

中图法分类号: TP18; TP391 文献标识码: A 文章编号: 1006-8961(2026)05-1425-26

论文引用格式: Yuan J Y, Zhao Q Q, Li Y J, Zhang Y L, Guo C L and Li C Y. 2026. Overview and prospects of underwater novel view synthesis. Journal of Image and Graphics, 31(5):1425-1450(袁捷禹, 赵倩倩, 李纾君, 张元林, 郭春乐, 李重仪. 2026. 水下新视角合成研究综述与展望. 中国图象图形学报, 31(5):1425-1450)[DOI:10.11834/jig.250469]

水下新视角合成研究综述与展望

袁捷禹, 赵倩倩, 李纾君, 张元林, 郭春乐, 李重仪*

南开大学计算机学院, 天津 300350

摘要: 水下新视角合成旨在从有限观测视角中重建完整的三维场景, 并生成任意视点的高质量图像, 为海洋观测与认知提供更丰富的空间信息。随着智慧海洋建设的深入推进, 高精度水下三维重建与新视角合成的需求日益迫切。然而, 水下成像环境中普遍存在的强散射效应和波长选择性吸收, 对传统计算机视觉方法提出了挑战, 使得在陆地场景中表现优异的新视角合成技术难以直接应用于水下环境。鉴于国内外水下新视角合成任务的研究综述相对稀少, 本文对水下新视角合成方法进行系统总结和评述。本文系统梳理了水下新视角合成的研究脉络与最新进展: 从物理机理角度解析了水下成像退化的根本原因及其对新视角合成的影响, 总结并对比了典型研究方法, 对代表性工作的技术原理、性能表现及其在应对光学效应、动态干扰与复杂场景恢复方面的优势与不足进行了深入分析, 并从几何精度、成像质量和计算效率等方面对国内外研究现状进行了综合评估。最后, 讨论了当前水下新视角合成所面临的核心挑战, 并展望了未来的发展方向。本综述旨在为相关研究提供系统化的技术参考与发展指引, 推动该领域的深入研究与实际应用。

关键词: 水下成像; 水下立体观测; 三维视觉表征; 新视角合成; 水下视觉复原; 体积渲染; 物理建模

Overview and prospects of underwater novel view synthesis

Yuan Jieyu, Zhao Qianqian, Li Yujun, Zhang Yuanlin, Guo Chunle, Li Chongyi*

College of Computer Science, Nankai University, Tianjin 300350, China

Abstract: Underwater novel view synthesis (UNVS) is an emerging research area at the intersection of computer vision, computer graphics, and marine science. It aims to reconstruct complete 3D scenes from sparse or limited observations and generate photorealistic images from arbitrary viewpoints. Compared with traditional underwater imaging techniques, UNVS provides geometrically consistent scene reconstructions and enables flexible visualization of marine environments, thereby offering richer spatial information to support applications in marine observation, ecological monitoring, resource exploration, and digital ocean construction. With the rapid development of smart ocean initiatives globally, the demand for high-quality underwater 3D reconstruction and novel view synthesis has increased. Despite the remarkable success of novel view synthesis methods, such as neural radiance fields (NeRF) and 3D Gaussian splatting (3DGS), in terrestrial environments, their direct application underwater remains severely limited. The underwater imaging environment is fundamentally

收稿日期: 2025-09-26; 修回日期: 2025-12-26; 预印本日期: 2026-01-03

* 通信作者: 李重仪 lichongyi@nankai.edu.cn

基金项目: 国家自然科学基金项目(62501312); 天津市自然科学基金项目(24JCJQC00020, 25JCQNJC01390); 博士后基金天津联合资助(2025T022TJ); 中央高校基本科研业务费专项资金资助(南开大学, 070-63243143, 070-63253223, 070-63253219)

Supported by: National Natural Science Foundation of China (62501312); Tianjin Municipal Natural Science Foundation (24JCJQC00020, 25JCQNJC01390); China Postdoctoral Science Foundation-Tianjin Joint Support Program(2025T022TJ); Fundamental Research Funds for the Central Universities(Nankai University, 070-63243143, 070-63253223, 070-63253219)

different from natural scenes because of its complex optical propagation effects. Light attenuation is wavelength-dependent, causing severe color distortion and loss of contrast. Forward and backward scattering caused by water molecules and suspended particles further degrade image quality by introducing haze-like effects, blur, and noise. In addition, the refraction at the water-camera interface, the caustics caused by surface waves, and dynamic disturbances (e. g., moving marine organisms) pose additional challenges to stable reconstruction and consistent novel view generation. These unique physical and environmental factors make UNVS a highly challenging problem that requires specialized solutions. Despite the growing research interest and emerging progress in UNVS, comprehensive surveys remain scarce, particularly those that systematically analyze recent developments and provide a holistic perspective of the field. To address this gap, this study presents a comprehensive review of the current state of UNVS. First, we revisit the physical principles of underwater imaging and analyze how scattering, absorption, and multipath transmission affect visual data acquisition and subsequent reconstruction quality. Understanding these mechanisms is crucial because they form the theoretical basis for integrating physics-based priors into learning-driven approaches. Second, we provide an overview of representative methodological advances. Physics-inspired methods attempt to explicitly model underwater light propagation and compensate for degradation, and data-driven approaches, particularly those based on NeRF and its variants, leverage deep neural networks to learn radiance fields directly from observations. Recent efforts have further extended 3DGS, which offers order-of-magnitude improvements in rendering efficiency and is particularly attractive for real-time applications. Hybrid approaches that integrate physical modeling, deep learning, and multimodal sensing (e. g., combining optical and acoustic data) have also been discussed because they show promise in addressing the limitations of single-modality methods. By analyzing representative studies from domestic and international communities, we highlight their technical principles, performance metrics, and practical applications. Specific attention is paid to how different approaches address underwater optical effects, dynamic interference, and the restoration of complex scenes (e. g., coral reefs and shipwrecks). Comparative evaluations are provided across several dimensions, including geometric accuracy, visual fidelity, rendering efficiency, and robustness under varying turbidity conditions. These comparisons reveal that although remarkable progress has been attained, a trade-off often exists between reconstruction quality and computational cost, and current solutions are still far from achieving the generalization ability required for real-world deployment in diverse underwater environments. On the basis of this review, we identify several open challenges and potential future directions. One key challenge lies in disentangling medium properties from scene geometry, a task that requires improved joint modeling of water volume and object surfaces. Another important research direction is domain adaptation and generalization; models trained under one turbidity condition or geographic region often fail when they are applied elsewhere, highlighting the need for cross-domain robustness. Data scarcity also remains a bottleneck because collecting large-scale, high-quality underwater datasets is expensive and technically demanding. This situation suggests opportunities for leveraging unsupervised, weakly supervised, or simulation-to-real transfer learning strategies. Furthermore, real-time performance is essential for integration with underwater robotics and autonomous underwater vehicles (AUVs), and such integration calls for the development of efficient architectures and hardware-accelerated implementations. Furthermore, long-term perspectives point to the integration of multimodal sensing, the incorporation of physical priors into neural rendering frameworks, and the design of scalable systems that can operate in open, dynamic, and large-scale marine environments. In conclusion, UNVS represents a rapidly growing but still challenging area in vision and graphics research. By systematically reviewing the physical foundations, methodological advances, and representative applications, this study aims to provide a comprehensive technical roadmap for researchers and practitioners. The insights summarized herein are expected to guide future investigations; foster interdisciplinary collaboration across oceanography, robotics, and artificial intelligence; and ultimately accelerate the transition of UNVS from laboratory prototypes to practical tools that support marine science, ecological protection, and the sustainable development of ocean resources.

Key words: underwater imaging; underwater stereo observation; 3D visual representation; novel view synthesis; underwater visual restoration; volume rendering; physics-based modeling

0 引言

海洋作为地球表面的主体,蕴藏着丰富的资源和独特的生态系统,是人类可持续发展的重要战略空间。在“海洋强国”战略指引下,“数字海洋”建设和智慧海洋技术体系正加速推进,以期实现海洋信息的数字化采集、智能化处理和可视化呈现(李学龙, 2024)。随着深海探测装备技术的快速发展,水下可见光成像系统因其成本低廉、操作便捷等优势,已广泛应用于海洋科学研究、水下考古以及海洋工程等领域。然而,传统的二维图像采集方式仅提供有限的空间信息,难以满足现代海洋科学对三维空间认知的迫切需求。水下新视角合成技术作为计算机视觉与海洋科学的交叉前沿技术,能够从有限的观测视角重建完整的三维场景并生成任意视点的高质量图像,为海洋空间的全方位感知和智能化分析提供了革命性的技术手段(聂婕等, 2022)。围绕水下新视角合成在智慧海洋建设中的总体框架,图1展示了这一技术从数据采集到实际应用的完整链条。发展高精度的水下新视角合成技术,对于推进水下视觉技术的进步和拓展其应用边界具有重要的理论意义和实用价值(吴立新等, 2020)。

近年来,以神经辐射场(neural radiance field, NeRF)和三维高斯溅(3D Gaussian splatting, 3DGS)为代表的新视角合成技术在自然陆地场景中取得了突破性进展,实现了照片级别的高质量渲染效果。

然而,水下环境的特殊性给新视角合成技术带来了前所未有的挑战,现有方法在水下场景中的应用仍面临诸多技术瓶颈。水下光学环境与理想的陆地成像条件存在显著差异(丛润民等, 2020)。水体中高密度、非均匀介质会导致波长依赖的衰减与散射,随着成像距离增加,前向与后向散射效应加剧,物体与介质信息高度耦合,形成“水云”般的半透明结构,从而在渲染中引入明显伪影(郭继昌等, 2017)。此外,水下能见度通常不足陆地的10%,低对比度、强色偏与光照不均匀等退化特征不可避免地固化在输入图像中,并进一步传递到重建结果,导致颜色空间扭曲与细节丢失(周玲等, 2025)。更为关键的是,这些退化破坏了多视角几何依赖的光度一致性假设。这些复杂的物理现象使得传统新视角合成方法在水下环境中的直接应用效果不佳,需要针对水下成像特点进行专门的技术适应和优化。

水下新视角合成作为一个新兴的交叉研究领域,虽然在近年来受到学术界和产业界的广泛关注,但相关研究工作仍相对稀少,技术发展尚处于起步阶段。现有研究主要集中在将陆地场景的成功方法直接迁移到水下环境,缺乏对水下成像物理机制的深入理解和系统性的技术创新。同时,该领域缺乏全面的技术梳理和系统性综述,研究者难以全面了解技术现状、发展脉络和关键挑战,这在一定程度上阻碍了技术的快速发展和学科体系的建立。

鉴于水下新视角合成技术的重要应用价值和当前研究的分散性,有必要对该领域进行系统性的调

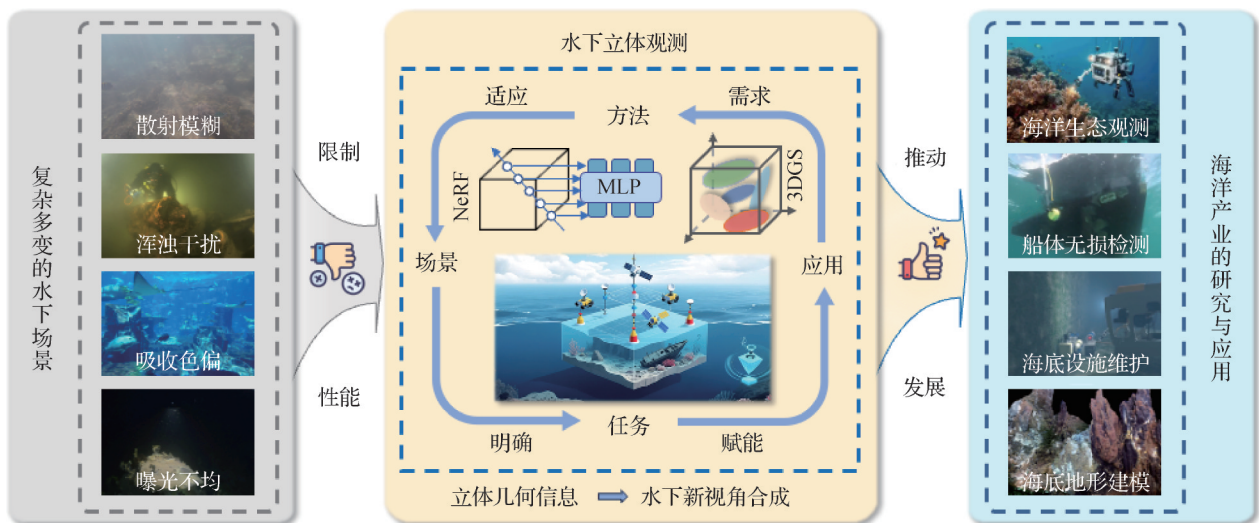


图1 水下新视角合成在智慧海洋中的研究与应用示意图

Fig. 1 Underwater novel view synthesis for research and applications in smart ocean

研和分析。通过对该领域的系统梳理,本文识别出以下关键技术演进趋势:1)从隐式到显式表征的范式转变,基于3DGS的显式方法因其实时渲染能力正逐步取代传统NeRF方法成为主流技术路线;2)物理建模与数据驱动深度融合,越来越多的方法将水下光学传输模型显式嵌入神经渲染框架,实现场景属性与介质效应的解耦;3)从逐场景优化到前馈泛化,三维几何大模型为水下场景的快速重建提供了新的技术可能。本文作为该领域的首篇中文综合性综述,在构建统一分析框架的同时,对该领域存在的关键问题和技术瓶颈进行深入讨论,并对未来发展趋势和研究方向进行前瞻性展望。通过这一系统性综述,旨在为相关研究者提供全面的技术图谱和发展指引,推动水下新视角合成技术的快速发展和广泛应用。

1 技术基础

1.1 水下成像退化分析

水下成像系统面临着复杂的光学挑战,这些挑战主要源于水介质与光线的相互作用。与清澈的大

气环境不同,水下光传播过程受到吸收和散射两种物理机制的显著影响,传播过程如图2所示。吸收效应表现为波长选择性衰减。红光在水中衰减最为严重,而蓝绿光相对保留较多,这一现象直接导致水下图像呈现典型的蓝绿色调偏移,且色彩失真程度随光程距离增加呈指数级加剧。散射效应由水分子和悬浮颗粒共同引起,包括前向散射和后向散射两个分量。前向散射使光线偏离原始传播路径,造成图像模糊和细节丢失;后向散射则将环境光引入成像光路,形成加性的背景雾,显著降低图像对比度。由于前向散射分量对图像退化的影响较小,为简化处理,上述光学效应可通过修正的辐射传输方程加以定量描述为

$$I = J \cdot e^{-\beta_a \cdot d} + A \cdot (1 - e^{-\beta_b \cdot d}) \quad (1)$$

式中, I 表示传感器接收到的辐射亮度, J 表示场景真实辐射亮度, β_a 与 β_b 为波长相关的衰减系数, d 为拍摄距离, A 为无穷远处的背景光亮度。该物理模型清晰地揭示了水下图像退化过程中距离依赖性和光谱选择性的本质特征,这些特征的综合作用导致水下图像普遍存在颜色失真、对比度下降以及细节模糊等典型问题。

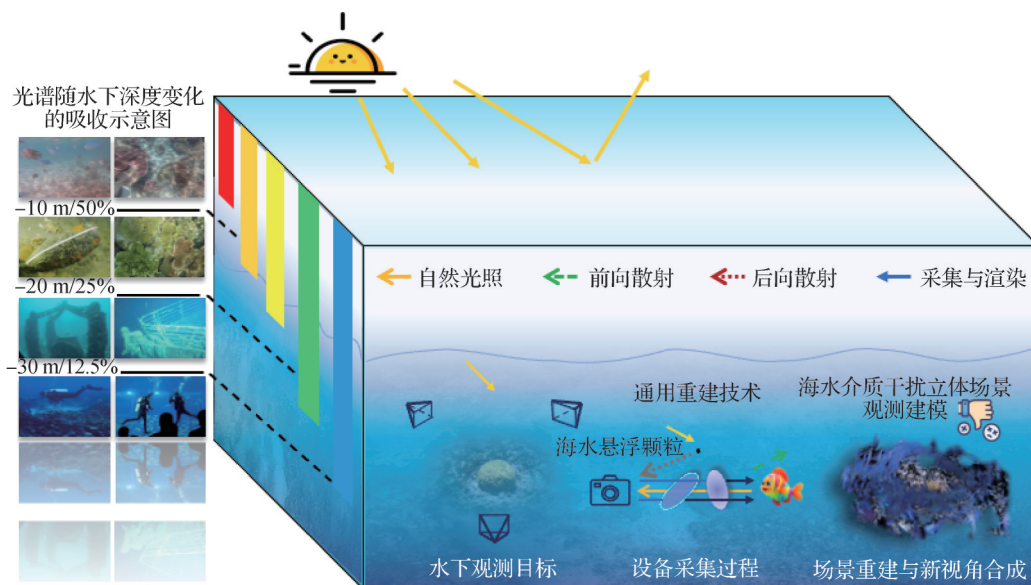


图2 水下光传播效应及其对三维场景建模与观测的影响示意图

Fig. 2 Schematic of underwater light propagation effects and their impact on 3D scene modeling and observation

水下环境的这些独特光学性质对新视角合成技术构成了根本性的技术挑战。现有的基于体渲染的方法,例如NeRF和3DGS等先进技术,通常基于光传输过程仅涉及发射和吸收的理想化假设,并采用

标准的体积渲染方程实现场景重建。然而,这一假设在水下环境中存在显著偏差,因为水介质作为参与介质通过连续的多次散射过程积极参与整个成像过程,引入了在清澈空气渲染中并不存在的复杂体

散射效应。

水下成像的退化特性对新视角合成技术造成两方面的核心干扰。首先是参与介质效应的建模困难。水介质具有显著的体积散射特性,光子在传播过程中经历多次散射事件,产生复杂的路径分布和角度依赖的散射相函数,而这种体积散射效应无法通过传统的表面渲染或简单的体密度模型得到准确表示。理想建模环境下,假设每个采样点的辐射仅来自局部发射,忽略了散射光的贡献,导致远距离物体被类似“体积雾”的散射层所遮蔽,最终在重建结果中产生不真实的漂浮伪影。其次是视角一致性的破坏。多视图三维重建的基本假设是同一三维点在不同视角下应表现出一致的本征辐射特性,但水下环境中深度依赖的衰减效应导致同一物体在不同观察距离和角度下呈现显著的外观变化。这种路径依赖的辐射变化违背了多视图一致性假设,使得多视图约束失效,进而导致几何重建精度下降和新视角合成质量恶化。

这些物理约束的存在使得传统新视角合成方法在水下环境应用中产生系统性的重建误差,具体表

现为色彩一致性缺失、几何重建精度不足以及距离相关的质量退化等多种问题。因此,针对水下场景的新视角合成技术迫切需要发展物理感知的建模框架,将水下光传输的复杂散射机制显式地纳入体渲染管线,实现场景固有属性与介质传输效应的有效解耦,从而获得高保真度的水下三维重建和新视角合成效果。

1.2 神经辐射场

神经辐射场(NeRF)为三维场景表示学习领域的革新性技术框架,通过多层感知机(multi-layer perceptron, MLP)对三维空间中连续辐射场的隐式建模,结合可微体积渲染的精密应用,实现了高保真度的新视角合成(Mildenhall等,2021)。该技术的核心创新在于将复杂三维场景的几何结构与外观信息统一编码为一个连续的五维函数。 $F:(\mathbf{x}, \mathbf{d}) \rightarrow (\sigma, \mathbf{c})$ 。通过可微分的体积渲染管线建立隐式场景表示与二维图像观测之间的直接映射关系,从而实现基于多视角图像监督的端到端场景学习。如图3所示,该图展示了从多视角输入图像到神经网络训练再到新视角合成的完整流程。

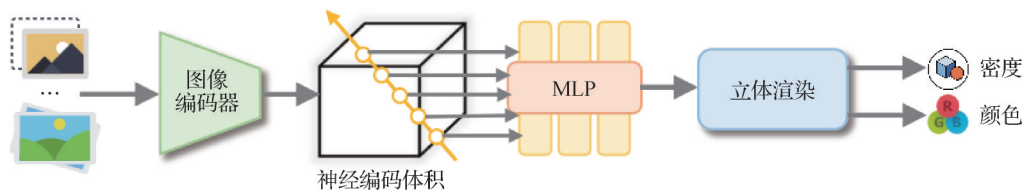


图3 基于NeRF的场景表征与对应可微神经渲染过程(Mildenhall等,2021)

Fig. 3 Scene representation based on NeRF and its differential neural rendering procedure(Mildenhall et al. ,2021)

从函数逼近理论的角度分析,NeRF将三维场景中的任意空间位置 $\mathbf{x} = (x, y, z) \in \mathbf{R}^3$ 以及对应的观察方向 $\mathbf{d} = (\theta, \varphi)$ (采用球坐标系中的方位角 θ 与仰角 φ 进行参数化)映射至由深度多层感知机 F_{θ} 参数化的连续函数空间。鉴于多层感知机在逼近高频函数时存在的频谱偏置问题,即网络倾向于优先学习低频分量而难以表达高频细节,NeRF引入位置编码机制提升网络对高频信号的表达能力。对于任意输入向量 $\mathbf{v} \in \mathbf{R}^d$,位置编码通过傅里叶特征映射定义为

$$\gamma(\mathbf{v}) = (\sin(2^0 \pi \mathbf{v}), \cos(2^0 \pi \mathbf{v}), \sin(2^1 \pi \mathbf{v}), \cos(2^1 \pi \mathbf{v}), \dots, \sin(2^{L-1} \pi \mathbf{v}), \cos(2^{L-1} \pi \mathbf{v})) \quad (2)$$

式中, L 为编码频率的最大阶数,该编码策略将输入映射到高维傅里叶特征空间,使网络能够有效学习

场景中的精细纹理和锐利边界等高频几何特征。

经过位置编码后,神经网络的前向传播过程可表示为

$$F_{\theta}(\gamma(\mathbf{x}), \gamma(\mathbf{d})) = (\sigma(\mathbf{x}), \mathbf{c}(\mathbf{x}, \mathbf{d})) \quad (3)$$

式中, $\sigma(\mathbf{x}) \in \mathbf{R}^+$ 表示空间位置 \mathbf{x} 处的体密度,量化了该点对入射光线的散射与吸收强度; $\mathbf{c}(\mathbf{x}, \mathbf{d}) \in \mathbf{R}^3$ 则表示沿特定视线方向 \mathbf{d} 观察时该点的发射辐射颜色。这种设计体现了重要的物理直觉:体密度 σ 作为几何属性仅依赖于空间位置,而射线辐射 \mathbf{c} 依赖于观察方向以建模视角相关的外观效应,如镜面反射、半透明等现象。

NeRF渲染的核心在于体积渲染(volume rendering)的物理模型,该方程描述了光线在参与介质中的传输过程。体积渲染将三维场景建模为连续的参

与介质,通过沿射线路径积分体密度和发射辐射来计算每个像素的最终颜色,从而实现从体积表示到二维图像的映射。对于从相机中心 o 出发、方向为 d 的射线 $r(t) = o + td$,体积渲染方程定义为

$$C(r) = \int_{t_n}^{t_f} T(t) \sigma(r(t)) c(r(t), d) dt \quad (4)$$

式中,透射率函数 $T(t) = \exp(-\int_{t_n}^t \sigma(r(s)) ds)$ 描述了光线传播至距离 t 时未被介质吸收的概率。该积分可以理解为3个步骤的组合:1)光线在每个位置都会收集到局部的颜色信息;2)收集到的颜色会根据透射率进行衰减,距离越远衰减越多;3)将沿途所有位置的衰减颜色累加起来,得到最终的像素颜色。将射线在区间 $[t_n, t_f]$ 内划分为 N 个采样点,假设每个区间 $[t_i, t_{i+1}]$ 内体密度和颜色保持常数,则体积渲染方程的离散形式为

$$\hat{C}(r) = \sum_{i=1}^N T_i (1 - \exp(-\sigma_i \delta_i)) c_i \quad (5)$$

式中, $\delta_i = t_{i+1} - t_i$ 为采样间距, $T_i = \prod_{j=1}^{i-1} \exp(-\sigma_j \delta_j)$ 为离散透射率, $1 - \exp(-\sigma_i \delta_i)$ 为局部不透明度 α_i 。

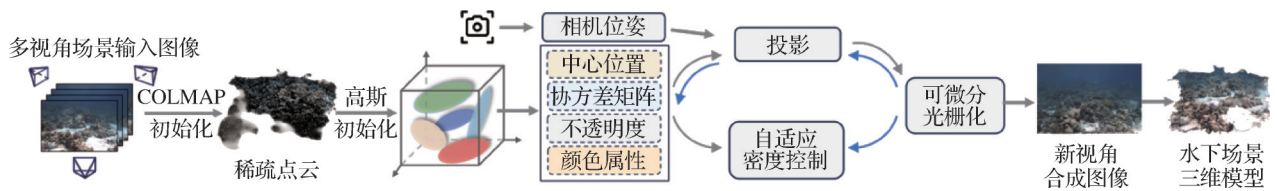


图4 3D高斯泼溅的属性初始化与渲染流程示意图(Kerbl等,2023)

Fig. 4 Overview of Gaussian attribute initialization and rendering pipeline for 3DGS(Kerbl et al., 2023)

三维高斯泼溅将场景表示为 N 个三维高斯基元的集合 $\mathcal{G} = \{G_i\}_{i=1}^N$,每个高斯基元 G_i 由以下参数完全确定:中心位置 $\mu_i \in \mathbf{R}^3$ 、三维协方差矩阵 $\Sigma_i \in \mathbf{R}^{3 \times 3}$ 、不透明度 $\alpha_i \in [0, 1]$ 、以及球谐系数 $SH_i \in \mathbf{R}^{(L+1)^2 \times 3}$ 用于编码视角相关的颜色。第 i 个高斯基元在位置 x 处的贡献定义为

$$G_i(x) = \exp\left(-\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i)\right) \quad (7)$$

为保证协方差矩阵的正定性并避免优化过程中的数值不稳定性,采用四元数—尺度分解的参数化方法,具体为

$$\Sigma_i = R_i S_i S_i^T R_i^T \quad (8)$$

这种离散化处理保证了渲染过程的完全可微性,支持基于梯度的端到端优化。

训练过程通过最小化渲染图像与真实图像之间的光度误差来优化网络参数,损失函数通常采用均方误差。具体为

$$L = \sum_{r \in \mathbf{R}} \left| \hat{C}(r) - C(r) \right|_2^2 \quad (6)$$

式中, \mathbf{R} 为训练射线集合。通过反向传播,梯度从像素级误差传递至射线上每个采样点的体密度和颜色预测,进而更新网络权重,使得学习到的隐式表示能够准确重现观测图像。

1.3 三维高斯泼溅

3D高斯泼溅(3DGS)是近年来在新视角合成领域取得突破性进展的一种新兴技术。作为显式场景表示方法,通过大量可微分的三维高斯基元对场景进行建模,结合基于点的差分渲染技术,能够实现高质量的实时新视角合成(Kerbl等,2023)。3DGS框架示意图如图4所示,该方法的核心思想是利用一组参数化的三维高斯分布函数显式表示场景的几何与外观信息,从而避免了隐式神经网络表示中频繁查询的计算开销,显著提升了渲染效率。

式中, R_i 为由单位四元数 q_i 参数化的旋转矩阵, $S_i = \text{diag}(s_x, s_y, s_z)$ 为尺度矩阵。这种分解确保了优化过程中协方差矩阵始终保持正定。

每个高斯基元的颜色通过球谐函数建模视角依赖性,具体为

$$c_i(d) = SH(sh_i, d) = \sum_{l=0}^L \sum_{m=-l}^l sh_{i,l}^m Y_l^m(d) \quad (9)$$

式中, $Y_l^m(d)$ 为球谐基函数, $SH_{i,l}^m \in \mathbf{R}^3$ 为对应的RGB系数。通常使用3阶球谐函数($L = 3$)平衡表达能力与计算复杂度。

渲染过程基于三维高斯分布投影到二维图像平面后仍保持高斯性质这一重要几何性质。给定相机

参数,三维高斯基元通过透视投影变换为二维高斯分布,其协方差矩阵通过雅可比矩阵计算。具体为

$$\Sigma' = \mathbf{J}\mathbf{W}\Sigma\mathbf{W}^T\mathbf{J}^T \quad (10)$$

式中, \mathbf{W} 为世界坐标到相机坐标的变换矩阵, \mathbf{J} 为透视投影的雅可比矩阵。

对于图像中的每个像素,计算所有相关高斯基元的贡献并进行深度排序的Alpha混合。具体为

$$C = \sum_{i \in N} T_i \alpha_i G_i(p) c_i(d) \quad (11)$$

式中, N 为影响该像素的高斯基元集合, $G_i(p)$ 为高斯基元在像素 p 处的二维高斯值, $T_i = \prod_{j=1}^{i-1} (1 - \alpha_j G_j(p))$ 为累积透射率。

为了进一步提高渲染质量和效率,三维高斯泼溅引入了自适应密度控制策略。这一策略基于视角空间中位置的梯度信息,对高斯基元的密度进行动态调整。在细节丰富的区域,例如物体边缘或复杂纹理区,系统会增加小尺度高斯基元的数量或通过分裂大尺度基元来增强细节的表达;而在细节较少的区域,系统则会减少高斯基元的密度,以降低计算开销。这种动态调整不仅确保了细节的准确呈现,还大大提高了渲染效率。此外,对于透明度较低或尺寸过大的基元,系统会定期进行修剪操作,进一步减少内存消耗和计算负担。通过自适应密度控制策略,渲染过程能够更精确地捕捉复杂场景中的高频细节,同时有效管理计算资源。

训练过程以优化像素级别的精度并保留图像结构上的细节。

重建损失函数由像素级损失(L_1 损失)与结构性相似性损失(structural similarity, SSIM)共同构成,定义为

$$L = (1 - \lambda)L_1 + \lambda L_{\text{SSIM}} \quad (12)$$

式中, $L_1 = |I - \hat{I}|_1$, $L_{\text{SSIM}} = 1 - \text{SSIM}(I, \hat{I})$, λ 为超参数权重,通常设为0.2。

相比于NeRF的体积渲染管线,3DGS通过显式参数表示与基于泼溅渲染的结合,在计算效率方面取得了显著优势。由于避免了神经网络的重复查询,且高斯基元的投影和混合过程高度并行化,该方法能够实现实时渲染。然而,显式表示也带来了内存消耗较大和大范围几何变化适应性困难等局限性。在水下场景应用中,需要特别考虑介质散射对高斯基元可见性和颜色的影响。

2 国内外代表方法

水下新视角合成作为计算机视觉与图形学的交叉前沿领域,目前仍处于快速发展阶段。现有研究主要围绕如何将先进的三维场景表示技术有效适配水下复杂光学环境展开,核心挑战集中在光学衰减建模、散射效应处理以及动态环境感知等方面。基于技术路径的不同,本文将现有方法归纳为3个主要研究方向:基于隐式神经表征的方法、基于显式几何表示的方法以及前馈网络方法,并从理论创新、技术特点和应用效果等维度进行系统性分析。其方法概括总览如图5所示。

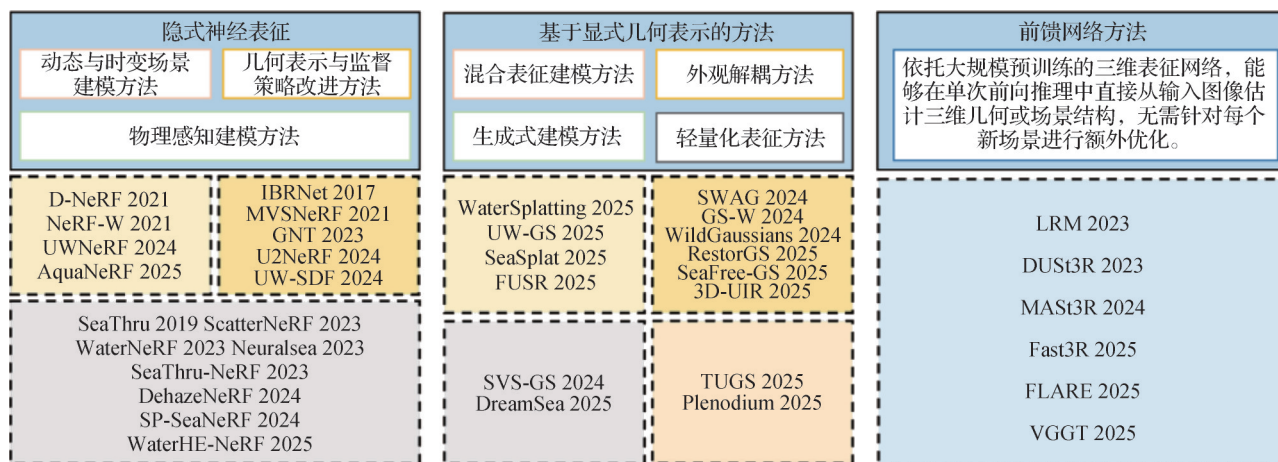


图5 国内外水下新视角合成代表方法总结概览

Fig. 5 Overview of representative underwater novel view synthesis methods

2.1 隐式神经表征

隐式神经表征以连续场函数对颜色、密度及光学传输过程进行统一建模,是水下场景表征的重要技术框架。其技术发展围绕物理真实性、时空完整性与表征能力3个维度展开。针对这些维度,现有方法可归纳为3类:物理感知建模方法融合水下光学传输机制以实现物理一致重建,动态与时变场景建模方法引入时间维度以应对环境动态扰动,几何表示与监督策略改进方法通过架构与表征创新提升建模能力。下面将围绕上述3类主导建模机制对相关方法进行系统梳理。

2.1.1 物理感知建模方法

基于物理模型驱动方法的核心思想是将水下成像的物理机制融入神经辐射场框架,通过隐式神经网络估算场景所需参数,由此建模光的吸收、散射等过程实现物理一致的场景重建。

1)技术溯源与理论启示。复杂介质环境下的物理感知建模最早在大气成像领域取得突破。Ramazzina 等人(2023)提出的 ScatterNeRF,通过将 Koschmieder 大气散射模型与神经辐射场相结合,首次验证了在散射介质中显式建模物理过程的可行性。DehazeNeRF(Chen 等,2024c)进一步将大气散射的物理过程显式集成到体积渲染管线中,实现了场景重建与去雾的统一优化。这些工作为水下场景建模提供了两个重要启示:(1)散射介质可通过物理模型参数化表征;(2)物理过程可嵌入神经渲染的可微分框架。然而,大气与水下环境存在本质差异:水下成像涉及波长依赖的衰减、更复杂的后向散射以及动态光照变化,这要求水下方法发展出更精细的物理建模策略。

2)早期全局建模范式。受到大气场景工作的推动,水下领域很快形成了自己的技术路线。WaterNeRF(Sethuraman 等,2023)首次提出了物理模型与数据驱动相结合的水下 NeRF 框架,针对水下成像中波长依赖的衰减与散射问题,在 NeRF 隐式表示的基础上利用轻量级多层感知机网络估计波长相关的衰减参数,然后利用估计出的参数根据成像模型反演得到经过颜色矫正后的场景。该方法的重要贡献在于验证了将经典水下成像物理模型嵌入神经辐射场框架的可行性,为后续研究奠定了基础性的技术路线。SeaThru-NeRF(Levy 等,2023)作为水下神经辐射场领域的重要代表性工作,进一步发展了物

理感知建模思路,系统性地将 SeaThru(Akkaynak 和 Treibitz,2019)水下成像模型融入 NeRF 框架。如图6所示,该方法通过拟合全局的水体衰减参数,实现对图像颜色的校正,并利用 NeRF 的场景表示能力完成三维重建和新视角合成。其特点在于强调物理一致性,适合水体条件较为均一的环境。然而,全局参数化假设的内在局限性逐渐显现。这一建模范式本质上将整个场景的水体视为空间均匀介质,忽略了真实水下环境中广泛存在的空间异质性。这些固有缺陷促使研究者转向探索能够捕捉空间变化特性的局部化建模方法。

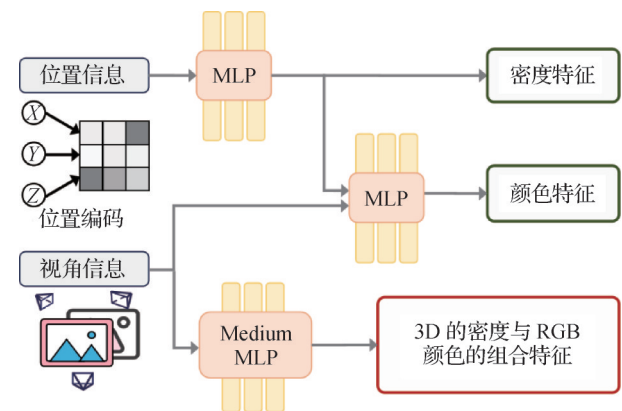


图6 水下隐式神经表征示意图

Fig. 6 Schematic of underwater implicit neural representation

3)精细化物理建模的多路径探索。针对全局建模的局限,后续研究从不同角度发展出更为精细的建模策略。Neuralsea(Zhang 和 Johnson-Roberson,2023)将神经反射场与物理成像模型相结合,能够同时学习水下环境的反照率、法向量和体积密度。该方法在神经场中引入逻辑回归函数,将水体与场景有效区分,并在训练过程中施加不同的光学传输模型。通过对复杂的水下反向散射效应进行近似建模,方法避免了直接估计所带来的困难,从而提升了采样效率与数值稳定性。SP-SeaNeRF(Chen 等,2024a)从另一角度突破了全局建模的限制,采用微型多层感知机网络预测空间变化的水柱参数。该方法将退化过程显式建模为真实颜色与散射颜色的组合,从而增强模型对散射效应的感知能力。这种散射分解策略使得模型能够感知并量化不同空间位置的散射强度差异,实现了从统一表征到空间自适应的重要转变。该方法的另一重要贡献在于对动态光照的处理,通过设计光照嵌入向量表征图像中的光

照偏差,有效缓解了动态光照对重建一致性的干扰。将每幅图像的瞬态光照分解为全局静态光照分量与图像特定的偏差项,使得模型能够在光照条件变化的开放水域环境中保持多视角一致性。为了提升计算效率,SP-SeaNeRF提出了一种新型采样模块,聚焦于对渲染贡献最大的采样点,相较于传统均匀采样策略实现了训练与推理速度的显著提升。WaterHE-NeRF (Zhou等,2025)针对水下NeRF方法在处理光学衰减和缺乏真实监督数据方面的局限,基于Retinex理论提出了一种新型的水射线匹配场,用于精确建模三维空间中的颜色、密度与照度衰减。该方法通过照度衰减机制同时生成退化与清晰的多视角图像,并结合重建损失与Wasserstein距离优化图像恢复过程。此外,WaterHE-NeRF采用直方图均衡作为伪地面实况监督,通过引入颜色分布先验约束从而在提升恢复精度的同时保持图像的细节与颜色分布一致性。解决了水下场景普遍缺乏配对真值数据的根本性问题。

物理感知建模方法的发展体现了从简单到复杂、从全局到局部的技术演进趋势。早期方法主要采用经典物理模型的直接应用,而后续工作逐步引入了更灵活的参数化策略和学习机制。这类方法的核心优势在于具有明确的物理可解释性,但在处理复杂多变的水下环境时仍面临建模精度与计算效率之间的平衡挑战。

2.1.2 动态与时变场景建模方法

传统的神经辐射场假设场景是静态的,然而真实水下环境往往存在游动鱼群、漂浮颗粒、波动水面以及时变光照等复杂动态因素,这些因素严重影响场景重建的准确性和一致性。

1)通用动态建模基础。早期动态神经辐射场研究为水下动态建模奠定了重要理论基础,D-NeRF (Pumarola等,2021)等经典工作开创了动态神经辐射场建模的先河,通过引入时间维度将静态的三维场景表示扩展为四维时空表示。这类方法通常采用变形场或时间编码的方式建模场景中物体的运动和变化,为后续处理复杂动态环境奠定了重要基础。Martin-Brualla等人(2021)则进一步扩展了动态建模的范围,提出NeRF-W(NeRF in the wild)方法专门处理真实自然环境中的挑战,如瞬态物体干扰、变化的光照条件等,通过引入外观嵌入和时变建模机制提升了复杂环境下的重建鲁棒性。

2)水下特定动态因素去除。在真实水下环境中,动态目标、悬浮颗粒以及不稳定光照往往对神经辐射场建模造成严重干扰,因此部分研究开始聚焦于动态与复杂场景的鲁棒性建模。针对真实环境中水面动态扰动带来的挑战,Xiong和Heidrich(2021)提出了一种单目相机三维重建方法,首次实现了动态水面与静态水下场景几何的联合估计方法。该方法基于可微分成像框架,将斯涅尔定律的折射建模、跨时间的三角化约束与多项正则化损失相结合,从而在单一端到端优化流程中同时迭代优化动态水气界面形态与水下场景的三维几何。该方法实现了对时变水气界面的高精度建模,并在实验室环境和野外复杂条件下都展现了良好的重建效果。面向更加复杂的自然水域,UWNeRF(Tang等,2024)提出了一种融合物理建模与神经表示的框架,旨在应对水下场景重建中距离依赖的可见性衰减,不稳定的光照条件与场景中的移动物体的挑战。如图7所示,该方法的核心框架由多个协同工作的神经模块构成,以实现复杂水下场景的精细化表征。其中,静态多层感知机负责建模场景中随时间保持不变的几何结构;动态多层感知机用于刻画运动目标,捕捉其随时间变化的密度与辐射属性;光照多层感知机则依据时间与观察方向重构水下局部光照的时变特性。上述模块的输出与共享的自适应色调映射器和水体物理参数共同参与优化,从而生成符合水下成像规律的高保真渲染结果,有效应对真实水下环境中几何动态性与光照不稳定性等多重挑战。该框架不仅显著提升了对水下动态变化与光照扰动的适应能力,还为水下效果的可控编辑提供了支持。AquaNeRF(Gough等,2025)则针对鱼群、漂浮颗粒等动态干扰提出了一种新的NeRF渲染器与优化方案。该方法的技术特点在于从渲染机制层面进行改进,通过每条射线估计单一表面并使用高斯加权函数维持介质透射率一致性,同时引入深度缩放增强近景梯度,显著减少鱼群、漂浮颗粒伪影,提升静态背景保真度。

动态与时变场景建模方法的发展体现了从通用动态处理到特定环境适配的技术演进路径。相比静态场景重建,动态建模需要额外处理时间维度的复杂性和计算开销,同时还要应对水下介质散射与动态目标的耦合效应。尽管在技术复杂度上面临挑战,这类方法为真实复杂水下环境的神经辐射场应

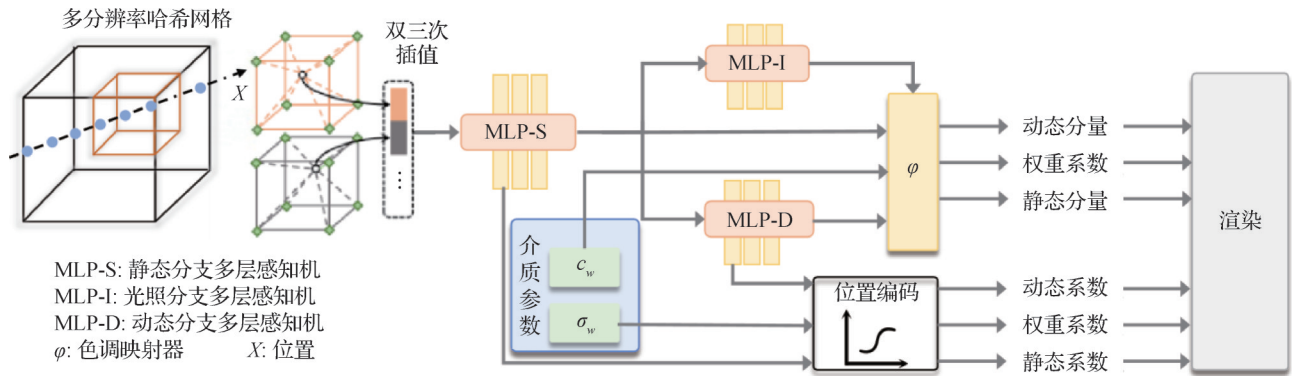


图7 基于NeRF水下动态场景建模框架 (Tang等, 2024)

Fig. 7 Framework of NeRF-based underwater dynamic scene modeling (Tang et al., 2024)

用奠定了重要技术基础,推动了该领域从理想化场景向实际应用场景的技术转移。

2.1.3 几何表示与监督策略改进方法

传统的神经辐射场通常采用逐像素处理策略,在处理复杂场景时面临特征表达能力不足的挑战。近年来,研究者从网络架构创新和几何表示改进两个角度提出了多种解决方案。

1)多视角特征聚合的架构演进。针对逐像素处理策略的表达局限,IBRNet (Wang等, 2021b)和MVSNeRF (Chen等, 2021)等工作率先探索了多视角特征聚合策略,通过融合多个视角的特征信息提升新视角合成的质量和泛化能力。这些早期工作验证了跨视角信息整合对提升场景表示能力的重要性,但其基于卷积的特征聚合方式在处理长距离依赖和复杂视角关系时仍存在局限。GNT (Varma等, 2023)将Transformer架构引入神经辐射场建模,提出了由View Transformer与Ray Transformer构成的双阶段架构,分别用于跨视角特征聚合与射线方向特征整合,从而实现了可泛化的场景表示与渲染。该架构通过注意力机制显著提升了跨视角特征交互的效率,为复杂场景下的多视角建模奠定了重要方法基础。然而,通用架构在面对水下场景特有的介质散射和颜色失真时,仍需要针对性的物理建模策略。U2NeRF (Gupta等, 2024)在此基础上提出了基于Transformer的自监督框架,在多视角几何约束下同时实现水下图像恢复与新视角合成。该方法建立在可泛化神经辐射场Transformer的基础上,通过沿极线方向聚合多视角特征,并使用堆叠的视图和射线Transformer沿射线方向进行特征整合。不同于传统NeRF的逐像素处理方式,U2NeRF将射线特征扩

展到二维补丁级别,显著增强了水下恢复的特征表达能力。将预测颜色分解为场景辐射、直接传输、后向散射与全局背景光4个部分,每个组分通过射线特征预测或使用变分自编码器从最近参考视图中获得,从而在缺乏真值监督的条件下实现有效的图像恢复。

2)从体密度到显式几何的表示革新。与网络架构创新并行发展的另一技术路线聚焦于几何表示范式的改进。神经有符号距离函数(signed distance function, SDF)作为一种重要的几何表示范式,在精确表面重建方面展现出显著优势。NeuS (Wang等, 2021a)、VolSDF (Yariv等, 2021)和UNISURF (Oechsle等, 2021)等经典工作建立了神经SDF的理论基础,通过学习连续的有符号距离场实现高质量的表面重建。这些方法相比传统的体密度表示能够提供更准确的几何边界和更清晰的表面细节,其核心优势在于将隐式的体积表示转化为显式的表面约束,使得几何重建过程具有更强的可控性和更高的精度。然而,SDF方法在通用场景中的成功并不能直接迁移到水下环境,水下成像的介质干扰和目标分割困难对几何重建提出了新的挑战。UW-SDF (Chen等, 2024d)将神经有符号距离函数技术引入水下重建领域,提出了专门的水下SDF框架。该方法在重建过程中引入混合几何先验,有效提升了神经SDF在水下环境中的重建精度与效率。考虑到水下多视角图像中目标分割一致性的技术难题,该工作进一步结合通用分割模型SAM (segment anything model) (Kirillov等, 2023),提出了一种少样本多视角目标分割策略,实现了对未知物体的快速且自动化的分割与重建。

多视角架构与几何表示创新方法通过引入先进的网络架构和精确的几何表示范式,推动了NeRF从单一表示向多元化表征的技术扩散。从通用方法向领域特化的发展趋势,即通过引入先进的网络架构和几何表示范式来突破传统NeRF在复杂环境下的表达局限,为后续跨领域技术融合和混合表示方法的发展奠定了重要基础。

2.2 基于显式几何表示的方法

显式几何表示通过参数化基元直接建模场景结构,在计算效率和实时渲染方面具有显著优势。其在水下场景中的技术发展呈现递进与互补的演进脉络,其中混合表征建模构建了几何与介质联合建模的基础架构,外观解耦在此基础上提升多视角外观一致性,生成式建模通过生成先验补偿数据稀缺的约束,轻量化表征则优化表示效率以满足实际部署。共同构成了从表示能力到实用部署的完整技术链路。下文将分别展开论述上述4类方法。

2.2.1 混合表征建模方法

在水下场景的几何与介质联合建模中,研究者们针对传统方法的不足提出了创新性的改进。传统的基于NeRF的体积渲染方法虽然能够同时建模几

何与水体介质,但在训练过程中存在较长的时间消耗,且缺乏实时渲染能力。尽管3DGS在渲染效率上具有优势,但其仅能显式表示几何结构,无法处理水体这类半透明散射介质的光学传播过程,因此在水下场景中面临表征能力的本质性缺陷。这两种表征范式各自的局限性促使研究者探索融合显式几何与隐式介质的混合表征策略。

1)混合表征框架的开创性探索。为了同时获得3DGS的渲染效率和NeRF的介质建模能力,WaterSplatting(Li等,2025)提出了一种创新的双重表示策略。如图8所示,该方法采用3DGS表示显式的几何结构,同时引入一个独立的体积场来建模散射介质,通过这种融合策略实现了几何与介质的解耦建模。WaterSplatting验证了混合表征在水下场景中的可行性,显式高斯基元负责高效的几何渲染,而隐式体积场则专门处理水体的散射与衰减效应。这种架构设计不仅能够精确重建水下场景,还能够通过移除散射介质恢复场景的真实外观,成功实现了水下场景的实时渲染,显著提升了效率和质量。然而,该方法在物理建模的精细化程度和对水下特定挑战的针对性处理方面仍有提升空间。

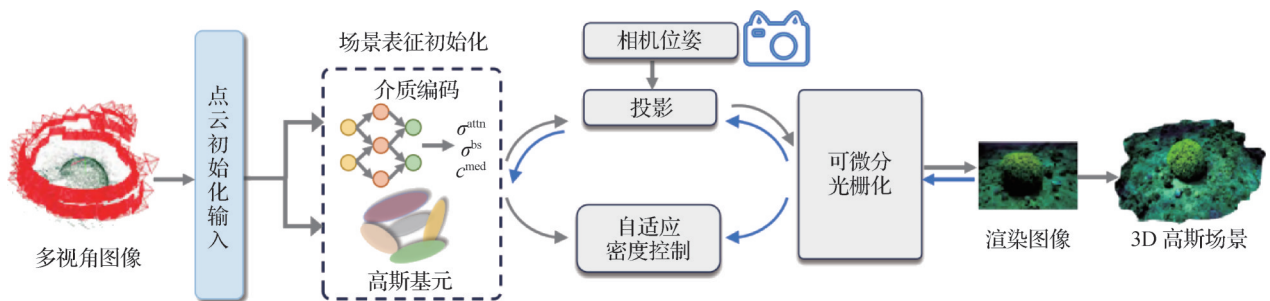


图8 显式约束下的介质表征三维重建框架图(Li等,2025)

Fig. 8 Schematic of splatting with medium 3D reconstruction under explicit constraints(Li et al. ,2025)

2)针对水下物理特性的专门化改进。尽管WaterSplatting证明了混合表征的有效性,后续研究开始针对水下环境的特定物理过程提出更精细的建模策略。UW-GS(Wang等,2025a)提出了一种更专为水下环境设计的方法,以进一步解决3DGS在水下光吸收、散射和动态目标处理中的不足。UW-GS通过引入距离相关的颜色建模来准确刻画水体衰减效应,并结合物理密度控制策略,提升远距离目标的清晰度。同时,该方法利用二值运动掩码有效处理动态物体,设计了适用于散射介质的专用损失函数,并通过伪深度图辅助优化训练过程,进一步提升了

水下场景重建的精度。UW-GS的核心在于将物理先验深度整合到优化过程中,使得介质建模更加符合水下光学传播规律。SeaSplat(Yang等,2025a)则从另一个角度推进了混合表征的发展,重点解决光波长依赖性和传播距离的耦合效应问题。通过同时约束几何与水体介质效应,SeaSplat不仅提升了新视角渲染的质量,还能够有效去除水体介质对场景的影响,恢复出场景的真实颜色,同时获得更为精确的深度图。这种强调联合优化的策略与UW-GS侧重物理密度控制的方法形成互补,展现了混合表征框架下不同的技术实现路径。

3)从监督依赖到自监督学习的演进。混合表征方法的另一重要演进方向体现在对监督信息的依赖程度上。FUSR (Hu 和 Liu, 2025)提出了双分支设计,分别通过深度分支和介质分支提高几何与水体介质建模的精度。深度分支利用MVS(multi-view stereo)代体积提升几何精度,而介质分支则基于物理成像模型估计水体散射参数,实现精确的渲染。FUSR方法的创新之处在于其无需深度标注便可进行训练,通过结合颜色多层感知机(MLP),能够恢复水下场景的真实颜色并生成高保真的几何表示。

混合表征建模方法的提出标志着从单一几何表示向复合场景建模的技术演进。通过这些架构层面的创新,这些方法有效克服了传统3DGS无法处理水体介质的根本性限制,为显式几何表示在水下环境中的应用奠定了坚实的技术基础。

2.2.2 外观解耦方法

在不受约束的场景多视角图像中,由于不同视角下场景介质密度与照度变化的视角不一致性,容易引发重影和伪影等问题。针对无约束自然环境的复杂光照和外观变化,研究者们首先在陆地场景中探索了多种外观建模策略。SWAG (Dahmani 等, 2024)通过调节高斯颜色和构建图像依赖的透明度变化,扩展了传统高斯建模的外观建模能力。针对自然场景中的光照变化和瞬态物体进行建模,使得方法能够更加鲁棒地应对这些挑战。通过调整颜色和透明度的动态变化,提高了合成图像在多变环境下的稳定性和质量。GS-W (Gaussians splatting in the wild) (Zhang 等, 2024a)聚焦于通过将外观分解为内在和动态特征来更精细地建模复杂场景。GS-W将外观建模分为固有的属性和环境相关的动态组分,使得在多变光照条件下保持更高的精度和一致性。该方法适用于处理多视角图像集合,能够有效应对不同环境光照变化对场景重建的影响。WildGaussians (Kulhanek 等, 2024)通过引入外观建模和不确定性预测机制,有效克服了复杂场景中的遮挡、光照变化和外观不一致问题。这种方法为每个高斯点嵌入了可训练的外观和图像特征,使得渲染结果能够动态适应不同视角和光照条件,同时通过不确定性估计进一步减少了动态遮挡对图像质量的干扰。在高遮挡、复杂光照以及大规模场景重建任务中表现尤为突出。

1)外观建模与水下挑战的适配。水下场景与

户外无约束场景具有相似性,都面临多视角不一致和动态外观变化的挑战,但水下环境的复杂性更为突出。在复杂水体介质对外观信息采集的干扰下,外观与几何信息的高度耦合使得图像合成和恢复过程愈加复杂。相比陆地场景主要受光照和遮挡影响,水下场景还需同时处理距离相关的颜色衰减和悬浮颗粒散射,这要求外观解耦方法不仅要分离视角变化引起的外观差异,更要区分物体真实颜色与介质退化效应。针对浅水区海底图像中普遍存在的“水焦散”效应,Zhang 等(2025c)提出 RecGS,将焦散作为可分离的低频结构进行建模。RecGS基于3DGS,通过迭代式场景构建,在每次迭代中从渲染结果与原图的残差中提取低频焦散模式(通过二维低通滤波获得),并逐步从训练图像中移除,从而有效消除焦散伪影,获得照明一致性更高的三维场景外观。针对水下真实外观难以估计的问题,SeaFree-GS(Liu 等, 2025)提出了一种退化感知的双重颜色建模策略来恢复水下成像中难以估算真实外观的固有颜色和一个模拟水体效应的退化颜色,并通过前后景分离与粗粒度深度约束,提升了几何结构的稳定性和跨视角的外观一致性。该方法特别适合处理水下复杂光照条件下的场景重建。

2)深度感知与外观解耦的深度融合。近年来研究者们提出了一类外观解耦与退化感知方法,旨在通过区分清晰视图与退化视图,并结合深度引导优化策略,提升水下场景的恢复与渲染质量。RestorGS (Qiao 等, 2025)针对传统3DGS在处理多视角不一致和退化场景时表现不佳的问题,提出了一种深度感知的高斯渲染框架,引入外观解耦与深度感知建模用于高效的场景恢复。该方法通过球谐函数学习区分清晰视图与退化视图,并结合深度引导机制约束退化效应的分布,使模型能够更准确地分离物体外观与介质影响。在水下、雾霾和低光环境下,RestorGS在保持实时渲染效率的同时,有效缓解了颜色漂移和几何失真问题。3D-UIR (Yuan 等, 2025)进一步将物理成像模型与外观建模机制结合在高斯泼溅框架中。相比RestorGS主要通过深度约束退化分布的策略,3D-UIR的核心区别在于将水下成像的物理过程显式整合到外观建模中,使得退化建模不仅符合数据分布,更遵循光学传播规

律。如图9所示,该方法构建了一个多分支协同的建模框架。首先,通过专门的高斯建模将物体外观与水介质效应进行显式分离。在此基础上,水下外观建模分支结合外观特征和嵌入来处理视图一致性难题,确保不同视角下物体真实颜色的稳定性。同时,散射介质建模分支采用物理驱动的策略,分别对后向散射和距离相关的衰减效应进行精细化建模。为了进一步提升几何精度和参数估计的可靠性,引入深度引导正则化优化机制,利用伪深度图约束空间分布,有效减少了退化建模中的伪影和

不确定性。所有的组件通过基于物理的水下图像形成模型在可微光栅化过程中进行统一整合,实现了端到端的联合优化。3D-UIR利用外观嵌入将物体纹理与介质效应(如衰减和散射)分别建模,同时引入深度监督与几何正则化来提升结构精度和减少伪影。凭借这种物理一致性的建模方式,3D-UIR能够在复杂水下环境中实现更高保真度的新视角渲染与场景恢复,为外观解耦方法从纯数据驱动向物理约束与数据驱动相结合的方向演进树立了重要范例。

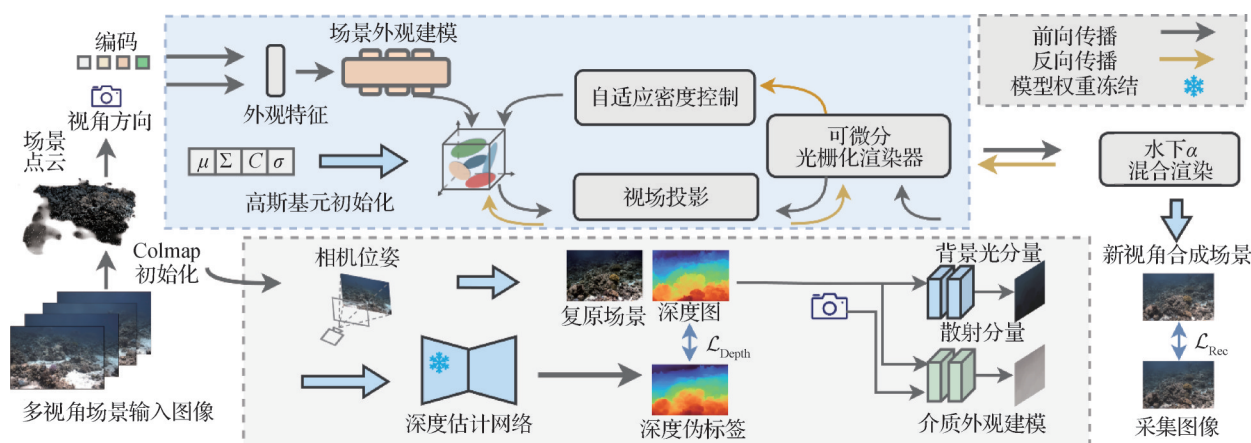


图9 基于高斯泼溅的水下场景外观建模示意图

Fig. 9 Schematic of underwater scene appearance modeling based on Gaussian splatting

外观解耦方法的发展展示了新视角合成由通用场景向水下特化场景的转变趋势。此类方法在外观层面上将物体的真实颜色与由水体散射、吸收引起的退化成分加以区分,并结合物理一致性约束,减少了因视角变化带来的光度不一致和伪影。为真实海况下高保真且鲁棒的新视角合成提供了可操作的技术路径。

2.2.3 生成式建模方法

随着生成模型与扩散模型技术的成熟,为水下场景重建提供了新的思路,以解决退化严重、标注稀少与视角稀疏的问题。通过引入生成先验与自监督约束,在缺乏显式几何或成对监督的条件下,依然能够实现高保真度的场景恢复与新视角合成。

1) 基于扩散先验的几何增强策略。SVS-GS (Chen等, 2024b)将二维扩散模型与3DGS相结合,通过分数蒸馏抽样(score distillation sampling, SDS)损失引入生成先验,提升稀疏视角条件下的几何一致性。同时,借助深度梯度先验与高斯平滑滤波,有

效抑制了伪影,显著改善了退化环境中的合成质量。然而,这一方法仍然依赖于结构从运动(structure from motion, SfM)进行位姿估计,且需要一定质量的输入图像作为基础,这在某些极端水下环境中仍存在局限性。

2) 从监督依赖到生成驱动的范式转变。针对SVS-GS仍需依赖SfM位姿估计和输入图像质量的局限性, DreamSea (Zhang等, 2025b)则提出了一种结合扩散模型与视觉基础模型(visual foundation model, VFM)的重建框架。该方法首先利用潜在扩散模型从未标注的RGB图像生成高保真海底RGBD数据,其中潜在嵌入通过分形分布建模地形细节;随后将生成的RGBD数据融合为三维地图,并在3DGS中引入二维扩散先验监督,实现无需SfM位姿估计、深度传感器或人工标注的高质量新视角渲染。

生成式建模方法有效缓解了水下场景数据获取困难和质量不足的问题。通过引入强大的生成先

验,这类方法在应对稀疏视角、退化成像以及复杂水下环境时表现出更强的鲁棒性和泛化能力,为高质量水下新视角合成提供了有力支持。

2.2.4 轻量化表征方法

3DGS作为一种显式建模方式,依赖于大规模高斯基元进行高质量渲染,带来显著的存储与计算资源消耗。在水下三维重建的实际应用中,无人潜航器和水下作业平台往往面临计算资源和存储容量的严格限制,如何在保证重建精度的同时降低计算与存储开销,成为系统实用化部署的关键技术挑战。TUGS(tensorized underwater Gaussian splatting)(Lian等,2025)提出了一种基于张量化高阶高斯的紧凑表示框架。该方法通过张量分解将高维高斯表示压缩为低维张量形式,显著减少参数数量与存储需求。同时,引入物理驱动的自适应水体估计模块,在渲染过程中联合建模光衰减与反向散射效应,从而实现几何与介质交互的一体化表示。此外,TUGS还设计了张量化的密度控制策略与专门的水下重建损失,在保证紧凑性的同时维持较高的几何精度与渲染质量,适合嵌入式与资源受限平台的部署。相比TUGS通过数学变换压缩现有表示,Plenodium(Wu等,2025)则从表征范式层面进行了改进,提出了一种新型的全光场介质表示。将方向与空间位置信息通过球谐函数联合编码,用于精细刻画水体介质的散射与吸收特性。与依赖于稀疏点云初始化的传统方法不同,Plenodium通过伪深度Gaussian补全机制增强COLMAP成的点云分布,并引入深度排序正则化损失,提升了几何一致性与深度结构的稳定性。依托这一创新的显式介质表示框架,Plenodium在复杂退化环境下实现了高保真的场景重建与高效渲染。轻量化表示方法通过创新的数据结构和表示范式,旨在在保持重建质量的同时显著降低计算复杂度和存储需求,提升系统的实用性和部署效率。

2.3 前馈网络方法

近年来,前馈式(feed-forward)三维重建技术逐渐成为传统逐场景优化方法的有力替代方案。该类方法依托于大规模预训练的三维表征网络,能够在单次前向推理中直接从输入图像估计三维几何或场景结构,无需针对每个新场景进行额外优化。相比传统几何表征方法,前馈式方法借助大规模数据集学习到的跨场景几何与外观先验,实现了快速而高

效的三维重建,无需针对每个场景进行额外优化,在计算效率与泛化能力上具有显著优势。

在三维重建与生成领域,直接从大规模三维数据集学习三维表示已成为显著趋势。这类方法通常采用大规模Transformer架构,基于多种输入生成三维表示,包括单幅图像、文本描述、有姿态多视角图像以及无姿态多视角图像等。LRM(large reconstruction model)(Hong等,2024)使用基于Transformer的前馈网络从单幅图像回归三维物体表示,通过在大规模数据集上的预训练实现了对复杂几何结构的快速重建。DUS3R(Wang等,2024)则提出以稠密点图(pointmap)为核心的表征方式,通过Transformer框架在图像对之间直接预测三维点云坐标及其对应关系,避免了传统基于SfM管道对相机模型与几何约束的依赖,其框架示意图如图10所示,通过Transformer编码器分别提取两个输入视图的深层特征表示,捕捉图像的语义和几何信息;随后,通过交叉注意力机制实现两个视图之间的特征交互与对应关系建立,这一设计使得网络能够隐式学习视图间的几何约束;最后,解码器直接在每个像素位置预测三维点云坐标,生成稠密的点图表示。这种端到端的设计避免了传统基于SfM管道对相机模型与几何约束的显式依赖,将相机位姿估计、深度预测与像素匹配等任务统一在一个框架中,显著简化了多视图重建流程。在此基础上,MASt3R(Leroy等,2024)引入显式立体约束与匹配注意力机制,将稠密特征匹配与点图预测统一优化,显著提升了跨场景的几何一致性与位姿精度。尽管上述方法在双视图场景中取得良好效果,但在扩展至大规模复杂场景时,仍需借助全局对齐步骤,该过程不仅计算开销大且稳定性不足。

针对基于DUS3R架构的方法在大规模场景扩展时仍需依赖全局对齐的问题,后续工作进行了改进。Fast3R(Yang等,2025b)通过高效的网络设计实现了对上千幅图像的同时推理,有效避免了传统全局对齐步骤带来的高时耗与不稳定性。另一类方法则尝试分解学习任务,如FLARE(Zhang等,2025a)先预测相机位姿,再估计场景几何,以降低整体学习难度。这种任务分解策略虽然在一定程度上缓解了学习复杂度,但仍需要多阶段处理,无法实现真正的端到端推理。Wang等人(2025b)进一步推进了前馈式重建的发展,提出了通用视觉几何大模型VGGT

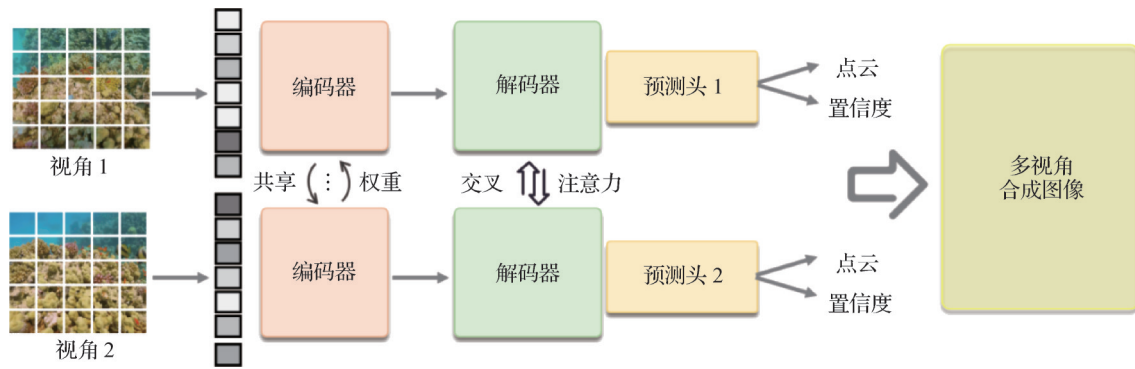


图 10 基于成对密集匹配的前馈式几何重建框架示意图(Wang 等, 2024)

Fig. 10 Schematic of feed-forward geometric reconstruction framework based on pairwise dense matching(Wang et al. ,2024)

框架。如图 11 所示,该框架通过统一的架构设计实现了多种几何属性的联合预测:输入端采用多视角图像序列,经过大规模 Transformer 编码器进行特征提取和跨视图信息融合;在解码阶段,框架通过多任务学习机制同时输出相机位姿、深度图、稠密点图及跨帧三维轨迹等多种几何属性。这种统一建模的设计使得不同几何任务能够相互促进、联合优化,相比

前述方法分别处理单一任务或需要多阶段推理, VGGT 实现了真正意义上的端到端多任务几何推理。在大规模三维标注数据集的支撑下, VGGT 在多项几何基准测试中已达到或超越现有优化型方法的性能,充分展示了前馈式重建在大模型范式下的发展潜力,代表了从特定任务网络向通用几何理解模型演进的重要里程碑。

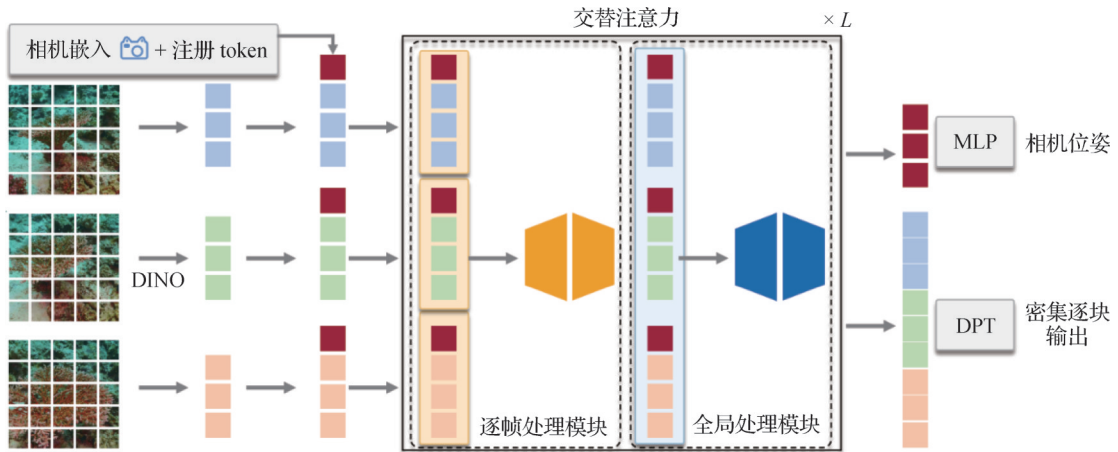


图 11 基于前馈式的点图重建框架示意图(Wang 等, 2025b)

Fig. 11 Framework of feed-forward pointmap reconstruction(Wang et al. ,2025b)

尽管前馈式方法在陆地场景中取得了突破性进展,但其对水下环境的适应性仍然有限。当前方法普遍缺乏对水下复杂光学过程(如介质散射、波长相关衰减、低能见度与动态照明等)的显式建模,容易在水下重建中出现精度下降、几何细节缺失和颜色失真等问题,难以满足工程化部署需求。相比之下,传统逐场景优化方法虽然可在特定环境中奏效,但需要为每个新场景重新训练,泛化性与时效性不足,难以支撑实际海洋作业的实时性与鲁棒性。因此,结合水下成像物理模型的前馈式网络具有广阔的发

展潜力。通过在网络架构中显式融入水下成像机理,并利用大规模跨场景预训练先验,有望在保证重建质量的同时显著提升计算效率,突破现有方法在泛化与效率上的双重瓶颈,从而推动水下三维重建与新视角合成的实用化发展。

3 性能评估与对比分析

3.1 数据集

数据集是推动水下新视角合成研究的重要基

础。根据获取方式的不同,现有数据集大体可以分为两类:一类是直接采集于水下环境的真实数据集,能够真实反映复杂的成像条件;另一类是基于物理模型或算法生成的合成数据集,在可控条件下提供精确的地面真值。

3.1.1 水下场景数据集

真实世界数据集能够充分反映复杂多变的环境条件,是验证模型在实际应用中鲁棒性和泛化能力的关键。然而,在散射介质中采集高质量、多视角的图像,并精确获取相机姿态和介质参数,往往需要专业设备与复杂流程。

1)静态场景数据集。SeaThru-NeRF 构建并使用了3个独特的真实世界水下多图像数据集,如图

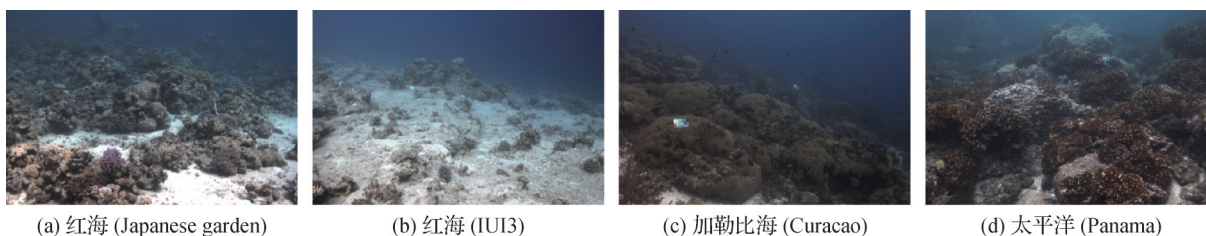


图12 SeaThru-NeRF数据集示例

Fig. 12 Samples of SeaThru-NeRF dataset ((a) red sea (Japanese garden); (b) red sea (IUI3); (c) Caribbean sea (Curacao); (d) Pacific sea (Panama))

2)动态场景数据集。针对水下环境中普遍存在的动态干扰,Tang等人(2024)提出了U-IW (underwater in the wild)数据集。如图13所示,该数据集由4个从互联网上收集的单目视频构成,拍摄于中等深度水域,使用消费级相机采集。其突出特点在于真实环境下的3类挑战:动态物体(如游动鱼群、摆动植物)、不稳定光照(由散射与光照变化引起)以及随距离增强的颜色衰减。此外,该数据集还提供了动态物体的识别与遮罩标注,便于验证方法在处理浮动生物与漂浮碎片时的有效性。因此,U-IW数据集为研究动态建模与稳健渲染提供了重要实验基础。

3)大规模前视成像场景数据集。针对水下前视方向视觉任务的特殊挑战和现有数据集规模有限的问题,FLSea(Randall和Treibitz,2023)数据集提供了目前最大规模的水下前视多传感器数据集。该数据集包含13个独立的数据序列,分别采集自地中海和红海的浅水区域,涵盖了不同的海洋环境条件。数据集采用双目立体相机和视觉惯性系统两种配置,

12所示,分别采集自红海、加勒比海(库拉索)和太平洋(巴拿马),覆盖了多样化的水体光学条件和场景复杂性。每个地点约包含20幅图像(红海20幅、库拉索20幅、巴拿马18幅),并选取其中3幅作为验证集。采集过程使用Nikon D850单反相机及Nauticam水下相机壳,并采用穹顶端口(dome port)(Treibitz等,2012)缓解折射对传统针孔相机模型的影响,姿态信息通过COLMAP(Schönberger和Frahm,2016)精确提取。图像以RAW格式获取,并经过下采样、线性白平衡及像素裁剪等预处理。SeaThru-NeRF数据集已成为该领域的基准数据集,广泛应用于相关算法的性能评测和对比,为验证方法在不同水体条件下的有效性提供了重要的测试基础。

总计包含5个立体数据集和8个视觉惯性数据集,提供了数千幅图像以及对应的地面真值深度图,示意图如图14所示。该数据集的独特价值在于其前视配置和大规模特性,涵盖了不同的水下能见度、光照条件、自然和人造结构以及动态相机运动,为水下新视角合成算法在前视导航场景下的性能验证提供了重要的数据基础。

3.1.2 合成数据集

合成数据集在受控条件下评估模型性能、进行大概摸的数据扩充和获取精确地面真实值方面具有不可替代的优势。

1)基于成像模型反演的合成方式。水下成像模型为构建配对的清晰—退化图像数据提供了重要的物理基础。该合成方式的核心思路是利用水下成像模型的正向过程(如SeaThru模型(Akkaynak和Treibitz,2019)),通过给定的水体参数和深度信息,从清晰场景图像计算生成对应的水下退化图像。具体而言,合成过程首先获取清晰的陆地场景图像作为真实场景辐射,并通过现有的新视角合成方法或

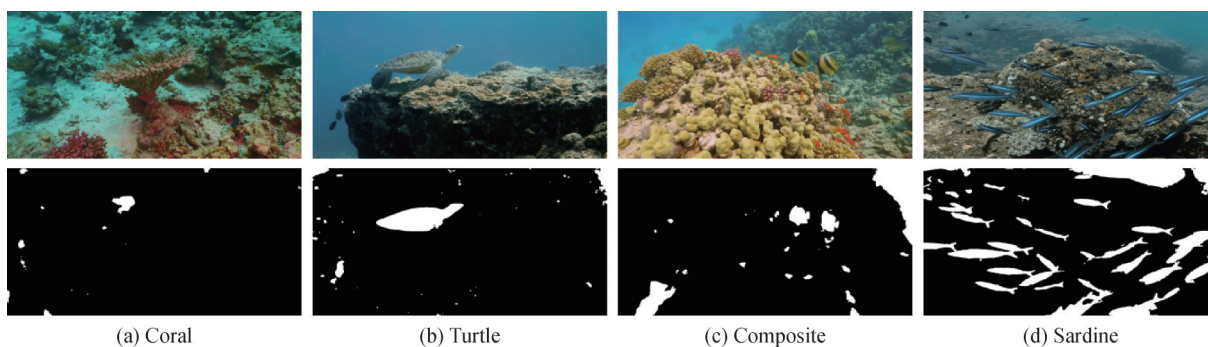


图13 U-IW数据集采集图像及运动掩码示例

Fig. 13 Samples of U-IW dataset((a)Coral;(b) Turtle; (c)Composite; (d)Sardine)

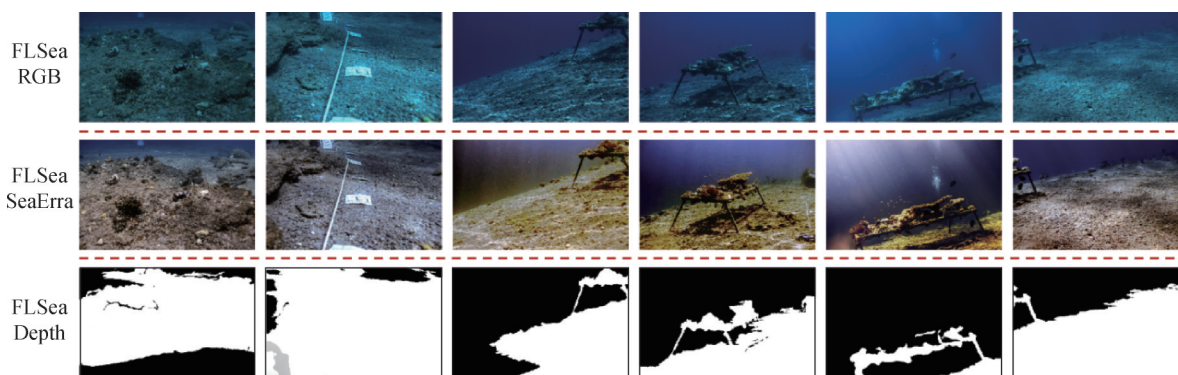


图14 FLSea数据集样例(RGB图像(上)、几何真值(中)和深度估计(下)结果)

Fig. 14 Samples of from FLSea dataset(RGB images (top), ground truth (middle), and depth estimation results (bottom))

深度估计技术获得对应的场景深度图。随后,根据特定的水下环境条件,设定波长依赖的衰减系数、后向散射系数和背景光强度等关键参数,通过成像模型计算得到对应的水下观测图像。该方法优势在于严格遵循水下成像的物理过程,确保了退化图像与真实水下成像机制的高度一致性,具有更强的理论基础和参数可解释性。

2)基于数据驱动的合成方式。现有的水下图像增强研究为水下新视角合成的数据构建提供了重要的技术基础。Li等人(2020)利用不同水体类型的衰减系数,通过水下成像物理模型对清晰图像进行退化处理,成功构建了大规模的水下图像数据集。对包含精确的相机内外参数和场景的三维几何信息的高质量陆地多视角图像数据集、对数据集中每个视角的清晰图像独立进行水下退化处理。在保持多视角间的物理一致性前提下,该方法能够从有限的陆地多视角数据快速生成涵盖不同水下环境的大批量训练数据。这类合成方式不仅大幅降低了数据获取成本,还提供了精确的地面真值(清晰场景、深度图和水体参数等),为水下新视角合成算法的监督学

习和定量评估提供了重要基础。

3)基于扩散模型的生成方式。生成式扩散模型为水下场景数据合成提供了全新的技术路径。利用大规模预训练扩散模型的强大生成能力,通过条件控制机制从有限的输入条件生成高质量、多样化的水下场景图像及其对应的密集标注信息。Zhang等人(2024b)提出的Atlantis方法开创了扩散模型在水下深度数据生成的先河。该方法首先构建包含水下图像、深度图和文本描述的三元组训练数据,然后训练专门的水下生成控制网络学习从陆地深度图到水下图像的条件映射关系。在数据生成阶段,利用陆地场景的深度信息作为几何结构指导,结合描述性文本提示,通过预训练的Stable Diffusion模型(Rombach等,2022)生成具有真实水下视觉效果图像。该方法巧妙地将陆地场景丰富的深度数据资源转化为水下环境的训练样本,有效缓解了水下深度估计数据稀缺的问题。在Atlantis基础上,Lin等人(2025)进一步提出了TIDE统一生成框架,实现了从单一文本条件到多类型密集标注的端到端生成。TIDE通过隐式布局共享机制确保文本到图像、文本

到深度、文本到语义掩码3个并行分支之间的空间一致性,同时引入时间自适应归一化技术实现跨模态特征交互,进一步提升生成数据的质量和一致性。相比于 Atlantis 需要预先提供深度图作为条件, TIDE 仅需简单的文本描述即可同时生成水下图像及其完整的密集标注信息。

基于扩散模型的生成方式在水下新视角合成数据构建方面展现出巨大潜力,为解决真实水下多视角数据稀缺问题提供了可行的技术路径。然而,要使该技术真正适用于高精度的水下三维重建任务,仍需在多视角几何一致性建模和水下光学物理约束两个关键方向上取得突破。未来的研究应重点关注如何将3D几何先验知识融入扩散生成过程,以及如何在大规模数据生成的同时保持严格的物理合理性,这将是推动水下新视角合成技术从数据增强向精确三维重建应用发展的重要研究课题。

3.2 主要评价指标

在水下新视角合成任务中,评价指标既是衡量方法优劣的重要依据,也是推动模型改进的关键环节。现有研究普遍采用来自图像合成与重建领域的经典指标,包括峰值信噪比(peak signal-to-noise ratio, PSNR)、结构相似性(structural similarity index, SSIM)和图像质量评价指标(learned perceptual image patch similarity, LPIPS)。其中,PSNR基于像素级误差衡量合成图像与参考图像的差异,能够反映整体失真程度,数值越高表示结果与真实图像越接近。但由于其仅考虑低层统计特征,往往难以准确捕捉视觉感知差异。SSIM则从亮度、对比度和结构信息出发,对两幅图像的相似性进行建模,更符合人眼感知规律,因此常作为PSNR的重要补充。相比之下,LPIPS利用深度神经网络提取的感知特征衡量视觉差异,能够更好地反映主观质量感受,在复杂场景下对细节和结构的保真度具有更强的区分能力。

尽管上述指标在通用的图像合成与新视角合成任务中得到广泛采用,但在水下环境下仍存在一定局限。PSNR和SSIM偏向低层统计差异,难以表征因光学退化导致的感知质量下降;LPIPS虽然能反映深层特征差异,但其特征提取依赖在自然场景上训练的模型,未必能完全适应水下特有的退化模式。因此,研究者提出了专门面向水下图像的无参考质量指标,如UIQM(underwater image quality measure)

(Panetta等,2016)和UCIQE(underwater color image quality evaluation)(Yang和Sowmya,2015)。这类指标从对比度、清晰度和色彩偏移等多个维度综合评估水下图像质量,能够更好地反映散射、模糊与颜色失真等问题,为复杂自然水下环境中的新视角合成结果提供了更具针对性的评价参考。

水下新视角合成评价体系需在几何一致性、感知质量和水下光学特性之间建立更加全面的度量框架,从而推动该领域向更真实、更鲁棒的方向发展。

3.3 实验对比

为了全面评估水下新视角合成领域的技术发展现状,本节基于标准化的SeaThru-NeRF(Levy等,2023)数据集对当前代表性方法进行系统性的对比研究。所选取的方法涵盖了该领域的主要技术路线:1)基于神经辐射场的方法(SeaThru-NeRF);2)基于显式高斯表示的经典方法(3DGS);3)专门面向水下环境的改进方法(RecGS(Zhang等,2025c)、SeaSplat(Yang等,2025a)、Water-Splatting(Li等,2025)、3D-UIR(Yuan等,2025)、Plenodium(Wu等,2025))。评估数据集包含Curacao、IUI3-RedSea、J. G. RedSea和Panama4个具有代表性的水下场景,这些场景在水下光学衰减程度、场景几何复杂度、成像距离及水质条件等关键因素方面具有显著差异,能够有效验证各类方法在多样化水下环境中的适应性和鲁棒性。

表1综合展示了不同技术路线的代表性方法在标准水下数据集上的定量性能表现。综合对比结果表明,各类方法在不同水下场景中显示出显著的性能差异和技术互补特征,反映了当前该领域技术发展的多元化特点。从数据统计结果来看,Plenodium方法作为新兴的水下重建技术,在多数评估指标上展现出最为突出的综合性能。该方法在Curacao场景中取得了最高的PSNR(32.469 dB)和SSIM(0.910),在IUI3-RedSea场景同样实现了最优的PSNR(28.548 dB),并在J. G. RedSea和Panama场景的LPIPS感知质量指标上均获得最佳表现(分别为0.169和0.157)。这一全面领先的性能表明,Plenodium在结合显式表示、物理建模与高效优化策略方面取得了重要突破。3D-UIR方法同样展现出卓越性能,特别是在结构相似性指标SSIM上表现突出,在Panama场景取得了最高的PSNR(29.816 dB)和SSIM(0.900),验证了其在跨视角一致性维护和几何重建精度方面的技术优势。Water-Splatting作为

融合物理建模与显式表示的代表性方法,在 Curacao 场景中在感知质量指标 LPIPS 和结构相似性指标上取得了显著的技术突破,验证了其结合显式高斯表示与独立体积场建模水下散射介质的创新技术路径。SeaSplat 作为物理驱动的水下重建方法,在 IUI3-RedSea 场景中在感知质量指标上取得了最优性能,表明了 3DGS 与水下物理成像模型深度融合的技术路线具有良好的发展前景。SeaThru-NeRF 作为早期基于神经辐射场的水下新视角合成方法,在某些特定场景中仍然展现出一定的技术优势,如在 Curacao 场景的 PSNR 和 Panama 场景的 LPIPS 指标上表现优秀。这一结果表明,神经辐射场在处理

复杂水下光学环境方面仍具有一定的技术价值。相比之下,未经水下专门优化的传统 3DGS 方法在整体性能上表现出显著局限性,这也从侧面验证了水下环境中专门化技术优化的必要性。综合对比结果显示,当前水下新视角合成领域各评估指标的最优性能往往分散于不同的技术路线之间,这一现象深度反映了水下视角合成任务的本质复杂性和多维度挑战性。这种技术现状表明,现有方法在应对水下光学衰减、介质散射以及几何重建等核心技术难题时各有侧重和特色,尚未形成统一的技术范式。这也从一个侧面说明了该领域仍处于快速发展阶段,具有广阔的技术创新空间。

表1 不同方法在 SeaThru-NeRF 数据集上的定量结果对比 (PSNR, SSIM, LPIPS)

Table 1 Quantitative evaluation of different methods on the SeaThru-NeRF dataset in terms of PSNR, SSIM, and LPIPS

方法	Curacao 场景			IUI3-RedSea 场景			J. G. RedSea 场景			Panama 场景		
	PSNR/dB	SSIM	LPIPS	PSNR/dB	SSIM	LPIPS	PSNR/dB	SSIM	LPIPS	PSNR/dB	SSIM	LPIPS
SeaThru-NeRF (Levy 等, 2023)	30.870	0.891	0.222	26.748	0.824	0.281	23.213	0.844	0.193	28.747	0.881	<u>0.162</u>
3DGS (Kerbl 等, 2023)	29.601	0.893	0.220	24.912	0.826	0.290	21.316	0.835	0.223	28.921	0.884	0.217
RecGS (Zhang 等, 2025c)	29.313	0.893	0.211	24.363	0.823	0.289	21.812	0.834	0.224	27.829	0.879	0.208
SeaSplat (Yang 等, 2025a)	30.768	0.898	<i>0.186</i>	27.493	<i>0.832</i>	0.213	22.698	0.845	0.196	28.580	<i>0.887</i>	0.180
Water-Splatting (Li 等, 2025)	30.583	<i>0.904</i>	<u>0.174</u>	27.593	0.831	0.275	<i>23.363</i>	<i>0.845</i>	<i>0.178</i>	28.751	0.881	<i>0.163</i>
3D-UIR (Yuan 等, 2025)	<u>30.978</u>	<u>0.907</u>	0.187	<u>28.301</u>	0.841	<u>0.252</u>	<u>23.371</u>	0.857	<u>0.187</u>	29.816	0.900	0.181
Plenodium (Wu 等, 2025)	32.469	0.910	0.169	28.548	<u>0.840</u>	<i>0.268</i>	24.059	<u>0.856</u>	0.169	<u>29.266</u>	<u>0.887</u>	0.157

注:加粗、下划线、斜体字体分别表示各列最优、次优、第三优结果。

表2从实用化应用的角度,综合评估了各类技术路线在渲染质量与计算效率之间的技术权衡特征。评估框架涵盖了视觉质量维度和计算效率维度(实时渲染性能FPS(frames per second)、训练时间成本),为评估不同技术路线的实际应用价值提供了全面的参考依据。从数据统计结果可以观察到,近年来基于显式表示的水下专用方法在整体性能上已经超越了早期基于NeRF的方法,这一发展趋势反映了该领域技术路线的重要转变。其中,Plenodium方法实现了质量与效率的突破性平衡,在所有测试场景上取得了最高的平均PSNR(28.585 dB)和最优的LPIPS(0.191),同时实现了超实时的渲染速度(240.75 帧/s)和相对较短的训练时间(24 min)。这一卓越表现标志着水下新视角合成技术在实用化部署方面取得了重要进展,其超实时的渲染性能甚至超越了传统的3DGS方法(149.36 帧/s),同时保持了

远超后者的重建质量。这种技术优势的形成主要得益于显式高斯表示在物理过程建模的直接性和渲染计算的高效性,这也为未来水下新视角合成技术的发展方向提供了重要的技术参考。从实用化应用的角度来看,不同技术路线在计算效率与渲染质量之间展现出了不同的技术特征。传统3DGS方法以极高的帧率展示了显式表示在实时渲染方面的天然优势,但在水下场景的质量表现上存在明显不足。水下专用方法中,3D-UIR和SeaSplat的渲染性能实现了质量与效率的良好平衡,这一成果为水下新视角合成技术的实际应用部署提供了可能。相比之下,早期的SeaThru-NeRF方法仅能实现极低帧率,这一结果反映了隐式表示在水下复杂环境中的计算效率瓶颈。从技术部署的角度来看,不同技术路线在训练效率方面展现出了显著差异。基于显式表示的方法普遍具有训练效率优势,其中传统3DGS仅需数

表2 现有方法在所有测试场景上的平均重建质量与运行效率对比

Table 2 Comparison of the average reconstruction quality and computational efficiency of existing methods across all scenes

方法	PSNR/dB	SSIM	LPIPS	FPS/(帧/s)	训练时间/min
SeaThru-NerF(Levy等,2023)	27.394	0.860	0.215	0.55	159
3DGS(Kerbl等,2023)	26.188	0.859	0.238	<u>149.36</u>	17
RecGS(Zhang等,2025c)	25.829	0.857	0.233	<i>146.62</i>	38
SeaSplat(Yang等,2025a)	27.385	<i>0.866</i>	<u>0.194</u>	42.69	85
Water-Splatting(Li等,2025)	27.573	0.865	<i>0.198</i>	35.80	29
3D-UIR(Yuan等,2025)	<u>28.116</u>	0.876	0.202	48.72	48
Plenodium(Wu等,2025)	28.585	<u>0.873</u>	0.191	240.75	<u>24</u>

注:加粗、下划线、斜体字体分别表示各列最优、次优、第三优结果。各指标结果为在多个数据集所有场景上的平均值。

分钟即可完成训练,而水下专用方法在引入物理建模后训练时间有所增加,但仍在实际可接受的范围内。相比之下,SeaThru-NeRF需要耗时数小时的较长训练时间,这一结果反映了不同技术路线在技术成熟度和部署成本之间的显著差异,为实际应用中的技术选型和资源配置提供了重要参考。

为了深入理解水下新视角合成领域的技术现状和发展瓶颈,本研究通过定性分析对主流技术路线在真实水下场景中的表现进行了系统性评估。分析内容涵盖了不同技术路线的新视角渲染效果、水下场景复原能力以及几何重建精度,旨在为该领域的未来发展提供有益的技术洞察。

图15展示了水下场景新视角合成与场景复原方法对比,各子图中,上面一行是新视角合成图像,下面一行是场景复原图像和深度图,“-”表示该方法不支持场景复原。通过对主流技术路线的定性分析,可以明确识别出水下光学介质的复杂性对当前新视角合成技术所带来的根本性挑战,具体表现为以下3个关键技术难点:

1)水体散射伪影与视角一致性问题。由于水体介质中的多重散射效应,合成图像中普遍存在显著的水云状伪影现象,该现象在场景背景区域和半透明物体边界处尤为突出。这一问题的根本原因在于传统方法缺乏对物体本征反射特性与水介质引起的视角依赖性光学效应(包括前向散射衰减和后向散射增强)的有效解耦机制。

2)几何重建精度不足。深度估计结果显示出严重的几何塌陷现象,特别是在物体边缘区域和远距离场景中表现出明显的不连续性和模糊性。这表明

现有方法在水下复杂光学条件下的三维几何重建能力存在根本性局限,难以准确区分水介质的光学边界与实际物体表面的几何边界。

3)图像细节保真度降低。合成视角图像普遍存在细节信息丢失和对比度显著下降的问题,这直接影响了新视角合成结果的视觉真实感和信息完整性,限制了其在实际应用中的有效性。

通过深入的对比分析,各方法在应对上述技术挑战时展现出显著的差异化特征。3DGS作为经典的高斯泼溅方法在陆地场景中表现优异,但在水下环境中暴露出明显的局限性。其渲染结果普遍存在严重的水云伪影和几何塌陷现象,这主要源于其缺乏对水体介质光学特性的物理建模能力。从深度图可以看出,3DGS方法在水下场景的几何重建方面存在显著的不连续性,特别是在远距离区域和物体边缘处表现出明显的塌陷问题。SeaThru-NeRF方法在传统NeRF基础上引入了水下光学模型,在一定程度上改善了水体散射伪影问题。然而,从实验结果可以观察到,在复杂几何场景的细节保真度方面仍有不足,特别是在处理多重散射效应时容易产生过度平滑的问题。其深度估计虽然相比3DGS有所改善,但在精细结构的重建方面仍存在模糊现象。Water-Splatting方法凭借其基于物理光学模型的外观-介质解耦策略,在抑制水云伪影方面表现出色。实验结果表明,其能够精确建模光线在水下介质中的传播过程,包括吸收衰减和散射增强效应,从而实现对物体本征外观特性的有效分离。在几何结构复杂的水下场景中,该方法能够保持更高的细节保真度,水下结构的完整性和清晰度得到显著提升。

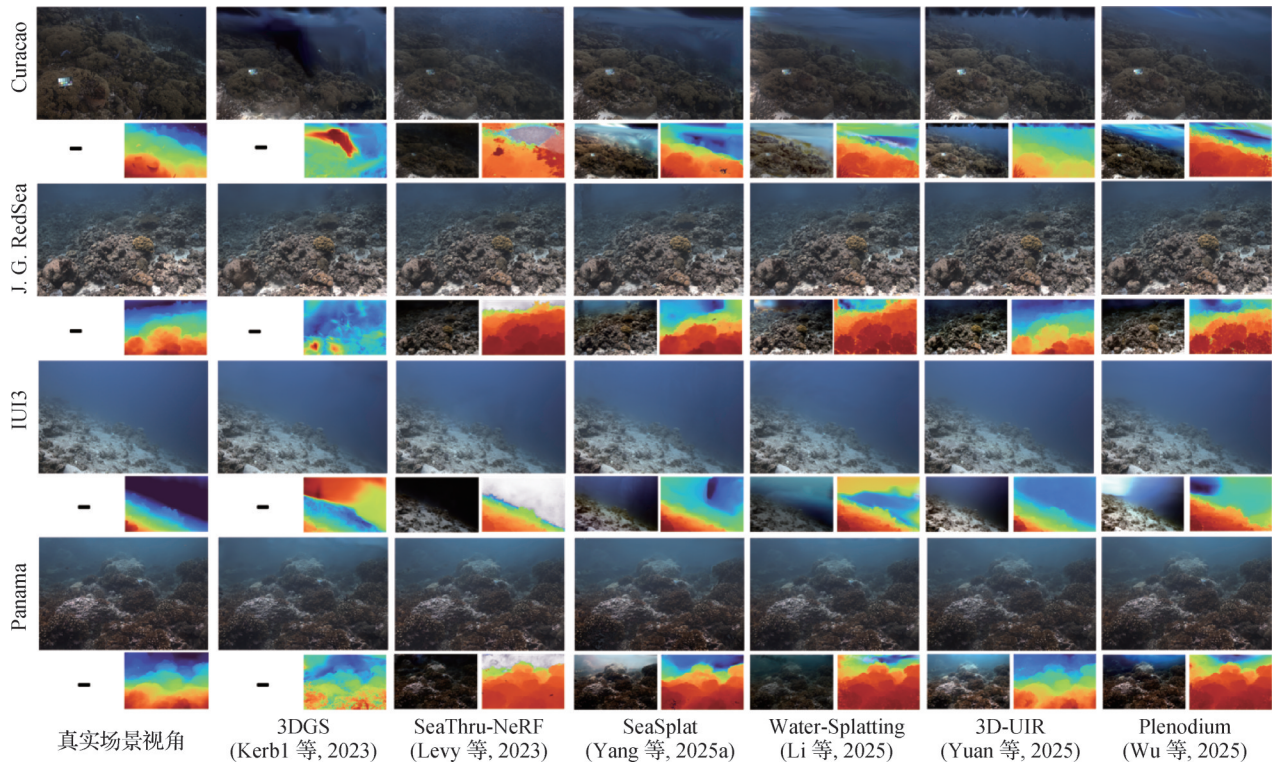


图15 水下场景新视角合成与场景复原方法对比

Fig. 15 Comparison of underwater novel view synthesis and scene restoration methods

Plenodium 在各个测试场景中实现了高质量的新视角合成,水云伪影得到有效抑制,场景细节保持清晰。然而由于其深度估计结果的几何连续性和准确性不佳,在面对复杂水下结构的场景复原结果不稳定。3D-UIR 方法在跨视角渲染一致性维护方面展现出卓越性能。从实验结果可以看出,该方法生成的深度图具有更高的空间连续性和几何准确性,有效解决了传统方法中普遍存在的深度塌陷和浮影伪影问题。这种改进不仅显著提升了新视角合成的视觉质量,更重要的是增强了重建结果的物理合理性,为后续的三维场景理解和分析提供了更可靠的几何基础。深度与新视角图像的强耦合关系使得其图像细节保真度也得到进一步提升。

综合分析结果表明,水下新视角合成正处于由通用方法向专用化方法转型的关键阶段,其中基于显式表示的技术在复杂光学环境下已展现出显著优势。然而,不同技术路线在具体问题的解决上仍各具特色,这不仅体现了方法发展的多样性,也预示着未来研究将在跨方法融合、光学建模精度与计算效率提升以及高质量与实时性能兼顾等方面孕育新的突破。

4 挑战与展望

尽管近年来水下新视角合成技术已有诸多探索,并取得了一定的发展与进步,但该领域仍面临诸多亟待解决的挑战,距离实际应用仍存在差距。结合当前研究现状,接下来将归纳主要挑战并展望未来发展趋势。

4.1 主要技术挑战

1) 场景动态性与时空一致性建模。水下环境的固有动态特性对新视角合成技术构成了根本性挑战。首先,海洋流体的连续运动导致场景几何在多视角数据采集过程中发生持续变化。海流、涌浪等水文现象使得同一物理结构在不同采集时刻呈现差异化的空间配置,破坏了传统多视图几何重建所依赖的静态场景假设。其次,海洋生物的随机运动进一步加剧了场景的时空不一致性。鱼群游动、海底植物摆动等生物活动在多视角图像中表现为非刚性变形和瞬态遮挡,现有方法往往将这些动态目标误识别为静态场景元素,导致重建结果出现严重的几何伪影和外观不一致。此外,悬浮颗粒的布朗运动在微观尺度上造成光学散射特性的随机波动,增加

了介质建模的复杂性。这些多尺度、多时序的动态因素相互耦合,使得传统基于静态假设的重建算法在水下环境中面临系统性失效的风险。

2)海洋环境光学参数的空间异质性。海洋光学环境呈现显著的空间异质性和时变特征,现有方法普遍假设的均匀介质模型与实际海洋环境存在本质差异。在空间维度上,不同海域的水体光学特性受到地理位置、水深分布、悬浮物浓度以及浮游生物密度等多重因素影响,呈现复杂的空间变化模式。深海与近岸、热带与极地、清水与混浊水体在光谱衰减系数、散射相函数等关键光学参数上存在数量级差异。在时间维度上,潮汐变化、季节更替以及天气条件等环境因子导致同一海域的光学特性呈现周期性和随机性波动。现有方法通常基于特定数据集训练得到的先验知识,在面对新的海洋环境时缺乏有效的跨域适应能力,模型泛化性能严重受限。这种环境光学参数的高度变异性要求新视角合成算法具备强大的自适应建模能力,能够在缺乏先验知识的条件下快速学习新环境的光学特征。

3)高质量多视角数据获取困难。水下多视角数据的获取面临技术成本高、采集难度大和质量控制难等多重制约。在技术层面,水下摄影设备需要具备高压防护、精确导航和稳定控制等复杂功能,设备成本和维护成本远高于陆地摄影系统。水下能见度有限、操作空间受限等因素进一步增加了精确控制相机轨迹的难度,难以获得满足重建算法要求的高重叠度、多角度覆盖的理想采集轨迹。在数据质量层面,现有公开数据集多来源于互联网图像收集或小规模实地采集,存在采集轨迹不规范、图像质量参差不齐和场景多样性不足等问题。与陆地场景拥有大规模高质量数据集形成鲜明对比,水下场景缺乏标准化的基准数据集,制约了算法性能的客观评估和公平比较。此外,合成数据与真实场景的域间差异问题尚未得到有效解决,基于简化物理模型生成的合成数据往往无法完全复现真实水下环境的复杂光学现象,导致在合成数据上训练的模型在真实场景中性能显著下降。

4.2 未来研究方向

1)复杂真实场景的时空动态建模。面向实际海洋应用的复杂野外场景建模将成为未来发展的重要方向。现有方法多基于相对理想的仿真场景或简化先验假设下进行验证,而真实海洋环境中复杂的时

空动态变化要求算法具备更强的鲁棒性。未来研究需要重点解决海洋流体运动的物理约束建模问题,确保各类三维场景表示方法在时间演化过程中遵循流体力学规律,实现渲染过程与物理建模的统一。同时,需要发展自适应动态场景分解技术,将静态结构、刚体运动和非刚性变形进行有效解耦,为复杂动态水下场景的精确重建奠定基础。

2)三维几何大模型的场景泛化。视觉基础模型为水下新视角合成提供了强大的跨模态先验,尤其是三维几何大模型。这类模型通过在海量三维几何数据上进行预训练,能够学习到丰富的几何先验和场景结构知识。未来研究将探索如何将这些通用的几何理解能力迁移到水下特定场景中,可行的技术路径包括在 DUS_t3R 和 MAS_t3R 等预训练模型基础上构建水下场景特定的适配层,通过少样本微调实现跨域迁移,或设计教师—学生框架将陆地场景预训练模型的几何理解能力蒸馏至水下专用网络。通过利用潜在空间的几何表示学习和扩散模型可控的三维生成机制,实现从稀疏观测到完整三维场景的高质量重建。关键技术挑战包括水下几何特征的有效编码、跨域几何知识的迁移策略,以及多尺度几何细节的层次化生成。

3)物理与语义协同的空间场景感知。未来的水下新视角合成技术需要从单纯的几何重建向语义感知的三维场景理解方向发展。精度的三维重建不仅要实现视觉上的逼真渲染,更要为下游的三维视觉感知任务提供有效支撑,包括海洋生物识别、地质结构分割和人工目标检测等应用。这要求发展语义嵌入的三维表示方法,例如为每个高斯基元附加可学习的语义特征向量以实现几何渲染与语义分割的联合输出,或结合 CLIP 和 SAM 等视觉语言模型构建支持任意文本查询的水下三维语义场,将语义分割、实例分割和目标检测等高级视觉理解能力直接集成到三维重建过程中。通过语义指导的几何约束和内容感知的渲染优化,实现从底层光学建模到高层语义理解的端到端统一,为智能海洋感知系统提供更加丰富和实用的三维场景描述。

4)轻量化表征建模与实时渲染。面向实际部署需求,场景建模的轻量化和三维模型压缩将成为关键技术方向。这包括发展紧凑的场景表示方法、设计高效的网络架构,以及探索渐进式渲染策略,在保证视觉质量的前提下显著降低计算和存储开销。特

别是针对水下机器人等边缘设备的有限计算资源,需要开发自适应复杂度控制算法,根据硬件能力和任务需求动态调整模型精度与计算负载的平衡。可通过基于渲染贡献度的重要性评估实现高斯基元剪枝与低比特量化,采用层级细节渲染策略在近景区域保持高保真渲染而远景区域使用简化表示,并针对水下机器人常用的嵌入式GPU设计专用算子加速方案。通过模型压缩、知识蒸馏等技术手段,实现高质量水下重建算法的设备端部署。

5) 标准化评估体系与基准数据集建设。针对水下新视角合成任务的特殊性质,现有的传统图像质量指标在新视角合成任务中存在明显局限性,往往无法准确反映人眼感知质量和几何重建精度。未来需要发展针对新视角合成特点的专用评估指标,构建融合几何精度、视觉感知和时序稳定性的多维度质量评价体系,综合考虑视角一致性、几何精度和时间稳定性等质量因素。同时,需建设多模态标准化基准数据集,包含真实采集的多视角水下图像、精确的相机参数标定、可靠的深度与语义标注,以及不同水质条件下的场景覆盖,为各类技术方法提供统一的比较基准。此外,需要构建面向应用的主观评估体系,结合海洋科学家、水下摄影师等专业用户的实际需求,建立更贴近应用场景的质量评估标准。

5 结 语

本文从水下新视角合成技术的发展现状出发,系统分析了水下成像的物理退化机制及其对三维重建的影响,阐述了神经辐射场和高斯泼溅等核心技术原理。从隐式神经表征、显式几何建模和前馈网络等角度全面梳理了现有代表性方法的技术路线和性能特点,深入分析了各类方法在处理水下光学效应和复杂场景方面的优势与局限。综合全文分析,当前该领域呈现出三大核心演进趋势:显式高斯表示正在取代隐式神经场成为主导范式,在渲染质量与效率上已全面超越早期NeRF方法。物理先验与学习能力的协同建模成为提升重建质量的关键,混合表征策略有效弥合了几何建模与介质传输之间的鸿沟。前馈式三维大模型为解决水下场景泛化难题提供了新思路,有望推动技术从实验室验证走向工程化部署。

当前水下新视角合成技术虽然已初步验证了在

受控环境下的可行性,但距离实际海洋应用仍存在显著差距。现有方法多局限于静态或准静态场景,对海洋生物游动、悬浮颗粒运动等动态因素的处理能力不足;大多数算法依赖于特定水体条件下的训练数据,跨海域泛化性能有待提升;公开数据集规模小、场景类型单一,难以支撑大规模模型训练。本文通过构建系统化的技术图谱和问题分析框架,旨在为相关研究者明确技术现状与发展瓶颈,推动该领域从概念验证向工程应用的转化,并为海洋机器人导航、水下考古数字化重建等实际应用场景提供技术参考。

参考文献 (References)

- Akkaynak D and Treibitz T. 2019. Sea-thru: a method for removing water from underwater images//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 1682-1691 [DOI: 10.1109/CVPR.2019.00178]
- Chen A P, Xu Z X, Zhao F Q, Zhang X S, Xiang F B, Yu J Y, et al. 2021. MVSNeRF: fast generalizable radiance field reconstruction from multi-view stereo//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, Canada: IEEE: 14104-14113 [DOI: 10.1109/ICCV48922.2021.01386]
- Chen L F, Xiong Y C, Zhang Y J, Yu R Y, Fang L and Liu D F. 2024a. SP-SeaNeRF: underwater neural radiance fields with strong scattering perception. *Computers and Graphics*, 123: #104025 [DOI: 10.1016/j.cag.2024.104025]
- Chen S, Zhou J L and Li L. 2024b. Optimizing 3D Gaussian splatting for sparse viewpoint scene reconstruction [EB/OL]. [2025-09-26]. <https://arxiv.org/pdf/2409.03213.pdf>
- Chen W T, Wang Y F, Kuo S Y and Wetzstein G. 2024c. DehazeNeRF: multi-image haze removal and 3D shape reconstruction using neural radiance fields//Proceedings of 2024 International Conference on 3D Vision (3DV). Davos, Switzerland: IEEE: 247-256 [DOI: 10.1109/3DV62453.2024.00039]
- Chen Z Y, Tang J Y, Wang G, Li S Q, Li X H, Ji X Y, et al. 2024d. UW-SDF: exploiting hybrid geometric priors for neural SDF reconstruction from underwater multi-view monocular images//Proceedings of 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Abu Dhabi, United Arab Emirates: IEEE: 14248-14255 [DOI: 10.1109/IROS58592.2024.10802499]
- Cong R M, Zhang Y M, Zhang C, Li C Y and Zhao Y. 2020. Research progress of deep learning driven underwater image enhancement and restoration. *Journal of Signal Processing*, 36(9): 1377-1389 (丛润民, 张禹墨, 张晨, 李重仪, 赵耀. 2020. 深度学习驱动的水下图像增强与复原研究进展. *信号处理*, 36(9): 1377-1389) [DOI: 10.16798/j.issn.1003-0530.2020.09.001]

- Dahmani H, Bennehar M, Piasco N, Roldão L and Tsishkou D. 2024. SWAG: splatting in the wild images with appearance-conditioned Gaussians//Proceedings of the 18th European Conference on Computer Vision. Milan, Italy: Springer: 325-340 [DOI: 10.1007/978-3-031-73116-7_19]
- Gough L, Azzarelli A, Zhang F and Anantrasirichai N. 2025. AquaNeRF: neural radiance fields in underwater media with distractor removal [EB/OL]. [2025-09-04]. <https://arxiv.org/pdf/2502.16351.pdf>
- Guo J C, Li C Y, Guo C L and Chen S J. 2017. Research progress of underwater image enhancement and restoration methods. *Journal of Image and Graphics*, 22(3): 273-287 (郭继昌, 李重仪, 郭春乐, 陈善继. 2017. 水下图像增强和复原方法研究进展. *中国图象图形学报*, 22(3): 273-287 [DOI: 10.11834/jig.20170301])
- Gupta V, Manoj S, Mukund Varma T and Mitra K. 2024. U2NeRF: unsupervised underwater image restoration and neural radiance fields [EB/OL]. [2025-09-26]. <https://arxiv.org/pdf/2411.16172.pdf>
- Hong Y C, Zhang K, Gu J X, Bi S, Zhou Y, Liu D F, et al. 2024. LRM: large reconstruction model for single image to 3D//Proceedings of the 12th International Conference on Learning Representations. Vienna, Austria: ICLR: 1-25
- Hu S Y and Liu Q. 2025. Fast underwater scene reconstruction using multi-view stereo and physical imaging. *Neural Networks*, 189: #107568 [DOI: 10.1016/j.neunet.2025.107568]
- Kerbl B, Kopanas G, Leimkuehler T and Drettakis G. 2023. 3D Gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4): #139 [DOI: 10.1145/3592433]
- Kirillov A, Mintun E, Ravi N, Mao H Z, Rolland C, Gustafson L, et al. 2023. Segment anything//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France: IEEE: 3992-4003 [DOI: 10.1109/ICCV51070.2023.00371]
- Kulhanek J, Peng S Y, Kukulova Z, Pollefeys M and Sattler T. 2024. WildGaussians: 3D Gaussian splatting in the wild//Proceedings of the 38th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: #670 [DOI: 10.5555/3737916.3738586]
- Leroy V, Cabon Y and Revaud J. 2024. Grounding image matching in 3D with MAST3R//Proceedings of the 18th European Conference on Computer Vision. Milan, Italy: Springer: 71-91 [DOI: 10.1007/978-3-031-73220-1_5]
- Levy D, Peleg A, Pearl N, Rosenbaum D, Akkaynak D, Korman S, et al. 2023. SeaThru-NeRF: neural radiance fields in scattering media//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 56-65 [DOI: 10.1109/CVPR52729.2023.00014]
- Li C Y, Anwar S and Porikli F. 2020. Underwater scene prior inspired deep underwater image and video enhancement. *Pattern Recognition*, 98: #107038 [DOI: 10.1016/j.patcog.2019.107038]
- Li H P, Song W X, Xu T A, Elsig A and Kulhanek J. 2025. WaterSplatting: fast underwater 3D scene reconstruction using Gaussian splatting//Proceedings of 2025 International Conference on 3D Vision. Singapore, Singapore: IEEE: 969-978 [DOI: 10.1109/3DV66043.2025.00094]
- Li X L. 2024. Water-related optics. *SCIENTIA SINICA Informationis*, 54(2): 227-280 (李学龙. 2024. 涉水光学. *中国科学: 信息科学*), 54(2): 227-280 [DOI: 10.1360/SSI-2022-0398])
- Lian S J, Zhang Z Y, Yang L T, Ren M Y, Liu D B and Li H. 2025. TUGS: physics-based compact representation of underwater scenes by Tensorized Gaussian [EB/OL]. [2025-09-26]. <https://arxiv.org/pdf/2505.08811.pdf>
- Lin H K, Liang D K, Qi Z H and Bai X. 2025. A unified image-dense annotation generation model for underwater scenes//Proceedings of 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 961-970 [DOI: 10.1109/CVPR52734.2025.00098]
- Liu S H, Gao N, Fu S W, Zhong X Q and Li H J. 2025. SeaFree-GS: reconstructing underwater 3D scenes with true appearances. *IEEE Signal Processing Letters*, 32: 2114-2118 [DOI: 10.1109/LSP.2025.3567853]
- Martin-Brualla R, Radwan N, Sajjadi M S M, Barron J T, Dosovitskiy A and Duckworth D. 2021. NeRF in the wild: neural radiance fields for unconstrained photo collections//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 7206-7215 [DOI: 10.1109/CVPR46437.2021.00713]
- Mildenhall B, Srinivasan P P, Tancik M, Barron J T, Ramamoorthi R and Ng R. 2021. NeRF: representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1): 99-106 [DOI: 10.1145/3503250]
- Nie J, Zuo Z J, Huang L, Wang Z G, Sun Z Y, Zhong G Q, et al. 2022. Marine oriented multimodal intelligent computing: challenges, progress and prospects. *Journal of Image and Graphics*, 27(9): 2589-2610 (聂婕, 左子杰, 黄磊, 王志刚, 孙正雅, 仲国强, 等. 2022. 面向海洋的多模态智能计算: 挑战、进展和展望. *中国图象图形学报*), 27(9): 2589-2610 [DOI: 10.11834/jig.211267])
- Oechsle M, Peng S Y and Geiger A. 2021. UNISURF: unifying neural implicit surfaces and radiance fields for multi-view reconstruction//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, Canada: IEEE: 5569-5579 [DOI: 10.1109/ICCV48922.2021.00554]
- Panetta K, Gao C and Agaian S. 2016. Human-visual-system-inspired underwater image quality measures. *IEEE Journal of Oceanic Engineering*, 41(3): 541-551 [DOI: 10.1109/JOE.2015.2469915]
- Pumarola A, Corona E, Pons-Moll G and Moreno-Noguer F. 2021. D-NeRF: neural radiance fields for dynamic scenes//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Rec-

- ognition (CVPR). Nashville, USA: IEEE: 10313-10322 [DOI: 10.1109/CVPR46437.2021.01018]
- Qiao Y J, Shao M W, Meng L Z and Xu K. 2025. RestorGS: depth-aware Gaussian splatting for efficient 3D scene restoration//Proceedings of 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 11177-11186 [DOI: 10.1109/CVPR52734.2025.01044]
- Ramazina A, Bijelic M, Walz S, Sanvito A, Scheuble D and Heide F. 2023. ScatterNeRF: seeing through fog with physically-based inverse neural rendering//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France: IEEE: 17911-17922 [DOI: 10.1109/ICCV51070.2023.01646]
- Randall Y and Treibitz T. 2023. FLSea: underwater visual-inertial and stereo-vision forward-looking datasets [EB/OL]. [2025-09-26]. <https://arxiv.org/pdf/2302.12772.pdf>
- Rombach R, Blattmann A, Lorenz D, Esser P and Ommer B. 2022. High-resolution image synthesis with latent diffusion models//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, USA: IEEE: 10674-10685 [DOI: 10.1109/CVPR52688.2022.01042]
- Schönberger J L and Frahm J M. 2016. Structure-from-motion revisited//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE: 4104-4113 [DOI: 10.1109/CVPR.2016.445]
- Sethuraman A V, Ramanagopal M S and Skinner K A. 2023. Water-NeRF: neural radiance fields for underwater scenes//Proceedings of the OCEANS 2023 — MTS/IEEE U.S. Gulf Coast. Biloxi, USA: IEEE: 1-7 [DOI: 10.23919/OCEANS52994.2023.10336972]
- Tang Y K, Zhu C X, Wan R J, Xu C and Shi B X. 2024. Neural underwater scene representation//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 11780-11789 [DOI: 10.1109/CVPR52733.2024.01119]
- Treibitz T, Schechner Y, Kunz C and Singh H. 2012. Flat refractive geometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(1): 51-65 [DOI: 10.1109/TPAMI.2011.105]
- Varma T M, Wang P, Chen X, Chen T, Venugopalan S and Wang Z Y. 2023. Is attention all that NeRF needs?//Proceedings of the 11th International Conference on Learning Representations. Kigali, Rwanda: ICLR: 1-22
- Wang H R, Anantrasirichai N, Zhang F and Bull D. 2025a. UW-GS: distractor-aware 3D Gaussian splatting for enhanced underwater scene reconstruction//Proceedings of 2025 IEEE/CVF Winter Conference on Applications of Computer Vision. Tucson, USA: IEEE: 3280-3289 [DOI: 10.1109/WACV61041.2025.00324]
- Wang J Y, Chen M H, Karaev N, Vedaldi A, Rupprecht C and Novotny D. 2025b. VGGT: visual geometry grounded transformer//Proceedings of 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 5294-5306 [DOI: 10.1109/CVPR52734.2025.00499]
- Wang P, Liu L J, Liu Y, Theobalt C, Komura T and Wang W P. 2021a. NeuS: learning neural implicit surfaces by volume rendering for multi-view reconstruction//Proceedings of the 35th International Conference on Neural Information Processing Systems. [s.l.]: Curran Associates Inc.: #2081 [DOI: 10.5555/3540261.3542342]
- Wang Q Q, Wang Z C, Genova K, Srinivasan P, Zhou H, Barron J T, et al. 2021b. IBRNet: learning multi-view image-based rendering//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 4688-4697 [DOI: 10.1109/CVPR46437.2021.00466]
- Wang S Z, Leroy V, Cabon Y, Chidlovskii B and Revaud J. 2024. DUS3R: geometric 3D vision made easy//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 20697-20709 [DOI: 10.1109/CVPR52733.2024.01956]
- Wu C G, Dong J X, Li C J and Tang J H. 2025. Plenodium: UnderWater 3D scene reconstruction with plenoptic medium representation [EB/OL]. [2025-09-26]. <https://arxiv.org/pdf/2505.21258/pdf>
- Wu L X, Chen Z H, Lin X P and Liu Y Z. 2020. Building the integrated observational network of "Transparent Ocean". *Chinese Science Bulletin*, 65(25): 2654-2661 (吴立新, 陈朝晖, 林霄沛, 刘永正. 2020. "透明海洋"立体观测网构建. *科学通报*, 65(25): 2654-2661) [DOI: 10.1360/TB-2020-0558]
- Xiong J H and Heidrich W. 2021. In-the-wild single camera 3D reconstruction through moving water surfaces//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, Canada: IEEE: 12538-12547 [DOI: 10.1109/ICCV48922.2021.01233]
- Yang D, Leonard J J and Girdhar Y. 2025a. SeaSplat: representing underwater scenes with 3D Gaussian splatting and a physically grounded image formation model//Proceedings of 2025 IEEE International Conference on Robotics and Automation. Atlanta, USA: IEEE: 7632-7638 [DOI: 10.1109/ICRA55743.2025.11128502]
- Yang J N, Sax A, Liang K J, Henaff M, Tang H, Cao A, et al. 2025b. Fast3R: towards 3D reconstruction of 1000+ images in one forward pass//Proceedings of 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 21924-21935 [DOI: 10.1109/CVPR52734.2025.02042]
- Yang M and Sowmya A. 2015. An underwater color image quality evaluation metric. *IEEE Transactions on Image Processing*, 24(12): 6062-6071 [DOI: 10.1109/TIP.2015.2491020]
- Yariv L, Gu J T, Kasten Y and Lipman Y. 2021. Volume rendering of neural implicit surfaces//Proceedings of the 35th International Conference on Neural Information Processing Systems. [s.l.]: Curran Associates Inc.: #367 [DOI: 10.5555/3540261.3540628]
- Yuan J Y, Li Y J, Zhang Y L, Guo C L, Tang X X, Wang R X, et al. 2025. 3D-UIR: 3D Gaussian for underwater 3D scene reconstruction via physics based appearance-medium decoupling [EB/OL]. [2025-09-26]. <https://arxiv.org/pdf/2505.21238.pdf>

- Zhang D B, Wang C M, Wang W T, Li P H, Qin M H and Wang H Q. 2024a. Gaussian in the wild: 3D Gaussian splatting for unconstrained image collections//Proceedings of the 18th European Conference on Computer Vision. Milan, Italy: Springer: 341-359 [DOI: 10.1007/978-3-031-73116-7_20]
- Zhang F, You S D, Li Y and Fu Y. 2024b. Atlantis: enabling underwater depth estimation with stable diffusion//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 11852-11861 [DOI: 10.1109/CVPR52733.2024.01126]
- Zhang S Z, Wang J Y, Xu Y H, Xue N, Rupprecht C, Zhou X W, et al. 2025a. FLARE: feed-forward geometry, appearance and camera estimation from uncalibrated sparse views//Proceedings of 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 21936-21947 [DOI: 10.1109/CVPR52734.2025.02043]
- Zhang T Y and Johnson-Roberson M. 2023. Beyond NeRF underwater: learning neural reflectance fields for true color correction of marine imagery. *IEEE Robotics and Automation Letters*, 8(10): 6467-6474 [DOI: 10.1109/LRA.2023.3307287]
- Zhang T Y, Zhi W M, Mangelson J and Johnson-Roberson M. 2025b. Infinite leagues under the sea: photorealistic 3D underwater terrain generation by latent fractal diffusion models [EB/OL]. [2025-09-26]. <https://arxiv.org/pdf/2503.06784.pdf>
- Zhang T Y, Zhi W M, Meyers B, Durrant N, Huang K N, Mangelson J, et al. 2025c. RecGS: removing water caustic with recurrent Gaussian splatting. *IEEE Robotics and Automation Letters*, 10(1): 668-675 [DOI: 10.1109/LRA.2024.3511418]
- Zhou J C, Liang T Y, Zhang D H, Liu S Y, Wang J S and Wu E Q. 2025. WaterHE-NeRF: water-ray matching neural radiance fields for underwater scene reconstruction. *Information Fusion*, 115: #102770 [DOI: 10.1016/j.inffus.2024.102770]
- Zhou L, Liu Q M, Jin K J, Zhao W Y and Zhang W D. 2025. Research progress of underwater image restoration and enhancement methods. *Journal of Image and Graphics*, 30(1): 51-65 (周玲, 刘庆敏, 金凯杰, 赵文义, 张卫东. 2025. 水下图像复原和增强方法研究进展. *中国图象图形学报*, 30(1): 51-65) [DOI: 10.11834/jig.240050]

作者简介

袁捷禹,男,博士后,主要研究方向为图像处理和三维计算机视觉。E-mail: jiejuyuan.cn@nankai.edu.cn

李重仪,通信作者,男,教授,博士生导师,主要研究方向为计算成像和底层视觉任务。E-mail: lichongyi@nankai.edu.cn

赵倩倩,女,博士研究生,主要研究方向为图像增强与复原。E-mail: qqzhao@mail.nankai.edu.cn

李纤君,男,硕士研究生,主要研究方向为水下三维重建与新视角合成。E-mail: 2111903@mail.nankai.edu.cn

张元林,男,硕士研究生,主要研究方向为水下图像增强与复原。E-mail: zason_zyl@163.com

郭春乐,男,副教授,博士生导师,主要研究方向为计算成像与生成式人工智能。E-mail: guochunle@nankai.edu.cn