Anantrasirichai, N.: BVI-Coral: Underwater scenes for 3D reconstruction. Zenodo (2024). <https://doi.org/10.5281/zenodo.11093417> . <https://doi.org/10.5281/zenodo.11093417>
- Akkaynak, D., Treibitz, T.: A revised underwater image formation model. In: the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6723–6732 (2018)
- Akkaynak, D., Treibitz, T.: Sea-Thru: A method for removing water from underwater images. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1682–1691 (2019)
- Akkaynak, D., Treibitz, T.: Sea-thru: A method for removing water from underwater images. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1682–1691 (2019)
- Aubram, D.: An Arbitrary Lagrangian-Eulerian Method for Penetration Into Sand at Finite Deformation. Shaker, Aachen, ??? (2013). <https://doi.org/10.14279/depositonce-3958>
- Anantrasirichai, N., Zhang, F., Bull, D.: Artificial intelligence in creative industries: Advances prior to 2025. *arXiv preprint arXiv:2501.02725* (2025)
- Boudiaf, A., Guo, Y., Ghimire, A., Werghi, N., De Masi, G., Javed, S., Dias, J.: Underwater image enhancement using pre-trained transformer. In: International Conference on Image Analysis and Processing, pp. 480–488 (2022)
- Bryson, M., Johnson-Roberson, M., Pizarro, O., Williams, S.B.: True color correction of autonomous underwater vehicle imagery. *Journal of Field Robotics* **33**(6), 853–874 (2016)
- Berman, D., Levy, D., Avidan, S., Treibitz, T.: Underwater single image color restoration using haze-lines and a new quantitative dataset. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2020)
- Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P., Hedman, P.: Mip-NeRF 360: Unbounded anti-aliased neural radiance fields. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),

- pp. 5460–5469 (2022). <https://doi.org/10.1109/CVPR52688.2022.00539>
- Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P., Hedman, P.: Zip-nerf: Anti-aliased grid-based neural radiance fields. In: 2023 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 19640–19648 (2023). <https://doi.org/10.1109/ICCV51070.2023.01804>
- Bazeille, S., Quidu, I., Jaulin, L., Malkasse, J.-P.: Automatic underwater image pre-processing. In: Proceedings of the SEA TECH WEEK Caractérisation du Milieu Marin (CMM), Brest, France (2006)
- Bing, X., Ren, W., Tang, Y., Yen, G.G., Sun, Q.: Domain adaptation for in-air to underwater image enhancement via deep learning. *IEEE Transactions on Emerging Topics in Computational Intelligence* (2023)
- Banerjee, S., Sanyal, G., Ghosh, S., Ray, R., Shome, S.N.: Elimination of Marine Snow effect from underwater image - An adaptive probabilistic approach. In: 2014 IEEE Students' Conference on Electrical, Electronics And Computer Science, pp. 1–4 (2014). <https://doi.org/10.1109/SCEECS.2014.6804438>
- Badran, M., Torki, M.: Daut: Underwater image enhancement using depth aware u-shape transformer. In: IEEE International Conference on Image Processing, pp. 1830–1834 (2023)
- Berman, D., Treibitz, T., Avidan, S.: Diving into haze-lines: Color restoration of underwater images. In: Proc. British Machine Vision Conference (BMVC), vol. 1, p. 2 (2017)
- Chiang, J.Y., Chen, Y.-C.: Underwater Image Enhancement by Wavelength Compensation and Dehazing. *IEEE Transactions on Image Processing* **21**(4), 1756–1769 (2012) <https://doi.org/10.1109/TIP.2011.2179666>
- Cong, X., Gui, J., Hou, J.: Underwater organism color fine-tuning via decomposition and guidance. In: AAAI Conference on Artificial Intelligence, vol. 38, pp. 1389–1398 (2024)
- Cong, Y., Gu, C., Zhang, T., Gao, Y.: Underwater robot sensing technology: A survey. *Fundamental Research* **1**(3), 337–345 (2021) <https://doi.org/10.1016/j.fmre.2021.03.002>
- Chen, L., Jiang, Z., Tong, L., Liu, Z., Zhao, A., Zhang, Q., Dong, J., Zhou, H.: Perceptual underwater image enhancement with deep learning and physical priors. *IEEE Transactions on Circuits and Systems for Video Technology* **31**(8), 3078–3092 (2020)
- 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)
- Chen, L., Li, W., Yang, Q., Tong, L., Chen, E., Huang, B., Li, R.: MUIR: Mamba for underwater image rendering. In: 2024 4th International Conference on Machine Learning and Intelligent Systems Engineering (MLISE), pp. 172–177 (2024). <https://doi.org/10.1109/MLISE62164.2024.10674249>
- Carlevaris-Bianco, N., Mohan, A., Eustice, R.M.: Initial results in underwater single image dehazing. In: OCEANS 2010 MTS/IEEE SEATTLE, pp. 1–8 (2010). <https://doi.org/10.1109/OCEANS.2010.5664428>
- Chen, Y.-W., Pei, S.-C.: Domain adaptation for underwater image enhancement via content and style separation. *IEEE Access* **10**, 90523–90534 (2022)
- Chao, L., Wang, M.: Removal of water scattering. In: 2010 2nd International Conference on Computer Engineering and Technology, vol. 2, pp. 2–35239 (2010). <https://doi.org/10.1109/ICCET.2010.5485339>
- Cheng, C., Wang, C., Yang, D., Liu, W., Zhang, F.: Underwater localization and mapping based on multi-beam forward looking sonar. *Frontiers in Neurorobotics* **15**, 801956 (2022) <https://doi.org/10.3389/fnbot.2021.801956>
- Chen, A., Xu, Z., Geiger, A., Yu, J., Su, H.: Tensorf: Tensorial radiance fields. *ECCV* (2022)
- Chen, J., Xu, Z., Pan, X., Hu, Y., Qin, C., Goldstein, T., Huang, L., Zhou, T., Xie, S., Savarese, S., Xue, L., Xiong,

- C., Xu, R.: BLIP3-o: A family of fully open unified multimodal models—architecture, training and dataset. arXiv preprint arXiv:2505.09568 (2025)
- Chen, L., Xu, Z., Wei, C., Xu, Y.: BDMUIE: Underwater image enhancement based on bayesian diffusion model. *Neurocomputing* **620**, 129274 (2025) <https://doi.org/10.1016/j.neucom.2024.129274>
- Chen, L., Xiong, Y., Zhang, Y., Yu, R., Fang, L., Liu, D.: Sp-seanerf: Underwater neural radiance fields with strong scattering perception. *Computers & Graphics* **123**, 104025 (2024) <https://doi.org/10.1016/j.cag.2024.104025>
- Cong, R., Yang, W., Zhang, W., Li, C., Guo, C.-L., Huang, Q., Kwong, S.: Pupan: Physical model-guided underwater image enhancement using gan with dual-discriminators. *IEEE Transactions on Image Processing* **32**, 4472–4485 (2023)
- Chen, X., Zhang, P., Quan, L., Yi, C., Lu, C.: Underwater image enhancement based on deep learning and image formation model. arXiv preprint arXiv:2101.00991 (2021)
- Drews Jr, P., Do Nascimento, E., Moraes, F., Botelho, S., Campos, M.: Transmission Estimation in Underwater Single Images. In: 2013 IEEE International Conference on Computer Vision Workshops, pp. 825–830. IEEE, Sydney, Australia (2013). <https://doi.org/10.1109/ICCVW.2013.113>
- DJI: DJI Terra. <https://enterprise.dji.com/dji-terra>. Accessed: 2026-03-20 (2026)
- Du, D., Li, E., Si, L., Zhai, W., Xu, F., Niu, J., Sun, F.: UIEDP: Boosting underwater image enhancement with diffusion prior. *Expert Systems with Applications* **259**, 125271 (2025) <https://doi.org/10.1016/j.eswa.2024.125271>
- Drews, P.L.J., Nascimento, E.R., Botelho, S.S.C., Montenegro Campos, M.F.: Underwater Depth Estimation and Image Restoration Based on Single Images. *IEEE Computer Graphics and Applications* **36**(2), 24–35 (2016) <https://doi.org/10.1109/MCG.2016.26>
- Desai, C., Tabib, R.A., Reddy, S.S., Patil, U., Mudanagudi, U.: Ruig: Realistic underwater image generation towards restoration. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 2181–2189 (2021)
- Doersch, C., Yang, Y., Vecerik, M., Gokay, D., Gupta, A., Aytar, Y., Carreira, J., Zisserman, A.: TAPIR: Tracking any point with per-frame initialization and temporal refinement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2023)
- Du, Y., Zhang, Y., Yu, H.-X., Tenenbaum, J.B., Wu, J.: Neural radiance flow for 4d view synthesis and video processing. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 14324–14334 (2021)
- Diamanti, E., Ødegård: Visual sensing on marine robotics for the 3d documentation of underwater cultural heritage: A review. *Journal of Archaeological Science* **166**, 105985 (2024) <https://doi.org/10.1016/j.jas.2024.105985>
- Emberton, S., Chittka, L., Cavallaro, A.: Hierarchical rank-based veiling light estimation for underwater dehazing. In: Proceedings of the British Machine Vision Conference 2015, pp. 125–112512. British Machine Vision Association, Swansea (2015). <https://doi.org/10.5244/C.29.125>
- Fattal, R.: Single image dehazing. *ACM Transactions on Graphics* **27**(3), 1–9 (2008) <https://doi.org/10.1145/1360612.1360671>
- Funt, B., Ciurea, F., McCann, J.: Retinex in Matlab (2000)
- Fabbri, C., Islam, M.J., Sattar, J.: Enhancing underwater imagery using generative adversarial networks. In: International Conference on Robotics and Automation, pp. 7159–7165 (2018)

- Fridovich-Keil, S., Yu, A., Tancik, M., Chen, Q., Recht, B., Kanazawa, A.: Plenoxels: Radiance Fields without Neural Networks. In: CVPR (2022)
- Fu, C., Liu, R., Fan, X., Zhou, T., Luo, Z., Zhang, L., Huang, X., Luo, Z.: Rethinking general underwater object detection: Datasets, challenges, and solutions. *Neurocomputing* **517**, 243–256 (2023) <https://doi.org/10.1016/j.neucom.2022.10.039>
- Fu, Z., Lin, H., Yang, Y., Chai, S., Sun, L., Huang, Y., Ding, X.: Unsupervised underwater image restoration: From a homology perspective. In: AAAI Conference on Artificial Intelligence, pp. 643–651 (2022)
- Fu, Z., Wang, W., Huang, Y., Ding, X., Ma, K.-K.: Uncertainty inspired underwater image enhancement. In: European Conference on Computer Vision, pp. 465–482 (2022)
- Fan, X., Wang, X., Ni, H., Xin, Y., Shi, P.: Water-adapted 3d gaussian splatting for precise underwater scene reconstruction. *Frontiers in Marine Science* **Volume 12 - 2025** (2025) <https://doi.org/10.3389/fmars.2025.1573612>
- Fu, X., Zhuang, P., Huang, Y., Liao, Y., Zhang, X.-P., Ding, X.: A retinex-based enhancing approach for single underwater image. In: 2014 IEEE International Conference on Image Processing (ICIP), pp. 4572–4576 (2014). <https://doi.org/10.1109/ICIP.2014.7025927>
- Gough, L., Azzarelli, A., Zhang, F., Anantrasirichai, N.: Aquanerf: Neural radiance fields in underwater media with distractor removal. In: IEEE International Symposium on Circuits and Systems (2025)
- Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces. In: Conference on Language Modeling (2024)
- Guo, X., Dong, Y., Chen, X., Chen, W., Li, Z., Zheng, F., Pun, C.-M.: Underwater Image Restoration via Polymorphic Large Kernel CNNs. In: 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 9731–9735 (2025). <https://doi.org/10.1109/ICASSP49660.2025.10890803> . <https://ieeexplore.ieee.org/document/10890803/>
- Gu, X., Fan, Z., Zhu, S., Dai, Z., Tan, F., Tan, P.: Cascade cost volume for high-resolution multi-view stereo and stereo matching. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2495–2504 (2020)
- Garcia-Garcia, B., Bouwmans, T., Rosales Silva, A.J.: Background subtraction in real applications: Challenges, current models and future directions. *Computer Science Review* **35**, 100204 (2020) <https://doi.org/10.1016/j.cosrev.2019.100204>
- Garg, D., Garg, N.K., Kumar, M.: Underwater image enhancement using blending of CLAHE and percentile methodologies. *Multimedia Tools and Applications* **77**(20), 26545–26561 (2018) <https://doi.org/10.1007/s11042-018-5878-8>
- Ghani, A.S.A., Isa, N.A.M.: Underwater image quality enhancement through composition of dual-intensity images and Rayleigh-stretching. In: 2014 IEEE Fourth International Conference on Consumer Electronics Berlin (ICCE-Berlin), pp. 219–220 (2014). <https://doi.org/10.1109/ICCE-Berlin.2014.7034265>
- Galdran, A., Pardo, D., Picón, A., Alvarez-Gila, A.: Automatic Red-Channel underwater image restoration. *Journal of Visual Communication and Image Representation* **26**, 132–145 (2015) <https://doi.org/10.1016/j.jvcir.2014.11.006>
- Gibson, K.B., Vo, D.T., Nguyen, T.Q.: An Investigation of Dehazing Effects on Image and Video Coding. *IEEE Transactions on Image Processing* **21**(2), 662–673 (2012) <https://doi.org/10.1109/TIP.2011.2166968>
- Guo, C., Wu, R., Jin, X., Han, L., Zhang, W., Chai, Z., Li, C.: Underwater Ranker: Learn which is better and how

- to be better. In: AAAI Conference on Artificial Intelligence, pp. 702–709 (2023)
- Guan, M., Xu, H., Jiang, G., Yu, M., Chen, Y., Luo, T., Zhang, X.: Diffwater: Underwater image enhancement based on conditional denoising diffusion probabilistic model. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **17**, 2319–2335 (2023)
- Guan, M., Xu, H., Jiang, G., Yu, M., Chen, Y., Luo, T., Song, Y.: WaterMamba: Visual State Space Model for Underwater Image Enhancement. *arXiv* (2024). <https://doi.org/10.48550/arXiv.2405.08419>
- Helan, S., Burie, J.-C., Bouwmans, T., Bazeille, S.: Object detection in underwater images. In: *Proceedings of the SEA TECH WEEK Caractérisation du Milieu Marin (CMM)*, Brest, France (2006)
- Huang, S.-C., Chen, B.-H., Wang, W.-J.: Visibility Restoration of Single Hazy Images Captured in Real-World Weather Conditions. *IEEE Transactions on Circuits and Systems for Video Technology* **24**(10), 1814–1824 (2014) <https://doi.org/10.1109/TCSVT.2014.2317854>
- Huang, Z., Li, J., Hua, Z., Fan, L.: Underwater image enhancement via adaptive group attention-based multiscale cascade transformer. *IEEE Transactions on Instrumentation and Measurement* **71**, 1–18 (2022)
- Huang, G., Lin, R., Li, Y., Bull, D., Anantrasirichai, N.: Bvi-mamba: video enhancement using a visual state-space model for low-light and underwater environments. In: *Machine Learning from Challenging Data 2025*, vol. 13460, pp. 74–81 (2025). SPIE
- Humbert, J., McMath, A., Robin, A., Espiau, B., Abadie, A.: The open-source camera trap for organism presence and underwater surveillance (OCTOPUS). *HardwareX* **15**, 00472 (2023) <https://doi.org/10.1016/j.ohx.2023.e00472>
- 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)
- Huang, S., Wang, K., Liu, H., Chen, J., Li, Y.: Contrastive semi-supervised learning for underwater image restoration via reliable bank. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 18145–18155 (2023)
- Huang, G., Wang, H., Qi, Z., Lu, W., Bull, D., Anantrasirichai, N.: From restoration to reconstruction: Rethinking 3d gaussian splatting for underwater scenes. *arXiv preprint arXiv:2509.17789* (2025)
- Huang, S.-C., Ye, J.-H., Chen, B.-H.: An Advanced Single-Image Visibility Restoration Algorithm for Real-World Hazy Scenes. *IEEE Transactions on Industrial Electronics* **62**(5), 2962–2972 (2015) <https://doi.org/10.1109/TIE.2014.2364798>
- Huang, G., Yang, Q., Qi, Z., Lin, R., Bull, D., Anantrasirichai, N.: Bayesian neural networks for one-to-many mapping in image enhancement. *Proceedings of the AAAI Conference on Artificial Intelligence* (2026)
- Hou, G., Zhao, X., Pan, Z., Yang, H., Tan, L., Li, J.: Benchmarking underwater image enhancement and restoration, and beyond. *IEEE Access* **8**, 122078–122091 (2020) <https://doi.org/10.1109/ACCESS.2020.3006359>
- Istenič, K., Gracias, N., Arnaubec, A., Escartín, J., Garcia, R.: Scale accuracy evaluation of image-based 3d reconstruction strategies using laser photogrammetry. *Remote Sensing* **11**(18), 2093 (2019) <https://doi.org/10.3390/rs11182093>
- Islam, M.J., Luo, P., Sattar, J.: Simultaneous enhancement and super-resolution of underwater imagery for improved visual perception. In: *16th Robotics: Science and Systems, RSS 2020* (2020)
- Islam, M.J., Xia, Y., Sattar, J.: Fast underwater image enhancement for improved visual perception. *IEEE Robotics and Automation Letters* **5**(2), 3227–3234 (2020) <https://doi.org/10.1109/LRA.2020.2974710>
- Islam, M.J., Xia, Y., Sattar, J.: Fast underwater image enhancement for improved visual perception. *IEEE Robotics and Automation Letters* **5**(2), 3227–3234 (2020)

- Jaffe, J.S.: Computer modeling and the design of optimal underwater imaging systems. *IEEE Journal of Oceanic Engineering* **15**(2), 101–111 (1990)
- Jiang, N., Chen, W., Lin, Y., Zhao, T., Lin, C.-W.: Underwater Image Enhancement With Lightweight Cascaded Network. *IEEE Transactions on Multimedia* **24**, 4301–4313 (2022) <https://doi.org/10.1109/TMM.2021.3115442>
- Jiang, Q., Chen, Y., Wang, G., Ji, T.: A novel deep neural network for noise removal from underwater image. *Signal Processing: Image Communication* **87**, 115921 (2020)
- Joshi, K.R., Kamathe, R.S.: Quantification of retinex in enhancement of weather degraded images. In: 2008 International Conference on Audio, Language and Image Processing, pp. 1229–1233 (2008). <https://doi.org/10.1109/ICALIP.2008.4590120>
- Jiang, Q., Kang, Y., Wang, Z., Ren, W., Li, C.: Perception-driven deep underwater image enhancement without paired supervision. *IEEE Transactions on Multimedia* (2023)
- Jiang, Z., Wang, H., Huang, G., Seymour, B., Anantrasirichai, N.: RUSplatting: Robust 3d gaussian splatting for sparse-view underwater scene reconstruction. In: Proceedings of the British Machine Vision Conference (BMVC) (2025). BMVA
- Jiang, Z., Wang, H., Huang, G., Seymour, B., Anantrasirichai, N.: SWAGSplatting: Semantic-guided water-scene augmented gaussian splatting. *arXiv:2509.00800* (2025)
- Ju, Y., Xiao, J., Zhang, C., Xie, H., Luo, A., Zhou, H., Dong, J., Kot, A.C.: Towards marine snow removal with fusing fourier information. *Information Fusion* **117**, 102810 (2025) <https://doi.org/10.1016/j.inffus.2024.102810>
- Jiang, J., Ye, T., Bai, J., Chen, S., Chai, W., Jun, S., Liu, Y., Chen, E.: Five a⁺ network: You only need 9k parameters for underwater image enhancement. *British Machine Vision Conference (BMVC)* (2023)
- Jiang, Q., Zhang, Y., Bao, F., Zhao, X., Zhang, C., Liu, P.: Two-step domain adaptation for underwater image enhancement. *Pattern Recognition* **122**, 108324 (2022)
- Kajiya, J.T.: The rendering equation. *ACM SIGGRAPH Computer Graphics* **20**(4), 143–150 (1986) <https://doi.org/10.1145/15886.15902>
- Khan, A., Ali, S.S.A., Malik, A.S., Anwer, A., Meriaudeau, F.: Underwater image enhancement by wavelet based fusion. In: 2016 IEEE International Conference on Underwater System Technology: Theory and Applications (USYS), pp. 83–88 (2016). IEEE
- Kapoor, M., Baghel, R., Subudhi, B.N., Jakhetiya, V., Bansal, A.: Domain adversarial learning towards underwater image enhancement. In: IEEE International Conference on Computer Vision Workshops, pp. 2241–2251 (2023)
- Kim, I., Choi, M., Kim, H.J.: UP-NeRF: Unconstrained Pose-Prior-Free Neural Radiance Fields. In: NeurIPS (2023)
- Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3D gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics* **42**(4) (2023)
- Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.* **42**(4), 139–1 (2023)
- Khan, M.R., Mishra, P., Mehta, N., Phutke, S.S., Vipparthi, S.K., Nandi, S., Murala, S.: Spectroformer: Multi-Domain Query Cascaded Transformer Network For Underwater Image Enhancement. In: 2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 1443–1452. IEEE, Waikoloa, HI, USA (2024). <https://doi.org/10.1109/WACV57701.2024.00148>
- Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo,

- W.-Y., Dollar, P., Girshick, R.: Segment anything. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 4015–4026 (2023)
- Khan, M.R., Negi, A., Kulkarni, A., Phutke, S.S., Vipparthi, S.K., Murala, S.: Phaseformer: Phase-based Attention Mechanism for Underwater Image Restoration and Beyond. In: 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 9618–9629 (2025). <https://doi.org/10.1109/WACV61041.2025.00931>. IEEE
- Khan, M.R., Negi, A., Kulkarni, A., Phutke, S.S., Vipparthi, S.K., Murala, S.: Phaseformer: Phase-based attention mechanism for underwater image restoration and beyond. In: 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 9618–9629 (2025). IEEE
- Kapoor, M., Prummel, W., Giraldo, J.H., Subudhi, B.N., Zakharova, A., Bouwmans, T., Bansal, A.: Graph-based moving object segmentation for underwater videos using semi-supervised learning. *Computer Vision and Image Understanding* **252**, 104290 (2025) <https://doi.org/10.1016/j.cviu.2025.104290>
- Kulhanek, J., Peng, S., Kukeleva, Z., Pollefeys, M., Sattler, T.: WildGaussians: 3D gaussian splatting in the wild. In: Proceedings of the 38th International Conference on Neural Information Processing Systems. NIPS '24. Curran Associates Inc., Red Hook, NY, USA (2025)
- Kapoor, M., Subudhi, B.N., Jakhetiya, V.: Principal graph neighborhood aggregation for underwater moving object detection. In: International Conference on Pattern Recognition (ICPR), pp. 398–412 (2024)
- Li, Y., Anantrasirichai, N.: Zero-TIG: Temporal consistency-aware zero-shot illumination-guided low-light video enhancement. In: Proceedings of the 33rd European Signal Processing Conference (EUSIPCO) (2025)
- Li, C., Anwar, S., Hou, J., Cong, R., Guo, C., Ren, W.: Underwater Image Enhancement via Medium Transmission-Guided Multi-Color Space Embedding. *IEEE Transactions on Image Processing* **30**, 4985–5000 (2021) <https://doi.org/10.1109/TIP.2021.3076367> arXiv:2104.13015 [cs]
- Li, C., Anwar, S., Hou, J., Cong, R., Guo, C., Ren, W.: Underwater image enhancement via medium transmission-guided multi-color space embedding. *IEEE Transactions on Image Processing* **30**, 4985–5000 (2021)
- Lin, R., Anantrasirichai, N., Huang, G., Lin, J., Sun, Q., Malyugina, A., Bull, D.: BVI-RLV: A fully registered dataset and benchmarks for low-light video enhancement. arXiv preprint arXiv:2407.03535 (2024)
- Li, C., Anwar, S., Porikli, F.: Underwater scene prior inspired deep underwater image and video enhancement. *Pattern Recognition* **98**, 107038 (2020) <https://doi.org/10.1016/j.patcog.2019.107038>
- Liu, H., Chau, L.-P.: Underwater image restoration based on contrast enhancement. In: 2016 IEEE International Conference on Digital Signal Processing (DSP), pp. 584–588 (2016). <https://doi.org/10.1109/ICDSP.2016.7868625>
- Ling, Z., Delnevo, G., Salomoni, P., Mirri, S.: Findings on machine learning for identification of archaeological ceramics: A systematic literature review. *IEEE Access* **12**, 100167–100185 (2024) <https://doi.org/10.1109/ACCESS.2024.3429623>
- Liang, Z., Ding, X., Wang, Y., Yan, X., Fu, X.: GUDCP: Generalization of underwater dark channel prior for underwater image restoration. *IEEE Transactions on Circuits and Systems for Video Technology* **32**(7), 4879–4884 (2022) <https://doi.org/10.1109/TCSVT.2021.3114230>
- Lin, Y., Dai, Z., Zhu, S., Yao, Y.: Gaussian-flow: 4D reconstruction with dynamic 3d gaussian particle. In: the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 21136–21145 (2024)
- Lin, C.-H., al.: BARF: Bundle-Adjusting Neural Radiance Fields. In: ICCV (2021)

- Liu, R., Fan, X., Zhu, M., Hou, M., Luo, Z.: Real-World Underwater Enhancement: Challenges, Benchmarks, and Solutions Under Natural Light. *IEEE Transactions on Circuits and Systems for Video Technology* **30**(12), 4861–4875 (2020) <https://doi.org/10.1109/TCSVT.2019.2963772>
- Liu, X., Gao, Z., Chen, B.M.: Ipmgan: Integrating physical model and generative adversarial network for underwater image enhancement. *Neurocomputing* **453**, 538–551 (2021)
- Liu, S., Gao, N., Fu, S., Zhong, X., Li, H.: SeaFree-GS: Reconstructing underwater 3d scenes with true appearances. *IEEE Signal Processing Letters* **32**, 2114–2118 (2025) <https://doi.org/10.1109/LSP.2025.3567853>
- Liu, S., Gao, N., Gu, Z., Dou, H., Deng, Y., Li, H.: Spatiotemporal degradation-aware 3d gaussian splatting for realistic underwater scene reconstruction. In: *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 141–150 (2025). <https://doi.org/10.1145/3746027.3754888>
- Li, C., Guo, C., Ren, W., Cong, R., Hou, J., Kwong, S., Tao, D.: An underwater image enhancement benchmark dataset and beyond. *IEEE Transactions on Image Processing* **29**, 4376–4389 (2019)
- Li, C., Guo, C., Ren, W., Cong, R., Hou, J., Kwong, S., Tao, D.: An underwater image enhancement benchmark dataset and beyond. *IEEE Transactions on Image Processing* **29**, 4376–4389 (2020)
- Lu, S., Guan, F., Zhang, H., Lai, H.: Underwater image enhancement method based on denoising diffusion probabilistic model. *Journal of Visual Communication and Image Representation* **96**, 103926 (2023) <https://doi.org/10.1016/j.jvcir.2023.103926>
- Lu, S., Guan, F., Zhang, H., Lai, H.: Speed-up ddpm for real-time underwater image enhancement. *IEEE Transactions on Circuits and Systems for Video Technology* **34**(5), 3576–3588 (2024) <https://doi.org/10.1109/TCSVT.2023.3314767>
- Liu, R., Jiang, Z., Yang, S., Fan, X.: Twin adversarial contrastive learning for underwater image enhancement and beyond. *IEEE Transactions on Image Processing* **31**, 4922–4936 (2022)
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: *IEEE International Conference on Computer Vision*, pp. 10012–10022 (2021)
- Lin, W.-T., Lin, Y.-X., Chen, J.-W., Hua, K.-L.: Pixmamba: Leveraging state space models in a dual-level architecture for underwater image enhancement. In: *Proceedings of the Asian Conference on Computer Vision (ACCV)*, pp. 3622–3637 (2024)
- Liu, S., Lu, J., Gu, Z., Li, J., Deng, Y.: Aquatic-gs: A hybrid 3d representation for underwater scenes. *arXiv preprint arXiv:2411.00239* (2024)
- Lee, B., Lee, H., Sun, X., Ali, U., Park, E.: Deblurring 3d gaussian splatting. In: *European Conference on Computer Vision*, pp. 127–143 (2024). Springer
- Levin, A., Lischinski, D., Weiss, Y.: A Closed-Form Solution to Natural Image Matting. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **30**(2), 228–242 (2008) <https://doi.org/10.1109/TPAMI.2007.1177>
- Li, H., Li, J., Wang, W.: A fusion adversarial underwater image enhancement network with a public test dataset. *arXiv preprint arXiv:1906.06819* (2019)
- Lu, H., Li, Y., Zhang, L., Serikawa, S.: Contrast enhancement for images in turbid water. *JOSA A* **32**(5), 886–893 (2015)
- Levy, D., Peleg, A., Pearl, N., Rosenbaum, D., Akkaynak, D., Korman, S., Treibitz, T.: Seathru-nerf: Neural radiance fields in scattering media. In: *the IEEE/CVF Conference on Computer Vision and Pattern Recognition*

- (CVPR), pp. 56–65 (2023)
- Li, C., Quo, J., Pang, Y., Chen, S., Wang, J.: Single underwater image restoration by blue-green channels dehazing and red channel correction. In: IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1731–1735 (2016)
- Li, J., Skinner, K.A., Eustice, R.M., Johnson-Roberson, M.: WaterGAN: Unsupervised Generative Network to Enable Real-Time Color Correction of Monocular Underwater Images. IEEE Robotics and Automation Letters **3**(1), 387–394 (2018) <https://doi.org/10.1109/LRA.2017.2730363>
- Li, H., Song, W., Xu, T., Elsig, A., Kulhanek, J.: WaterSplatting: Fast underwater 3d scene reconstruction using gaussian splatting. International Conference on 3D Vision (2025)
- Li, Z., Tang, C., Huang, R., Han, C.: TAFormer: A transmission-aware transformer for underwater image enhancement. IEEE Transactions on Circuits and Systems for Video Technology **35**(1), 770–784 (2025) <https://doi.org/10.1109/TCSVT.2024.3417165>
- Liu, P., Wang, G., Qi, H., Zhang, C., Zheng, H., Yu, Z.: Underwater image enhancement with a deep residual framework. IEEE Access **7**, 94614–94629 (2019)
- Lu, H., Zhang, Y., Li, Y., Zhou, Q., Tadoh, R., Uemura, T., Kim, H., Serikawa, S.: Depth Map Reconstruction for Underwater Kinect Camera Using Inpainting and Local Image Mode Filtering. IEEE Access **5**, 7115–7122 (2017) <https://doi.org/10.1109/ACCESS.2017.2690455>
- Marques, T.P., Albu, A.B.: L2uwe: A framework for the efficient enhancement of low-light underwater images using local contrast and multi-scale fusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 538–539 (2020)
- Mohd Azmi, K.Z., Abdul Ghani, A.S., Md Yusof, Z., Ibrahim, Z.: Natural-based underwater image color enhancement through fusion of swarm-intelligence algorithm. Applied Soft Computing **85**, 105810 (2019) <https://doi.org/10.1016/j.asoc.2019.105810>
- Mualem, N., Amoyal, R., Freifeld, O., Akkaynak, D.: Gaussian splashing: Direct volumetric rendering underwater. arXiv preprint arXiv:2411.19588 (2024)
- Martin-Brualla, R., Radwan, N., Sajjadi, M.S.M., Barron, J.T., Dosovitskiy, A., Duckworth, D.: NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections. In: CVPR (2021)
- McGlamery, B.: A computer model for underwater camera systems. In: Ocean Optics VI, vol. 208, pp. 221–231 (1980). SPIE
- Müller, T., Evans, A., Schied, C., Keller, A.: Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. ACM Trans. Graph. (SIGGRAPH) (2022)
- Müller, T., Evans, A., Schied, C., Keller, A.: Instant neural graphics primitives with a multiresolution hash encoding. ACM Transactions on Graphics (Proc. SIGGRAPH) **41**(4), 102–110215 (2022)
- Malyugina, A., Huang, G., Ruiz-Libreros, E., Leslie, B., Anantrasirichai, N.: Marine snow removal using internally generated pseudo ground truth. In: 33rd European Signal Processing Conference (2025)
- Mertens, T., Kautz, J., Van Reeth, F.: Exposure Fusion: A Simple and Practical Alternative to High Dynamic Range Photography. Computer Graphics Forum **28**(1), 161–171 (2009) <https://doi.org/10.1111/j.1467-8659.2008.01171.x>
- Menna, F., Nocerino, E., Remondino, F.: Flat versus hemispherical dome ports in underwater photogrammetry. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol.

- XLII-2/W3, pp. 481–487 (2017). <https://doi.org/10.5194/isprs-archives-XLII-2-W3-481-2017>
- Mohan, S., Simon, P.: Underwater Image Enhancement based on Histogram Manipulation and Multiscale Fusion. *Procedia Computer Science* **171**, 941–950 (2020) <https://doi.org/10.1016/j.procs.2020.04.102>
- Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. In: *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 405–421 (2020)
- Mu, P., Xu, H., Liu, Z., Wang, Z., Chan, S., Bai, C.: A generalized physical-knowledge-guided dynamic model for underwater image enhancement. In: *ACM International Conference on Multimedia*, pp. 7111–7120 (2023)
- Nissar, M., Farhadifard, F., Radolko, M., Lukas, U.: A human vision neuroscience-driven deep neural framework for change detection in underwater scenes. In: *Computer Vision, Pattern Recognition, Image Processing, and Graphics. Communications in Computer and Information Science*, vol. 2522, pp. 427–438. Springer, ??? (2026)
- Nocerino, E., Menna, F., Gruen, A., Troyer, M., Capra, A., Castagnetti, C., Rossi, P., Brooks, A.J., Schmitt, R.J., Holbrook, S.J.: Coral reef monitoring by scuba divers using underwater photogrammetry and geodetic surveying. *Remote Sensing* (2020)
- Narasimhan, S.G., Nayar, S.K.: Chromatic framework for vision in bad weather. In: *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00562)*, vol. 1, pp. 598–6051 (2000). <https://doi.org/10.1109/CVPR.2000.855874>
- Narasimhan, S.G., Nayar, S.K.: Vision and the atmosphere. *International journal of computer vision* **48**, 233–254 (2002)
- Narasimhan, S.G., Nayar, S.K.: Interactive (de) weathering of an image using physical models. In: *IEEE Workshop on Color and Photometric Methods in Computer Vision*, vol. 6, p. 1 (2003). France
- Oquab, M., Darcet, T., Moutakanni, T., Vo, H.V., Szafraniec, M., Khalidov, V., Fernandez, P., HAZIZA, D., Massa, F., El-Nouby, A., Assran, M., Ballas, N., Galuba, W., Howes, R., Huang, P.-Y., Li, S.-W., Misra, I., Rabbat, M., Sharma, V., Synnaeve, G., Xu, H., Jegou, H., Mairal, J., Labatut, P., Joulin, A., Bojanowski, P.: DINOv2: Learning robust visual features without supervision. *Transactions on Machine Learning Research* (2024)
- Pizer, S.M., Amburn, E.P., Austin, J.D., Cromartie, R., Geselowitz, A., Greer, T., Zuiderveld, A.: *Adaptive Histogram Equalization and Its Variations* (1987)
- Peng, Y.-T., Cosman, P.C.: Underwater Image Restoration Based on Image Blurriness and Light Absorption. *IEEE Transactions on Image Processing* **26**(4), 1579–1594 (2017) <https://doi.org/10.1109/TIP.2017.2663846>
- Peng, Y.-T., Cao, K., Cosman, P.C.: Generalization of the dark channel prior for single image restoration. *IEEE Transactions on Image Processing* **27**(6), 2856–2868 (2018)
- Peng, Y.-T., Chen, Y.-R., Chen, G.-R., Liao, C.-J.: Histoformer: Histogram-based transformer for efficient underwater image enhancement. *IEEE Journal of Oceanic Engineering* **50**(1), 164–177 (2025) <https://doi.org/10.1109/JOE.2024.3474919>
- Pumarola, A., Corona, E., Pons-Moll, G., Moreno-Noguer, F.: D-NeRF: Neural Radiance Fields for Dynamic Scenes. In: *CVPR* (2021)
- Panetta, K., Gao, C., Agaian, S.: Human-visual-system-inspired underwater image quality measures. *IEEE Journal of Oceanic Engineering* **41**(3), 541–551 (2015)
- Pizer, S.M., Johnston, R.E., Ericksen, J.P., Yankaskas, B.C., Muller, K.E.: Contrast-limited adaptive histogram equalization: Speed and effectiveness. In: [1990] *Proceedings of the First Conference on Visualization in*

- Biomedical Computing, pp. 337–345 (1990). <https://doi.org/10.1109/VBC.1990.109340>
- Ponzi, V., Napoli, C.: Graph neural networks: Architectures, applications, and future directions. *IEEE Access* **13**, 62870–62891 (2025) <https://doi.org/10.1109/ACCESS.2025.3558752>
- Prado, E., Rodríguez-Basalo, A., Cobo, A., Ríos, P., Sánchez, F.: 3d fine-scale terrain variables from underwater photogrammetry. *Remote Sensing* (2020)
- Park, K., Sinha, U., Barron, J.T., Bouaziz, S., Goldman, D.B., Seitz, S.M., Martin-Brualla, R.: Nerfies: Deformable neural radiance fields. In: *International Conference on Computer Vision (ICCV)*, pp. 5865–5874 (2021)
- Pramanick, A., Sarma, S., Sur, A.: X-caunet: Cross-color channel attention with underwater image-enhancing transformer. In: *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3550–3554 (2024). IEEE
- Piérard, G., Van Droogenbroeck, M.: A methodology to evaluate strategies predicting rankings on unseen domains. *CoRR* **abs/2505.15595** (2025)
- Piérard, G., Van Droogenbroeck, M.: What is the optimal ranking score between precision and recall? we can always find it and it is rarely F1. *CoRR* **abs/2511.22442** (2025)
- Peng, L., Zhu, C., Bian, L.: U-shape transformer for underwater image enhancement. *IEEE Transactions on Image Processing* (2023)
- Peng, Y.-T., Zhao, X., Cosman, P.C.: Single underwater image enhancement using depth estimation based on blurriness. In: *2015 IEEE International Conference on Image Processing (ICIP)*, pp. 4952–4956 (2015). <https://doi.org/10.1109/ICIP.2015.7351749>
- Qiao, Y., Shao, M., Meng, L., Xu, K.: RestorGS: Depth-aware gaussian splatting for efficient 3d scene restoration. In: *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11177–11186 (2025). <https://doi.org/10.1109/CVPR52734.2025.01044>
- Qu, Z., Vengurlekar, O., Qadri, M., Zhang, K., Kaess, M., Metzler, C., Jayasuriya, S., Pediredla, A.: Z-Splat: Z-Axis Gaussian Splatting for Camera-Sonar Fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2024)
- Ramazzina, A., Bijelic, M., Walz, S., Sanvito, A., Scheuble, D., Heide, F.: Scatternerf: Seeing through fog with physically-based inverse neural rendering. In: *International Conference on Computer Vision (ICCV)*, pp. 17957–17968 (2023)
- Radolko, M., Farhadifard, F., Lukas, U.: Dataset on underwater change detection. In: *OCEANS 2016 MTS/IEEE Monterey* (2016). <https://doi.org/10.1109/OCEANS.2016.7761129>
- Rahnama, R., Hill, D., Mills, A.: Subsea inspection supported by ai-driven computer vision. In: *Proceedings of ADIPEC* (2025). <https://doi.org/10.2118/229719-MS>
- Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., *et al.*: Learning transferable visual models from natural language supervision. In: *International Conference on Machine Learning*, pp. 8748–8763 (2021). PMLR
- Rao, Y., Liu, W., Li, K., Fan, H., Wang, S., Dong, J.: Deep color compensation for generalized underwater image enhancement. *IEEE Transactions on Circuits and Systems for Video Technology* **34**(4), 2577–2590 (2024) <https://doi.org/10.1109/TCSVT.2023.3305777>
- Rahman, S., Li, A.Q., Rekleitis, I.: SVIn2: An underwater slam system using sonar, visual, inertial, and depth sensor. In: *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1861–1868

- (2019). <https://doi.org/10.1109/IROS40897.2019.8967703>
- Ruan, J., Li, J., Xiang, S.: VM-UNet: Vision mamba unet for medical image segmentation. arXiv preprint arXiv:2402.02491 (2024)
- Rout, D., Ray, N., Das, M.: Underwater visual surveillance: A comprehensive survey. *Ocean Engineering* **314**, 119556 (2024) <https://doi.org/10.1016/j.oceaneng.2024.119556>
- Ren, T., Xu, H., Jiang, G., Yu, M., Zhang, X., Wang, B., Luo, T.: Reinforced swin-convs transformer for simultaneous underwater sensing scene image enhancement and super-resolution. *IEEE Transactions on Geoscience and Remote Sensing* **60**, 1–16 (2022) <https://doi.org/10.1109/TGRS.2022.3144882>
- Sadasivan, A., *et al.*: A systematic survey of graph convolutional networks for artificial intelligence applications. *WIREs Data Mining and Knowledge Discovery* **15**(2), 70012 (2025) <https://doi.org/10.1002/widm.70012>
- Shaker, E., Baker, M.R., Mahmood, Z.: The impact of image enhancement and transfer learning techniques on marine habitat mapping. *Gazi University Journal of Science* **36**(2), 592–606 (2023) <https://doi.org/10.35378/gujs.973082>
- Seitz, S.M., Curless, B., Diebel, J., Scharstein, D., Szeliski, R.: A comparison and evaluation of multi-view stereo reconstruction algorithms. In: *Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR)*, pp. 519–528 (2006)
- Shin, Y.-S., Cho, Y., Pandey, G., Kim, A.: Estimation of ambient light and transmission map with common convolutional architecture. In: *OCEANS 2016 MTS/IEEE Monterey*, pp. 1–7 (2016). <https://doi.org/10.1109/OCEANS.2016.7761342>
- Storlazzi, C., Dartnell, P., Hatcher, G.A., Gibbs, A.E.: End of the chain? rugosity and fine-scale bathymetry from existing underwater digital imagery using structure-from-motion (sfm) technology. *Coral Reefs* (2016)
- Schonberger, J.L., Frahm, J.-M.: Structure-from-motion revisited. In: *the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4104–4113 (2016)
- Sethi, R., Indu, S.: Fusion of Underwater Image Enhancement and Restoration. *International Journal of Pattern Recognition and Artificial Intelligence* (2019) <https://doi.org/10.1142/S0218001420540075>
- Schechner, Y.Y., Karpel, N.: Clear underwater vision. In: *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, vol. 1, p. (2004). IEEE
- Sedlazeck, A., Koser, K., Koch, R.: 3d reconstruction based on underwater video from roV *Kiel 6000* considering underwater imaging conditions. In: *Proceedings of OCEANS 2009-Europe*, pp. 1–10 (2009)
- She, M., Nakath, D., Song, Y., Köser, K.: Refractive geometry for underwater domes. *ISPRS Journal of Photogrammetry and Remote Sensing* **183**, 525–540 (2022) <https://doi.org/10.1016/j.isprsjprs.2021.11.006>
- Spampinato, C., Palazzo, S., Giordano, D., Kavasidis, I., Lin, F.-P., Lin, Y.-T.: Understanding fish behavior during typhoon events in real-life underwater environments. *Multimedia Tools and Applications* **70**(1), 199–224 (2014) <https://doi.org/10.1007/s11042-012-1228-0>
- Skinner, K.A., Ruland, E.I., Johnson-Roberson, M.: Automatic color correction for 3d reconstruction of underwater scenes. In: *IEEE International Conference on Robotics and Automation* (2017)
- Sethuraman, A.V., Ramanagopal, M.S., Skinner, K.A.: WaterNeRF: Neural radiance fields for underwater scenes. In: *OCEANS 2023 - MTS/IEEE U.S. Gulf Coast*, pp. 1–7 (2023). <https://doi.org/10.23919/OCEANS52994.2023.10336972>
- Snively, N., Seitz, S.M., Szeliski, R.: Photo tourism: Exploring photo collections in 3d. In: *ACM SIGGRAPH*

2006 Papers, pp. 835–846 (2006)

Samboko, H.T., Schurer, S., Savenije, H.H.G., Makurira, H., Banda, K., Winsemius, H.: Evaluating low-cost topographic surveys for computations of conveyance. *Geoscientific Instrumentation, Methods and Data Systems* **11**(1), 1–23 (2022) <https://doi.org/10.5194/gi-11-1-2022>

Song, W., Wang, Y., Huang, D., Liotta, A., Perra, C.: Enhancement of Underwater Images With Statistical Model of Background Light and Optimization of Transmission Map. *IEEE Transactions on Broadcasting* **66**(1), 153–169 (2020) <https://doi.org/10.1109/TBC.2019.2960942>

Shen, Z., Xu, H., Luo, T., Song, Y., He, Z.: UDAformer: Underwater image enhancement based on dual attention transformer. *Computers & Graphics* **111**, 77–88 (2023) <https://doi.org/10.1016/j.cag.2023.01.009>

Tan, R.T.: Visibility in bad weather from a single image. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2008). <https://doi.org/10.1109/CVPR.2008.4587643>

Tarel, J.-P., Hautiere, N.: Fast visibility restoration from a single color or gray level image. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 2201–2208. IEEE, Kyoto (2009). <https://doi.org/10.1109/ICCV.2009.5459251>

Tang, Y., Iwaguchi, T., Kawasaki, H., Sagawa, R., Furukawa, R.: Autoenhancer: Transformer on u-net architecture search for underwater image enhancement. In: Proceedings of the Asian Conference on Computer Vision, pp. 1403–1420 (2022)

Tang, Y., Kawasaki, H., Iwaguchi, T.: Underwater image enhancement by transformer-based diffusion model with non-uniform sampling for skip strategy. In: ACM International Conference on Multimedia, pp. 5419–5427 (2023)

Teague, J., Scott, T.: Underwater photogrammetry and 3d reconstruction of submerged objects in shallow environments by roV and underwater gps. *Journal of Marine Science Research and Technology* (2017)

Tang, Y., Zhu, C., Wan, R., Xu, C., Shi, B.: Neural underwater scene representation. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11780–11789 (2024). <https://doi.org/10.1109/CVPR52733.2024.01119>

Uplavikar, P.M., Wu, Z., Wang, Z.: All-in-one underwater image enhancement using domain-adversarial learning. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 1–8 (2019)

Vasamsetti, S., Mittal, N., Neelapu, B.C., Sardana, H.K.: Wavelet based perspective on variational enhancement technique for underwater imagery. *Ocean Engineering* **141**, 88–100 (2017) <https://doi.org/10.1016/j.oceaneng.2017.06.012>

Vrochidis, A., Tzovaras, D., Krinidis, S.: Enhancing 3D reconstructions in underwater environments: The impact of image enhancement on model quality. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLVIII-2/W10-2025, pp. 317–324 (2025). <https://doi.org/10.5194/isprs-archives-XLVIII-2-W10-2025-317-2025>

Vrochidis, A., Tzovaras, D., Krinidis, S.: Enhancing three-dimensional reconstruction through intelligent colormap selection. *Sensors* **25**(8), 2576 (2025) <https://doi.org/10.3390/s25082576>

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

Wang, Z., Bovik, A.C.: A universal image quality index. *IEEE signal processing letters* **9**(3), 81–84 (2002)

- Wang, Z., Bovik, A.C.: Mean squared error: Love it or leave it? a new look at signal fidelity measures. *IEEE signal processing magazine* **26**(1), 98–117 (2009)
- Wang, Z., Cun, X., Bao, J., Zhou, W., Liu, J., Li, H.: Uformer: A general u-shaped transformer for image restoration. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 17662–17672 (2022)
- Wright, A.E., Conlin, D.L., Shope, S.M.: Assessing the accuracy of underwater photogrammetry for archaeology: A comparison of structure from motion photogrammetry and real time kinematic survey at the east key construction wreck. *Journal of Marine Science and Engineering* **8**(11), 849 (2020) <https://doi.org/10.3390/jmse8110849>
- Wittmann, J., Chatterjee, S., Sure, T.: Robust marker detection and identification using deep learning in underwater images for close range photogrammetry. *ISPRS Open Journal of Photogrammetry and Remote Sensing* **13**, 100072 (2024) <https://doi.org/10.1016/j.ophoto.2024.100072>
- Wen, J., Cui, J., Yang, G., Zhao, B., Zhai, Y., Gao, Z., Dou, L., Chen, B.M.: Waterformer: A global–local transformer for underwater image enhancement with environment adaptor. *IEEE Robotics & Automation Magazine* **31**(1), 29–40 (2024) <https://doi.org/10.1109/MRA.2024.3351487>
- Wen, J., Cui, J., Zhao, Z., Yan, R., Gao, Z., Dou, L., Chen, B.M.: Syreanet: A physically guided underwater image enhancement framework integrating synthetic and real images. In: IEEE International Conference on Robotics and Automation (ICRA), pp. 5177–5183 (2023). <https://doi.org/10.1109/ICRA48891.2023.10161260>
- Wang, Y., Guo, J., Gao, H., Yue, H.: UIEC²-Net: CNN-based underwater image enhancement using two color space. *Signal Processing: Image Communication* **96**, 116250 (2021) <https://doi.org/10.1016/j.image.2021.116250>
- Wang, H., Köser, K., Ren, P.: Large foundation model empowered discriminative underwater image enhancement. *IEEE Transactions on Geoscience and Remote Sensing* **63**, 1–17 (2025) <https://doi.org/10.1109/TGRS.2025.3525962>
- Wang, Y., Liu, H., Chau, L.-P.: Single Underwater Image Restoration Using Adaptive Attenuation-Curve Prior. *IEEE Transactions on Circuits and Systems I: Regular Papers* **65**(3), 992–1002 (2018) <https://doi.org/10.1109/TCSI.2017.2751671>
- Wang, S., Ma, K., Yeganeh, H., Wang, Z., Lin, W.: A patch-structure representation method for quality assessment of contrast changed images. *IEEE Signal Processing Letters* **22**(12), 2387–2390 (2015)
- Woodham, R.J.: Photometric method for determining surface orientation from multiple images. *Optical Engineering* **19**(1), 139–144 (1980)
- Wang, Z., Shen, L., Yu, Y., Hui, Y.: UIERL: Internal-External Representation Learning Network for Underwater Image Enhancement. *IEEE Transactions on Multimedia* **26**, 9252–9267 (2024) <https://doi.org/10.1109/TMM.2024.3387760>
- Wen, H., Tian, Y., Huang, T., Gao, W.: Single underwater image enhancement with a new optical model. In: 2013 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 753–756 (2013). <https://doi.org/10.1109/ISCAS.2013.6571956>
- Wang, B., Xu, H., Jiang, G., Yu, M., Ren, T., Luo, T., Zhu, Z.: Uie-convformer: Underwater image enhancement based on convolution and feature fusion transformer. *IEEE Transactions on Emerging Topics in Computational Intelligence* **8**(2), 1952–1968 (2024)
- Wu, G., Yi, T., Fang, J., Xie, L., Zhang, X., Wei, W., Liu, W., Tian, Q., Wang, X.: 4D gaussian splatting for real-time dynamic scene rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 20310–20320 (2024)
- Wen, J., Yang, G., Zhao, B., Huang, D., Lei, L., Zhang, B., Gao, Z., Chen, X., Chen, B.M.: A semi-supervised

- domain-adaptive framework for real-world underwater image enhancement. *IEEE Transactions on Geoscience and Remote Sensing* **63**, 1–15 (2025) <https://doi.org/10.1109/TGRS.2025.3590798>
- Wei, Y., Zhang, Y., Li, K., Wang, F., Tang, S., Zhang, Z.: Leveraging vision-language prompts for real-world image restoration and enhancement. *Computer Vision and Image Understanding* **250**, 104222 (2025) <https://doi.org/10.1016/j.cviu.2024.104222>
- Xia, H., Bao, B., Liao, F., Chen, J., Wang, B., Li, Z.: A patch-based method for underwater image enhancement with denoising diffusion models. *IEEE Transactions on Cybernetics* **55**(1), 269–281 (2025) <https://doi.org/10.1109/TCYB.2024.3482174>
- Xian, W., Bao, J., Zhang, T., Chen, D., Wen, F., Guo, B.: Space-time Neural Irradiance Fields for Free-Viewpoint Video. In: *CVPR* (2021)
- 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 (CVPR)*, pp. 22358–22367 (2024). <https://doi.org/10.1109/CVPR52733.2024.02110>
- Xiao, F., Yuan, F., Huang, Y., Cheng, E.: Turbid underwater image enhancement based on parameter-tuned stochastic resonance. *IEEE Journal of Oceanic Engineering* **48**(1), 127–146 (2023) <https://doi.org/10.1109/JOE.2022.3190517>
- Yi, J., Bi, Q., Zheng, H., Huang, H., Zhan, H., Shen, Y., Ji, W., Huang, Y., Li, Y., Wu, X., Zheng, Y.: AtlantisGS: Underwater sparse-view scene reconstruction via gaussian splatting. In: *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 7805–7814 (2025). <https://doi.org/10.1145/3746027.3755125>
- Yang, H.-Y., Chen, P.-Y., Huang, C.-C., Zhuang, Y.-Z., Shiau, Y.-H.: Low Complexity Underwater Image Enhancement Based on Dark Channel Prior. In: *2011 Second International Conference on Innovations in Bio-inspired Computing And Applications*, pp. 17–20 (2011). <https://doi.org/10.1109/IBICA.2011.9>
- Yu, Z., Chen, A., Huang, B., Sattler, T., Geiger, A.: Mip-splatting: Alias-free 3d gaussian splatting. In: *the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 19447–19456 (2024)
- Ye, T., Chen, S., Liu, Y., Ye, Y., Chen, E., Li, Y.: Underwater light field retention: Neural rendering for underwater imaging. In: *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 488–497 (2022)
- Yan, S., Chen, X., Wu, Z., Tan, M., Yu, J.: Hybrur: A hybrid physical-neural solution for unsupervised underwater image restoration. *IEEE Transactions on Image Processing* (2023)
- Yang, Z., Gao, X., Zhou, W., Jiao, S., Zhang, Y., Jin, X.: Deformable 3d gaussians for high-fidelity monocular dynamic scene reconstruction. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2024)
- Yaqoob, M., Ishaq, M., Ansari, M.Y., *et al.*: Advancing paleontology: A survey on deep learning methodologies in fossil image analysis. *Artificial Intelligence Review* **58**, 83 (2025) <https://doi.org/10.1007/s10462-024-11080-y>
- Yang, L., Kang, B., Huang, Z., Xu, X., Feng, J., Zhao, H.: Depth anything: Unleashing the power of large-scale unlabeled data. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2024)
- Yan, Z., Low, W.F., Chen, Y., Lee, G.H.: Multi-scale 3d gaussian splatting for anti-aliased rendering. In: *the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20923–20931 (2024)
- Yang, D., Leonard, J.J., Girdhar, Y.: SeaSplat: Representing Underwater Scenes with 3D Gaussian Splatting and a Physically Grounded Image Formation Model. In: *International Conference on Robotics and Automation* (2025)

- Yao, Y., Luo, Z., Li, S., Fang, T., Quan, L.: Mvsnet: Depth inference for unstructured multi-view stereo. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 767–783 (2018)
- Yang, W., Lin, Y., Lim, C.H., Tao, Z., Leng, J.: Experimental comparison between nerfs and 3d gaussian splatting for underwater 3d reconstruction. In: 2024 China Automation Congress (CAC), pp. 6633–6638 (2024). <https://doi.org/10.1109/CAC63892.2024.10864941>
- Yuan, J., Li, Y., Zhang, Y., Guo, C., Tang, X., Wang, R., Li, C.: 3D-UIR: 3D gaussian for underwater 3d scene reconstruction via physics-based appearance-medium decoupling. arXiv preprint arXiv:2505.21238 (2025)
- Yang, M., Sowmya, A.: An underwater color image quality evaluation metric. IEEE Transactions on Image Processing **24**(12), 6062–6071 (2015)
- Yin, J., Wang, Y., Guan, B., Zeng, X., Guo, L.: Unsupervised underwater image enhancement based on disentangled representations via double-order contrastive loss. IEEE Transactions on Geoscience and Remote Sensing (2024)
- Yang, G., Wen, J., Zhao, B., Li, Q., Huang, Y., Lei, L., Chen, X., Lam, A.H.F., Chen, B.M.: End-to-end underwater multi-view stereo for dense scene reconstruction. In: IEEE International Conference on Robotics and Automation (ICRA), pp. 7616–7623 (2025). <https://doi.org/10.1109/ICRA55743.2025.11128539>
- Yang, X., Xie, W., Peng, S., Fu, Y., Fan, W., Yang, B., Dong, X.: 4D gaussian splatting for high-fidelity dynamic reconstruction of single-view scenes. Neurocomputing **640**, 130262 (2025). <https://doi.org/10.1016/j.neucom.2025.130262>
- Yan, H., Zhang, Z., Xu, J., Wang, T., An, P., Wang, A., Duan, Y.: Uw-cyclegan: Model-driven cyclegan for underwater image restoration. IEEE Transactions on Geoscience and Remote Sensing (2023)
- Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.-H.: Restormer: Efficient transformer for high-resolution image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5728–5739 (2022)
- Zhao, C., Cai, W., Dong, C., Hu, C.: Wavelet-based fourier information interaction with frequency diffusion adjustment for underwater image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8281–8291 (2024)
- Zhou, S., Chang, H., Jiang, S., Fan, Z., Zhu, Z., Xu, D., Chari, P., You, S., Wang, Z., Kadambi, A.: Feature 3dgs: Supercharging 3d gaussian splatting to enable distilled feature fields. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 21676–21685 (2024)
- Zheng, Z., Chen, Y., Zeng, H., Vu, T.-A., Hua, B.-S., Yeung, S.-Y.K.: MarineInst: A foundation model for marine image analysis with instance visual description. In: Proceedings of the European Conference on Computer Vision (ECCV) (2024)
- Zhang, S., Duan, Y., Li, D., Zhao, R.: Mamba-ue: Enhancing underwater images with physical model constraint. arXiv preprint arXiv:2407.19248 (2024)
- Zhang, W., Dong, L., Pan, X., Zhou, J., Qin, L., Xu, W.: Single Image Defogging Based on Multi-Channel Convolutional MSRCR. IEEE Access **7**, 72492–72504 (2019) <https://doi.org/10.1109/ACCESS.2019.2920403>
- Zhang, J., Han, F., Han, D., Yang, J., Zhao, W., Li, H.: Integration of sonar and visual-inertial systems for slam in underwater environments. IEEE Sensors Journal **24**(10), 16792–16804 (2024) <https://doi.org/10.1109/JSEN.2024.3384301>
- Zhang, Z., Hu, W., Lao, Y., He, T., Zhao, H.: Pixel-gs: Density control with pixel-aware gradient for 3d gaussian splatting. In: European Conference on Computer Vision, pp. 326–342 (2024). Springer

- Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586–595 (2018)
- Zhao, X., Jin, T., Qu, S.: Deriving inherent optical properties from background color and underwater image enhancement. *Ocean Engineering* **94**, 163–172 (2015) <https://doi.org/10.1016/j.oceaneng.2014.11.036>
- Zhang, W., Liu, Q., Lu, H., Wang, J., Liang, J.: Underwater image enhancement via wavelet decomposition fusion of advantage contrast. *IEEE Transactions on Circuits and Systems for Video Technology*, 1–1 (2025) <https://doi.org/10.1109/TCSVT.2025.3545595>
- Zhuang, P., Li, C., Wu, J.: Bayesian retinex underwater image enhancement. *Engineering Applications of Artificial Intelligence* **101**, 104171 (2021) <https://doi.org/10.1016/j.engappai.2021.104171>
- Zhu, P., Liu, Y., Wen, Y., Xu, M., Fu, X., Liu, S.: Unsupervised underwater image enhancement via content-style representation disentanglement. *Engineering Applications of Artificial Intelligence* **126**, 106866 (2023)
- Zhang, W., Li, X., Xu, S., Li, X., Yang, Y., Xu, D., Liu, T., Hu, H.: Underwater Image Restoration via Adaptive Color Correction and Contrast Enhancement Fusion. *Remote Sensing* **15**(19), 4699 (2023) <https://doi.org/10.3390/rs15194699>
- Zhou, J., Liang, T., Zhang, D., Liu, S., Wang, J., Wu, E.Q.: Waterhe-nerf: Water ray matching neural radiance fields for underwater scene reconstruction. *Information Fusion* **115**, 102770 (2025) <https://doi.org/10.1016/j.inffus.2024.102770>
- Zhang, A., Palaoag, T.: Underwater fish tracking algorithm based on ViBE detection and covariance matrix. In: *International Conference on Image, Video and Signal Processing (IVSP)* (2024)
- Zhang, Z., Peng, R., Hu, Y., Wang, R.: Geomysnet: Learning multi-view stereo with geometry perception. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21508–21518 (2023)
- Zhou, J., Pang, L., Zhang, D., Zhang, W.: Underwater image enhancement method via multi-interval subhistogram perspective equalization. *IEEE Journal of Oceanic Engineering* **48**(2), 474–488 (2023) <https://doi.org/10.1109/JOE.2022.3223733>
- Zhou, J., Sun, J., Li, C., Jiang, Q., Zhou, M., Lam, K.-M., Zhang, W., Fu, X.: Hclr-net: Hybrid contrastive learning regularization with locally randomized perturbation for underwater image enhancement. *International Journal of Computer Vision*, 1–25 (2024)
- Zhou, Y., Xu, H., Jiang, G., Ren, T., Yu, M., Chen, H.: UIE-SFIFormer: Underwater image enhancement based on physical-guided spatial-frequency interaction transformer. *IEEE Journal of Oceanic Engineering* **50**(2), 727–742 (2025) <https://doi.org/10.1109/JOE.2024.3458109>
- Zhang, Y., Yuan, J., Cai, Z.: Dcgf: Diffusion-color guided framework for underwater image enhancement. *IEEE Transactions on Geoscience and Remote Sensing* (2024)
- Zhang, S., Zhao, S., An, D., Liu, J., Wang, H., Feng, Y., Li, D., Zhao, R.: Visual slam for underwater vehicles: A survey. *Computer Science Review* (2022)
- Zhang, S., Zhao, S., An, D., Liu, J., Wang, H., Feng, Y., Li, D., Zhao, R.: Visual SLAM for underwater vehicles: A survey. *Computer Science Review* **46**, 100510 (2022) <https://doi.org/10.1016/j.cosrev.2022.100510>
- Zhuang, J., Zheng, Y., Guo, B., Yan, Y.: Globally deformable information selection transformer for underwater image enhancement. *IEEE Transactions on Circuits and Systems for Video Technology* (2024)

Zhang, T., Zhi, W., Johnson-Roberson, M.: Infinite leagues under the sea: Realistic 3d underwater terrain generation augmented by visual foundation models. In: ICLR 2025 Workshop on Foundation Models in the Wild (2025)

Zhang, T., Zhi, W., Meyers, B., Durrant, N., Huang, K., Mangelson, J., Barbalata, C., Johnson-Roberson, M.: RecGS: Removing Water Caustic with Recurrent Gaussian Splatting. IEEE Robotics and Automation Letters (2024)

Łacka, M., Łubczonek, J.: Methodology for creating a digital bathymetric model using neural networks for combined hydroacoustic and photogrammetric data in shallow water areas. Sensors **24**(1), 175 (2024) <https://doi.org/10.3390/s24010175>

ARTICLE IN PRESS