

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(XXXX)XX-0001-29

论文引用格式: Zhang Qi, Wu Zi'ang, Xue Wanli, Zhu Qi, Tian Chunwei, Zuo Wangmeng. Generative network-based underwater image enhancement: a survey[J/OL]. Journal of Image and Graphics, XXXX: 1-29. DOI: 10.11834/jig.260516. (张琦, 吴子昂, 薛万利, 朱旗, 田春伟, 左旺孟. 生成式网络的水下图像增强方法综述[J/OL]. 中国图象图形学报, XXXX: 1-29. DOI: 10.11834/jig.260516.) [DOI: 10.11834/jig.260516]

生成式网络的水下图像增强方法综述

张琦¹, 吴子昂², 薛万利³, 朱旗⁴, 田春伟², 左旺孟²

1. 哈尔滨工业大学(威海)经济管理学院, 威海 264209; 2. 哈尔滨工业大学计算学部, 哈尔滨 150006; 3. 天津理工大学计算机科学与工程学院, 天津 300384; 4. 南京航空航天大学人工智能学院, 南京 211106

摘要: 水下图像增强(underwater image enhancement, UIE)旨在从受吸收、散射、悬浮颗粒及非均匀照明影响的退化图像中恢复清晰、自然的场景图像, 以满足下游视觉任务需求。近年来, 生成式网络已成为 UIE 研究的重要范式, 但现有方法在生成主干、监督方式、物理约束、推理效率及评测口径上差异显著, 造成各方法间缺乏可比性。本文围绕基于生成式网络的 UIE 方法开展系统性综述: 从水下成像退化机理出发, 分析传统物理先验、数据驱动学习与生成式建模之间的演进关系; 将现有方法归纳为生成对抗网络(generative adversarial network, GAN)、Transformer 与 Diffusion 三类范式, 并从监督方式、生成机制、物理/语义约束及部署成本等维度进行跨方法对比; 结合水下图像增强基准数据集(underwater image enhancement benchmark dataset, UIEB)、U45 与 TEST-U90 等数据集的定量结果, 探讨峰值信噪比(peak signal-to-noise ratio, PSNR)、结构相似性指数(structural similarity index measure, SSIM)、水下彩色图像质量评价(underwater color image quality evaluation, UCIQE)、水下图像质量评价(underwater image quality measure, UIQM)等指标之间的分离现象及其对方法选择的影响。实验结果表明, GAN、Transformer 与 Diffusion 分别适用于实时增强、全局一致性建模与高保真恢复等不同需求。未来 UIE 研究应进一步关注复杂退化耦合、跨域泛化、评价协议标准化、轻量化部署及可信可控生成等关键问题。本文可为生成式 UIE 方法的分类理解、实验设计与工程部署提供理论依据与实践指导。

关键词: 水下图像增强; 生成式网络; 生成对抗网络; Transformer; 扩散模型

Generative network-based underwater image enhancement: a survey

Zhang Qi¹, Wu Zi'ang², Xue Wanli³, Zhu Qi⁴, Tian Chunwei², Zuo Wangmeng²

1. School of Economics and Management, Harbin Institute of Technology (Weihai), Weihai 264209, China; 2. Faculty of Computing, Harbin Institute of Technology, Harbin 150006, China; 3. School of Computer Science and Engineering, Tianjin University of Technology, Tianjin 300384, China; 4. College of Artificial Intelligence, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

Abstract: Underwater image enhancement (UIE) aims to recover clear scenes from degraded underwater images affected by selective absorption, scattering, suspended particles, and non-uniform illumination. Because these degradations are coupled and depend on water type, imaging distance, and lighting, UIE is more difficult than ordinary image enhancement

收稿日期: 2026-08-16; 修回日期: 2026-09-17

基金项目: 国家自然科学基金(项目编号: 62576123); 黑龙江省自然科学基金青年项目(B类)(项目编号: YQ2025F003); 山东省青年科技人才托举工程(项目编号: SDAST2026QTB050); 哈尔滨工业大学青年科学家工作室项目; 中央高校基本科研业务费专项资金项目(项目编号: AUGA5710012325)

Supported by: National Natural Science Foundation of China (62576123); Natural Science Foundation of Heilongjiang Province (YQ2025F003); Young Talent of Lifting engineering for Science and Technology in Shandong, China (SDAST2026QTB050); Youth Scientist Workshop of HIT; Fundamental Research Funds for the Central Universities (AUGA5710012325)

and directly affects downstream tasks such as detection, segmentation, navigation, and three-dimensional (3D) reconstruction. Early physics-based and prior-based methods were interpretable but unstable, while deep models that optimize pixel losses often produced results that were numerically acceptable but inconsistent with clear underwater scenes. This review focuses on generative-network-based UIE and compares how different generative paradigms balance restoration fidelity, global consistency, controllability, and deployment cost. We build a unified taxonomy with three families: generative adversarial network (GAN), Transformer, and Diffusion. GAN methods emphasize adversarial distribution matching, and representative methods such as underwater generative adversarial network (UWGAN) and fast underwater image enhancement-generative adversarial network (FUNIE-GAN) show how paired supervision, unpaired translation, structural constraints, physical consistency, and color regularization were introduced to address data scarcity and stabilize generation. Transformer methods replace local convolutional receptive fields with global dependency modeling; U-Shape Transformer, transmission-aware transformer for underwater image enhancement (TAFormer) and related transmission-space alternatives show how self-attention, window attention, frequency-aware priors, and state-space alternatives are used to preserve whole-image color coherence and long-range structure. Diffusion methods treat enhancement as iterative denoising; denoising diffusion probabilistic model (DDPM), denoising diffusion implicit models (DDIM), latent diffusion, and UIE-specific designs such as physical-aware diffusion (PA-Diff) show how multi-step restoration and conditional modeling can improve restoration fidelity, while skip sampling, latent-space inference, multi-scale supervision, and distillation can improve efficiency or training stability. Across these paradigms, the review also compares paired, unpaired, and semi-supervised training; one-pass and iterative generation; and physical, semantic, structural, color, and task-driven constraint. To facilitate comparison, this survey synthesizes mainstream datasets and evaluation practices. Reference-based benchmarks such as UIEB, LSUI, UFO, and paired test splits such as Test-UIEB support peak signal-to-noise ratio (PSNR), structural-similarity index measure (SSIM), and learned perceptual image patch similarity (LPIPS); real-scene and no-reference datasets such as U45, and SQUID expose behavior under domain shift and allow underwater color image quality evaluation (UCIQE), underwater image quality measure (UIQM), natural image quality evaluator (NIQE), and blind/referenceless image spatial quality evaluator (BRISQUE) to be interpreted alongside runtime, memory cost, parameters, floating-point operations (FLOPs), and downstream gains. The surveyed literature suggests that no single paradigm dominates all scenarios. GAN-based methods are more suitable when latency is critical: they are fast, lightweight, and suitable for edge platforms, but they are also sensitive to adversarial instability, loss weighting, and extreme turbidity, where hallucinated textures, oversaturation, or structural drift may appear. Transformer-based methods usually provide a better compromise between global consistency and structural preservation; they handle large-scale color cast and non-uniform illumination more reliably than purely convolutional backbones, and prior-guided variants further improve color, phase, and frequency consistency. In the reported results, efficient sparse wavelet transformer for underwater image enhancement (USWFormer) is relatively stable on Test-UIEB, Test-LSUI, and several no-reference datasets, although the quadratic cost of full attention and the memory burden of high-resolution window attention still constrain deployment. Diffusion-based methods offer a higher restoration ceiling and a natural way to add physical or semantic constraints throughout the denoising chain: in benchmarks such as UIEB-T90, methods like diffusion models for underwater image enhancement (DiffUIE) show strong reconstruction quality, but the price is inference latency, because iterative sampling can remain substantially slower than single-pass GAN or Transformer models even after skip scheduling or latent diffusion acceleration. Another recurring finding is metric disagreement: PSNR and SSIM reward pixel-aligned reconstruction, UCIQE and UIQM emphasize underwater perceptual quality, and NIQE and BRISQUE measure naturalness from another angle. The surveyed results should therefore be interpreted jointly across quality, efficiency, and downstream utility rather than through a single leaderboard. Evaluation practice is still fragmented. Paired datasets usually favor reference-based metrics, while real-scene datasets show that high scores do not always correspond to visually natural outputs or stable downstream gains. Differences in test splits, resizing rules, hardware platforms, and paired or unpaired evaluation also affect comparability. A unified reporting protocol should include full-reference and no-reference metrics, runtime and memory cost, and at least one downstream task result. The field is moving toward hybrid and multi-objective designs: a fast GAN or Transformer front end can provide coarse enhancement, and a diffusion or physically constrained refinement stage can

improve color balance, texture fidelity, and structural consistency. Future work should focus on explicit compound-degradation and uncertainty modeling, cross-domain datasets and standardized evaluation protocols, lightweight or hierarchical deployment, and safe fusion of multimodal priors such as physical parameters and task feedback without sacrificing physical plausibility. This review provides a taxonomy for understanding the field and a practical reference for choosing methods under different operating constraints.

Key words: underwater image enhancement; generative network; generative adversarial network; Transformer; diffusion model

论文引用格式: Zhang Q, Wu Z A, Xue W L, Zhu Q, Tian C W and Zuo W M. Generative network-based underwater image enhancement: a survey. *Journal of Image and Graphics* (张琦, 吴子昂, 薛万利, 朱旗, 田春伟, 左旺孟. 生成式网络的水下图像增强方法综述. 中国图象图形学报) [DOI: 10.11834/jig.260516]

0 引言

在海洋勘测、水下机器人导航、生态监测、考古调查以及国防安全等场景中,水下图像增强作为基础性任务,其质量直接影响目标检测、图像分割和三维重建等下游任务的性能(Zhou 等,2023)。相比于陆地成像,水下成像不仅受更复杂的退化因素影响,同时还对水体类型与光照条件高度敏感(Saoud 等,2024)。具体为,水体对光的选择性吸收会使红光优先衰减并引发蓝绿偏色,悬浮颗粒导致的前向散射会削弱细节清晰度,后向散射会抬升背景亮度并降低对比度,而非均匀照明又会进一步放大局部失真。这些因素相互耦合,使得水下图像增强面临有别于陆地图像增强的特有技术挑战(Zhuang 等,2025)。

水下图像增强的目标是从退化的水下图像中恢复清晰的场景图像。早期水下图像增强主要依赖物理建模。相关研究包括基于图像融合的增强方法(Ancuti 等,2012;Ancuti 等,2018)和结合去雾与直方图先验的方法(Li 等,2016)。为解决参数获取困难的问题,研究者引入对清晰图像及其相关属性的合理假设即先验约束,进而估计模型参数,降低水下图像恢复逆问题的不适定性(He 等,2011)。例如,先验通过暗通道统计特性估计透射率和背景光,从而为后续图像增强提供参数依据。此外,基于物理模型、暗通道先验和 Retinex 理论的图像去雾研究,也为散射退化图像中的透射率估计、背景光修正和

亮度恢复提供了重要参照(禹晶 等,2011;刘海波 等,2015)。相关去雾综述及改进的暗通道方法进一步表明,导向滤波、Retinex 亮度恢复和暗通道修正能够改善光晕、亮度偏暗和背景光或透射率估计不稳定等问题,但这些改进在面对不同水体类型和光照条件时仍难以保持稳定。这类方法可解释性较强,但其手工先验高度依赖特定场景假设,一旦水体类型、照明条件或悬浮颗粒分布发生变化,先验容易失效(罗虎 等,2026)。随后,一些传统机器学习方法,如支持向量机、随机森林等,尝试通过数据学习自动提取退化特征,以弥补先验假设在复杂场景下的不足,从而提升模型对多变水况的适应性;但由于高成本的特征工程和有限的表达能力,这些方法在实际应用中仍存在较大局限性。深度学习方法的引入降低了对人工特征工程的依赖(田春伟 和 左旺孟,2024),使水下图像增强能够由退化输入直接预测清晰输出,从而实现端到端学习(Li 等,2020a;Fu 和 Cao,2020;Peng 等,2023;陈龙和丁丹丹,2022)。卷积神经网络在层级特征提取、端到端重建等方面的研究(田春伟 等,2025),也为水下图像增强网络设计提供了方法支撑。然而,此类方法多以最小化像素级误差(如 L1、L2 损失)为主要目标,虽然能获得较高的像素一致性,但在整体统计特征上仍可能偏离真实清晰水下图像;感知损失相关研究也表明,单纯像素误差难以充分刻画视觉细节与高层感知差异(Johnson 等,2016),这一问题在重度退化与跨数据集场景中更加突出。

因此,近几年兴起的生成式模型将目标从最小化像素误差扩展为学习退化域与清晰域之间的概率分布映射,使增强结果在颜色分布、纹理统计与整体感知质量上更接近真实清晰图像。GAN 的理论基础来自生成器与判别器之间的对抗学习,二者在训练过程中相互博弈、动态优化,推动生成样本不断逼近真实数据分布(Goodfellow 等,2014)。这一对抗

机制在水下图像增强任务中展现出显著潜力,以 FUnIE-GAN 为代表的方法借助该框架有效提升了视觉真实感,同时兼顾了实时部署的算力需求 (Islam 等, 2020b)。Transformer 的理论根基建立于自注意力机制之上,该机制通过建模序列元素间的长距离依赖关系,实现了对全局上下文信息的深度捕获 (Vaswani 等, 2017)。在 UIE 领域, U-Shape Transformer 等方法进一步将这种全局建模能力转化为对颜色分布与结构一致性的精确约束,从而显著改善了水下图像中常见的色彩失真与细节模糊问题 (Peng 等, 2023)。扩散模型采用了渐进式生成策略,即通过前向加噪与反向去噪的迭代过程,逐步引导数据回归真实分布,确保了生成过程的稳定性 (Ho 等, 2020)。在 UIE 应用中, DM-water 等方法将这一渐进式恢复机制引入水下场景,从而同时实现了高保真恢复与高效率的运行 (Tang 等, 2023)。尽管生成式方法展现出显著潜力,但在真实应用中仍面临训练与推理成本、跨水体与跨设备泛化能力、物理一致性以及下游任务收益稳定性等挑战。

本文的主要贡献可概括如下。

1) 本文围绕生成式方法,系统梳理了水下图像增强领域的主流技术路线,重点分析以 GAN、Transformer 和 Diffusion 为代表的三类典型范式,从退化机理、生成机制、监督方式与失效模式等角度梳理了水下图像增强方法的演进脉络。

2) 本文进一步整理了水下图像增强领域的主流数据集、评价指标与实验协议,讨论了各模型、各指标之间的差异。

3) 本文总结了生成式水下图像增强面临的关键挑战与未来方向,包括复合退化建模、跨域泛化、物理约束显式建模等问题。

后文结构安排如下:第一章梳理相关工作与技术演进;第二章在统一框架下进行方法分类,对不同方法进行训练范式与生成机制层面的比较,并讨论评测目标与综合权衡;第三章根据不同方法在公开数据集上的表现进行定量分析与对比;最后给出结论并讨论后续研究方向。

1 相关工作

1.1 基于物理模型的早期方法

早期方法依赖物理建模和基于先验假设的参数

估计,通过求解逆问题从退化图像中恢复清晰场景。例如, He 等人 (2011) 提出的暗通道先验 (dark channel prior, DCP) 为图像去雾提供了理论基础,该先验通过暗通道统计特性估计透射率和背景光; Drews 等人 (2013) 提出的水下暗通道先验 (underwater dark channel prior, UDCP) 通过裁剪蓝绿通道,使先验假设更接近真实水下光谱的特点,从而改善颜色校正效果。相关改进方法可在一定程度上缓解光晕、偏暗及背景光或透射率估计不稳定等问题,但其稳定性仍受水体类型和光照条件限制 (吴迪 和 朱青松, 2015; 陈书贞 等, 2016)。总体为,这类方法可解释性较强,但其手工先验高度依赖特定场景假设,一旦水体类型、照明条件或悬浮颗粒分布发生变化,先验容易失效。

手工先验的固有局限促使研究者从数据本身寻找更具泛化能力的解决方案, Anwar 等人 (2020) 以及 Cong 等人 (2024) 的综述均指出,先验假设在复杂水下环境中存在明显局限,难以适应水体类型和光照条件的多变情形,随着深度学习基础理论框架的日趋完善,后续研究因此逐步转向数据驱动与生成式学习 (Tian 等, 2025a)。然而,人工采集高质量配对数据需要在同一场景中同步获得严格配对的退化图像与清晰参考图像,这在真实水况与动态光照条件下难以实现。为缓解配对数据不足问题, Li 等人 (2018) 提出 WaterGAN, 通过物理成像仿真将陆地图像合成为水下退化图像,并借助对抗学习提升合成样本真实感,从而为监督训练提供规模化数据。但合成退化图像与真实退化图像之间的域偏移仍难以完全消除,且由于该流程主要服务于像素误差驱动的监督训练,基于合成数据训练的模型在真实场景下仍可能产生统计偏差。这表明,无论基于物理仿真还是真实采集,配对监督信息的获取都面临难以逾越的成本与质量瓶颈。

1.2 基于 GAN 的方法

由于合成数据与真实数据之间的域偏移难以消除,真实配对采集成本高昂,研究者进一步探索无监督域翻译路径,无需配对监督即可实现退化图像到清晰图像的域映射。早期无监督方法以域翻译为基本思路,逐渐摆脱对配对数据的依赖。例如, Fabbri 等人 (2018) 提出的 UGAN 将循环一致性生成对抗网络 (cycle-consistent generative adversarial network, CycleGAN) 的无配对域翻译思想引入到水下图像增

强中,实现了从退化图像到清晰图像的直接映射。Islam 等人(2020b)提出的 FUnIE-GAN 在 UGAN 无配对映射的框架上,设计了一个轻量级全卷积网络,大幅降低了模型参数量和计算复杂度,更好地适配实时部署的场景。但基于对抗训练的无监督框架仍存在模式坍塌和梯度消失等问题,在重度退化场景下可能出现伪影或纹理幻觉;带梯度惩罚的 Wasserstein 生成对抗网络(Wasserstein generative adversarial network with gradient penalty, WGAN-GP)等稳定化训练策略通过判别器约束与梯度惩罚缓解此类不稳定性,同时为后续对抗式 UIE 方法提供了重要的训练理论基础(Gulrajani 等,2017)。尽管如此,基础 GAN 框架的固有不稳定性与生成质量的可控性不足,仍制约着增强质量的进一步提升。

为提升生成过程的稳定性与可控性,后续工作陆续引入结构、物理与颜色层面的辅助约束。Wang 等人(2019)提出的 UWGAN 引入结构保持损失,抑制几何结构漂移;Cong 等人(2023)提出的 PUGAN 引入物理一致性约束,使生成图像更符合真实水下物理成像规律;Liu 等人(2026)提出的 DA-GAN 采用多颜色空间协同损失以缓解颜色漂移问题。这些改进在一定程度上提升了 GAN 框架的图像增强质量,但无监督方法在复杂退化场景下的泛化能力仍存在明显不足。

上述辅助约束在一定程度上提升了 GAN 框架的图像增强质量,但无监督方法完全依赖未配对数据,在极端退化条件下仍暴露出明显的泛化不足,这在标注成本高、真实配对数据稀缺的场景下尤为突出。

1.3 基于 Transformer 的方法

上述基于 GAN 的方法主干网络均为卷积结构,局部感受野从根本上限制了模型对整幅图像范围内颜色与结构语义的统一建模能力。当水下图像因水体吸收、非均匀照明等因素出现跨区域色偏与全局结构失真时,卷积网络的局部视野难以捕获这种远距离依赖关系。然而,无论是无监督还是半监督方法,其主干网络仍受限于局部感受野,难以在整幅图像范围内统一颜色与结构语义。随着应用场景向跨设备、跨水型与非均匀照明条件扩展,图像中的大范围色偏和全局结构失真逐渐成为制约增强效果的瓶颈。在图像复原领域,Transformer 则凭借自注意力机制天然具备长程依赖建模能力与强大的表示能

力,为这一难题提供了新的主干方案(Vaswani 等,2017;Xiao 等,2026)。Peng 等人(2023)据此提出 U-Shape Transformer,利用其全局建模能力实现退化图像到清晰图像的一致性映射。标准 Transformer 的注意力权重由数据驱动学习得到,对水下成像过程中光传播衰减、波长选择性吸收等物理规律缺乏显式编码,因而在色偏严重或纹理退化的区域难以保持稳定的恢复方向。

为增强模型对水下退化特性的感知能力,后续工作开始尝试将外部先验知识注入 Transformer 的注意力计算过程。Li 等人(2025)提出的 TAFORMER 将传输先验引入注意力计算,用于刻画水下传播衰减差异。Peng 等人(2025)提出的 Histoformer 利用直方图统计实现色彩分布校正。Khan 等人(2025)提出的 Phaseformer 利用相位信息增强结构稳定性。Tian 等人(2025b)提出的 SFormer 引入频域信噪比以辅助细节恢复。Zhuang 等人(2025)提出的 GS-Transformer 则强化浑浊场景下的关键区域建模。Shang 等人(2025)通过亮度引导的自注意力机制,解决了水下光照非均匀导致的局部过暗与过曝共存问题。这些方法表明,先验注入与全局建模的结合有助于提升复杂水下场景中的增强稳定性。在增强效果得到改善的同时,上述方法的计算效率问题也随之凸显。它们大多在原始空间分辨率上进行注意力计算,自注意力的计算复杂度随输入图像尺寸增加而迅速上升,限制了其在高分辨率水下图像处理场景中的适用性。

为缓解计算压力,研究者们进一步将频域分析与高效主干相组合,以降低高分辨率输入下的计算开销。Khan 等人(2024)提出的 Spectroformer 通过多域级联查询将空间建模与频谱信息结合;Mishra 等人(2025)提出的 USWFormer 利用稀疏小波表示降低高分辨率计算开销;Zhao 等人(2024a)提出的小波/傅里叶交互与频率扩散调节则将频域先验与生成式去噪过程相结合,为后续跨范式融合提供了可行路径。上述频域优化策略虽在一定程度上缓解了 Transformer 在高分辨率输入下的计算压力,但该类方法仍属于单次前向传播框架,在极重度退化场景下仍难以达到理想的增强效果。

1.4 基于 Diffusion 的方法

单次前向传播框架在退化场景下的性能瓶颈,催生了对多步迭代生成范式的探索。扩散模型通

过逐步去噪逼近目标分布,展现出更高的保真度上限,因而被引入水下图像增强任务。DDPM 为逐步加噪和反向去噪提供了稳定生成框架(Ho 等, 2020), DDIM 则通过隐式采样压缩反向链路,为扩散模型加速提供了基础(Song 等, 2021a)。在此框架下,研究者从采样效率与计算架构两个层面推进加速策略。在采样效率层面,Tang 等人(2023)提出 DM-water,通过跳步采样减少扩散反向过程的计算量。Wu 等人(2025)进一步将预训练扩散先验与显式场景分解结合,实现少步乃至单步潜空间采样,以改善推理效率。在计算架构层面,Yao 等人(2025)提出 UWLDM,将扩散过程迁移到潜空间以降低显存和推理开销。上述加速策略主要围绕推理效率展开优化,对水下成像物理规律与生成结果语义一致性的关注相对有限。

与加速推理导向的工作相比,另一类基于扩散的方法则侧重增强过程的物理可信性与语义可控性。Qing 等人(2025)提出 DiffUIE,利用全局先验提升细节恢复稳定性。Bach 等人(2024)提出 UW-DiffPhys,通过物理分支提高成像一致性。Ding 等人(2025)提出 ALDiff-UIE,将扩散模型与对抗学习结合以增强视觉真实感。Liu 等人(2024)提出基于定制对比语言-图像预训练(contrastive language-image pre-training, CLIP)分类器的水下图像增强扩散模型 CLIP-UIE,引入语义一致性约束以抑制语义漂移。Zhao 等人(2024b)提出 PA-Diff,从物理感知条件建模角度约束扩散恢复过程。Zhang 等人(2025)通过颜色先验引导扩散过程,弥补了常规扩散模型在水

下颜色恢复中缺乏物理引导的不足。Bi 等人(2025)通过物理先验与颜色先验联合建模,缓解了水下配对数据稀缺导致的扩散先验偏差问题。Hu 等人(2025)提出 DATDM,通过动态注意力和条件建模提升复杂退化场景中的适配能力。上述工作共同表明,扩散模型的价值不仅在于高保真恢复,更在于其天然支持物理与语义约束的多条件注入,为构建可控可信的增强系统提供了新的可能性。

1.5 多目标优化与跨范式融合趋势

上述 GAN、Transformer 与 Diffusion 范式在推理速度、全局一致性与高保真恢复方面各具优势与局限,促使研究者探索多目标权衡与跨范式协同。在模型结构优化层面,UIE 方法设计已从单模型单指标最优转向质量、时延及任务收益的联合优化,由单一范式走向跨范式融合。

围绕这一转向,Ma 等人(2024a)提出模糊神经网络架构搜索(fuzzy neural architecture search, fuzzy NAS),将不确定性建模引入架构搜索以提升候选结构评估的稳健性;其后续提出的 Pareto-wise 排序将精度与复杂度的非支配比较前移到搜索阶段(Ma 等, 2024b);而正交梯度 one-shot 训练通过梯度解耦可缓解共享权重场景下的多模型遗忘(Ma 等, 2025);Long 等人(2025)提出层次融合中的不一致性抑制机制,验证了层次融合与一致性约束相结合对性能与稳定性的协同增益。上述多目标优化与架构搜索思想在候选结构筛选、复杂度约束和多分支融合稳定性方面的方法论探索,可为后续 UIE 模型自动化设计提供参考。

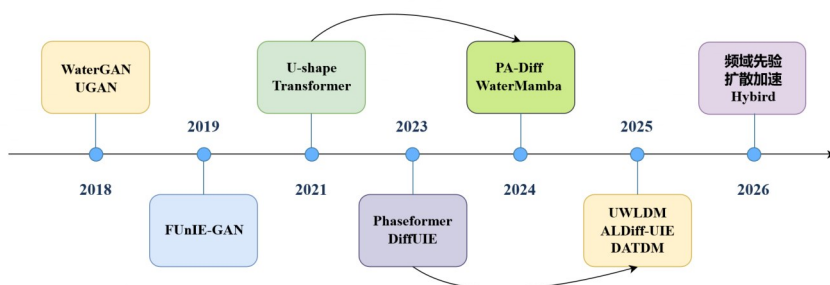


图 1 2018 - 2026 年 UIE 技术演进与跨范式融合趋势

Fig. 1 Evolution of UIE Technologies and Cross-Paradigm Fusion Trends, 2018-2026

如图 1 所示,UIE 技术自 2018 年以来经历了从对抗生成到全局建模再到高保真扩散的范式演进:

早期以 WaterGAN 和 UGAN 为代表的 GAN 方法奠定了数据驱动基础;FUnIE-GAN 等轻量化工作推动了

实时部署;2021年U-Shape Transformer引入全局依赖建模,突破了局部感受野限制;2023年前后Phaseformer与DiffUIE分别从频域先验和扩散去噪角度提升恢复质量;2024—2025年PA-Diff、WaterMamba、UWLDM、ALDiff-UIE及DATDM等方法在物理约束、

状态空间效率与潜空间加速方面持续深化;面向2026年的研究趋势则进一步呈现出频域先验、扩散加速与Hybrid架构的跨范式融合特征。上述时间线清晰展示了UIE方法从单一范式最优逐步转向质量、时延与任务收益联合优化的发展脉络。

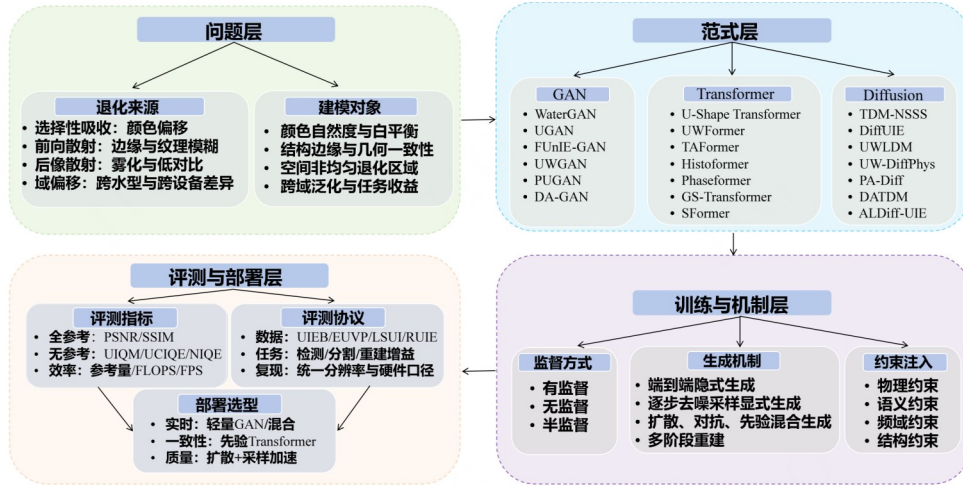


图2 水下图像增强方法分类框架

Fig. 2 Taxonomy Framework for Underwater Image Enhancement Methods

2 基于生成式网络的水下图像增强方法

本章将从水下图像退化的主要来源出发,从生成主干、监督方式、生成机制和约束目标四个层面系统梳理基于生成式网络的水下图像增强方法。已有UIE综述通常围绕物理模型、数据集、评价指标、损失函数和网络结构展开讨论(Anwar和Li,2020; Cong等,2024);本文则进一步聚焦于生成式网络在图像增强中的作用机制,即GAN侧重通过对抗学习逼近清晰域分布,Transformer侧重通过全局依赖建模维持颜色与结构一致性,Diffusion侧重通过逐步去噪实现高保真和可控恢复。

因此,第三章围绕“生成机制如何服务于水下退化恢复”展开分类:GAN范式对应从像素映射到分布对齐的增强思路;Transformer范式对应从局部卷积到全局依赖建模的主干演进;Diffusion范式对应从单次端到端映射到多步条件生成的恢复机制。后续各节将分别从理论基线、图像复原通用机制和UIE代表方法三个层次展开。

如图2所示,本文从退化来源、生成范式、训练

机制与评测部署四个层面构建UIE方法分类框架,贯通问题定义到工程落地的分析脉络;在此基础上,图3展示了三类范式在典型架构上呈现明显差异:GAN通过生成器与判别器对抗逼近清晰分布,Transformer基于全局自注意力实现编码-解码建模,Diffusion则通过逐步去噪实现高保真条件恢复。

2.1 水下图像退化建模

水下退化过程可以用以下成像模型描述(Anwar和Li,2020; Cong等,2024):

$$\mathbf{x} = \mathbf{y}e^{-\beta^d(v_d)d} + \mathbf{B}^*(1 - e^{-\beta^d(v_d)d}) \quad (1)$$

式中, $\mathbf{x}$ 为观测到的退化图像, $\mathbf{y}$ 为清晰场景图像; β^d 与 β^b 分别为直接透射与后向散射的衰减系数; v_d 与 v_b 分别对应两类光传输分量的波长; d 为目标与相机的距离; $\mathbf{B}^*$ 为背景光。第一项描述场景辐射在传输过程中的衰减,第二项描述后向散射带来的背景光叠加。水下图像增强的目标是在这个公式所描述的退化条件下,从观测到的退化图像 $\mathbf{x}$ 中恢复清晰的场景图像 $\mathbf{y}$ 。

从建模视角分析,UIE的核心挑战在于同一观测可对应多种潜在清晰解。公式中的透射项与散射项都受深度和水体参数影响,而这些变量通常无法

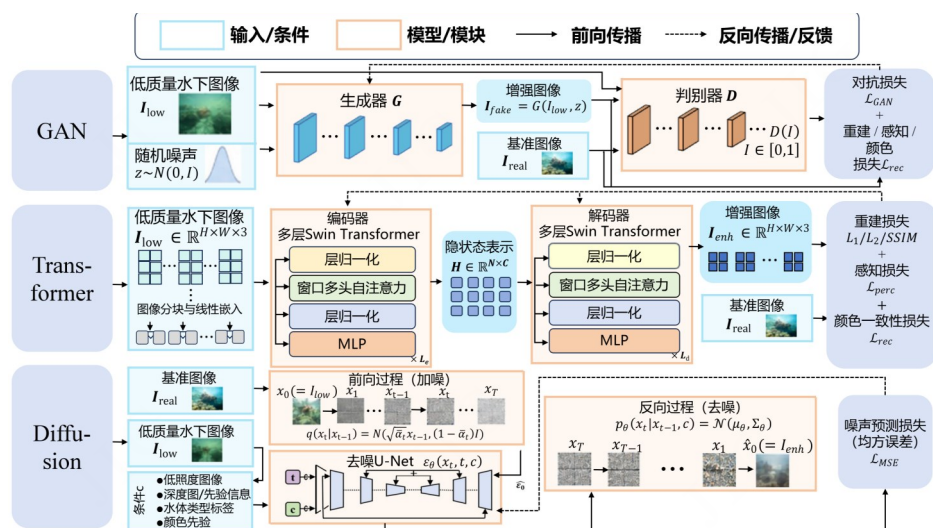


图3 水下图像增强典型架构

Fig. 3 Typical Architectures for Underwater Image Enhancement

直接观测,因此反演过程天然存在不适定性(Anwar和Li, 2020; Cong等, 2024):若仅以像素误差为优化目标,模型容易学到保守平均解;若仅以视觉锐度为优化目标,则易引入伪纹理。因此,实际建模对象至少包含四类变量:颜色分布(偏色校正与白平衡);结构边缘(细节保持与几何一致性);空间非均匀性(局部照明补偿)以及跨设备跨水型稳定性。

水下光学图像处理研究也表明,水下图像清晰化需要同时处理光吸收、散射、颜色失真和成像设备条件等因素,单纯套用陆地图像去雾模型难以覆盖全部水下退化情形(郭银景等, 2021)。与此同时,海雾图像去除研究中关于暗区域估计、相对总变差和透射率优化的思路,也可为海洋散射类场景中的能见度恢复提供方法启发(黄鹤等, 2021)。

从网络结构角度分析,物理参数的嵌入主要有三种模式:1)在GAN架构中,PUGAN(Cong等, 2023)设计了专门的参数估计子网络,显式学习衰减系数与深度图,进而推导透射率,并通过双流交互增强子网络将物理参数与生成特征融合;2)在Transformer架构中,CGFormer(Liang等, 2026)采用双分支设计,物理恢复分支通过估计背景光和透射率进行模型反演,提供颜色和亮度先验以引导主分支的特征学习;3)在Diffusion架构中,PG-TIE(Zhang等, 2026)通过物理引导先验生成模块估计透射率图和背景光,作为条件特征输入扩散Transformer。这些设计表明,物理参数在网络中的位置和作用方式各不相同,如作为独立分支、作为先验引导、作为条件

输入。这种物理参数在网络结构中的嵌入深度差异,直接影响了模型在跨场景泛化时的物理可解释性和场景适应性。

基于以上退化分析,后续比较将围绕三条主线展开:1)视觉质量提升与伪影抑制之间的平衡,2)结构与物理一致性的可控性,3)在计算资源约束下能否带来稳定的下游任务收益,如检测、分割、重建等。

2.2 基于GAN的水下图像增强方法

2.2.1 GAN理论基线

如图4所示,Goodfellow等提出的原始GAN通过生成器与判别器的极小极大博弈直接学习数据分布(Goodfellow等, 2014):

$$\min_G \max_D V(D, G) = E_{x \sim p_d} [\log D(x)] + E_{z \sim p_z} [\log (1 - D(G(z)))] \quad (2)$$

式中, G 为生成器, D 为判别器, p_d 为真实数据分布, p_z 为噪声先验。该式第一项表示判别器对真实样本的识别能力,第二项表示生成器通过伪样本“欺骗”判别器的能力;两者博弈的目标是让生成分布逼近真实分布。由于该博弈在训练中容易出现梯度不稳定,后续工作开始从损失函数形态上改进。

为缓解模式坍塌与训练不稳定,Wasserstein生成对抗网络(Wasserstein generative adversarial network, WGAN)以Wasserstein距离替代传统散度,为优化提供更平滑的信号(Arjovsky等, 2017)。WGAN-GP进一步通过梯度惩罚约束判别器的Lipshitz连续性,从而改善训练稳定性(Gulrajani等,

2017)。

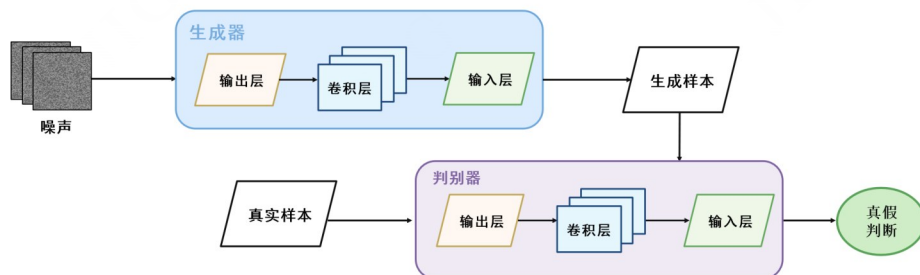


图4 生成对抗网络基本结构示意图

Fig. 4 Schematic of the Basic Structure of Generative Adversarial Networks

条件生成对抗网络(conditional generative adversarial network, cGAN)将退化水下图像作为条件输入、清晰图像作为生成目标,把无条件生成转化为面向特定输入的条件映射(Mirza 与 Osindero, 2014); pix2pix 进一步将这一思路用于图像到图像翻译(Isola 等, 2017)。

CycleGAN 通过循环一致性约束,使跨域映射后的样本尽可能回到原域,从而减少内容漂移和结构扭曲(Zhu 等, 2017)。在图像复原中,感知损失还可利用预训练网络的高层特征度量视觉差异,缓解像素损失导致的过度平滑(Johnson 等, 2016)。

Karras 等首次提出的 StyleGAN 通过风格映射与逐层调制提升生成质量(Karras 等, 2019)。StyleGAN2 进一步抑制伪影并改进训练稳定性(Karras 等, 2020),后续工作则继续强化平移与时域一致性(Karras 等, 2021)。自适应实例归一化(adaptive instance normalization, AdaIN)利用目标风格的通道统计量调制内容特征,为 UIE 中的颜色与纹理控制提供基础(Huang 和 Belongie, 2017)。

2.2.2 GAN 水下图像增强应用

UIE 中采用 GAN 的一个关键动机,是现实数据很难获得稳定的退化—清晰配对。WaterGAN 通过显式模拟水下成像退化过程,生成退化图像及其清晰参考,以构造配对训练样本(Li 等, 2018)。

与之对应, Fabbri 等人提出的 UGAN 等无配对方法更强调分布对齐,弱化对像素级严格配对的依赖(Fabbri 等, 2018)。这类方法通常联合对抗、循环一致性、同一性、重建、物理、颜色和结构等约束,不同方法的具体损失组合有所差异。上述约束可概括为统一的 UIE-GAN 损失框架:

$$L_g = L_a + \lambda_r L_r + \lambda_{ph} L_{ph} + \lambda_{co} L_{co} + \lambda_{st} L_{st} \quad (3)$$

式中, L_a 为对抗损失,用于逼近域整体分布; L_r 为重建损失,约束内容不被改写; L_{ph} 为物理一致性损失,限制结果满足成像规律; L_{co} 为颜色损失,抑制颜色漂移; L_{st} 为结构损失,对齐边缘和几何关系; λ_r 、 λ_{ph} 、 λ_{co} 、 λ_{st} 为对应权重系数。这些损失项的相对权重显著影响模型性能:对抗项过强易导致不真实纹理,重建项过强则使结果过度平滑。

Islam 等人提出的 FUNIE-GAN 聚焦工程可用性,通过轻量化编码器—解码器、局部判别器以及像素和感知约束换取高吞吐推理(Islam 等, 2020b)。在域外输入下,轻量主干也可能把局部噪声误判为纹理,产生色偏、边缘抖动或伪影。

GAN 范式在 UIE 中扮演低时延前端增强器的角色:Li 等人(2018)提出的 WaterGAN 侧重补足配对数据, Fabbri、Wang、Cong 与 Liu 等人(2026)提出的 UGAN/UWGAN/PUGAN/DA-GAN 侧重增强无配对学习的稳定性(Fabbri 等, 2018; Wang 等, 2019; Cong 等, 2023), Islam 等人(2020b)提出的 FUNIE-GAN 侧重把推理时延降低到可部署水平。该类方法推理效率高、部署链路成熟,适合在线感知前端;但对数据覆盖和损失权重较为敏感,在重度浑浊、强非均匀照明或域外输入下更容易出现伪影、纹理幻觉和过饱和。

2.3 基于 Transformer 的水下图像增强方法

2.3.1 Transformer 理论基线

如图 5, Vaswani 等(2017)提出的 Transformer 通过自注意力机制建立长程依赖关系。对于输入特征 X , 注意力计算过程如下:

$$A(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (4)$$

式中, Q, K, V 分别为查询、键和值矩阵, d_k 为键向量维度, QK^T 为查询与键的相似度矩阵。相似度除以 $\sqrt{d_k}$ 用于缩放点积, softmax 将其归一化为注意力权重, 再与 V 相乘得到加权聚合结果, 因此每个位置都能根据与其他位置的相关性汇聚全局信息。

缩放点积注意力通过查询与键的相似度计算全局依赖, 并对值进行加权聚合; 多头机制进一步从不同子空间学习互补关系。因此, Transformer 能够在整幅图像范围内统一建模颜色与结构语义 (Vaswani 等, 2017)。

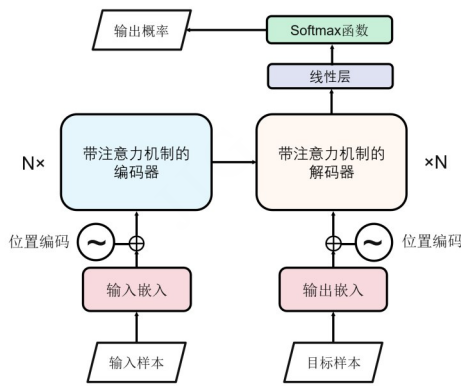


图 5 Transformer 编码器-解码器架构示意图

Fig. 5 Schematic of the Transformer Encoder-Decoder Architecture

Dosovitskiy 等 (2021) 提出的视觉 Transformer (vision transformer, ViT) 通过图像分块序列化实现纯注意力视觉建模, 但全局注意力计算复杂度为 $O(N^2)$ 。Liu 等提出的 Swin Transformer 采用窗口与移位窗口机制, 窗口注意力可形式化为:

$$A_w(X) = \bigoplus_{m=1}^M A(X_m) \quad (5)$$

式中, X_m 为第 m 个局部窗口特征, M 为窗口总数, $\bigoplus$ 表示拼接操作。该机制将全局注意力分解为窗口内独立计算, 复杂度由全局二次增长降为窗口内二次增长并近似线性扩展 (Liu 等, 2021)。

Mehta 等提出的 MobileViT 将卷积局部先验与全局自注意力结合, 以降低轻量场景下纯注意力主干的训练和部署成本 (Mehta 和 Rastegari, 2022)。Uformer 通过 U 形结构和局部窗口注意力服务于图像复原 (Wang 等, 2022)。Restormer 进一步验证了

高效注意力、通道建模和多尺度编码—解码框架的价值 (Zamir 等, 2022)。

2.3.2 Transformer 水下图像增强应用

Transformer 在 UIE 中的主要作用是增强全局建模能力, 抑制大范围色偏和非均匀照明带来的全局颜色与结构失配。其核心计算基于自注意力机制 (Vaswani 等, 2017), 通过全局特征交互建模颜色分布与结构信息, 从而缓解跨区域色偏和非均匀照明问题; 但高分辨率输入会带来较高的计算和显存开销。

作为将 Transformer 主干引入水下图像增强的早期代表, Peng 等人 (2023) 提出的 U-Shape Transformer 保留编码器-解码器和跳连结构, 仅将卷积主干换成注意力主干。该结构的功能分工如下: 编码端提取全局上下文, 解码端重构像素级表示, 跳跃连接则保留边缘与细节信息。后续方法的共同策略是在注意力分数中添加外部先验信息。TAFormer 将传输图作为先验, 强调不同深度处的衰减差异 (Li 等, 2025)。Histoformer 借助直方图或统计分布修正颜色偏移 (Peng 等, 2025)。Phaseformer 将相位一致性作为结构约束, 以抑制边缘漂移 (Khan 等, 2025)。SFormer 使用频域信噪比区分噪声与细节 (Tian 等, 2025b)。GS-Transformer 则显式建模受浑浊影响的关键特征, 从而在强退化下保持更稳的恢复方向 (Zhuang 等, 2025)。这些先验注入机制可以统一概括为:

$$A_h = \text{Softmax}\left(\frac{QK^T + \phi(h)}{\sqrt{d}}\right)V \quad (6)$$

式中, $\phi(h)$ 是任意形式的外部先验偏置, h 可以是传输图 t 、颜色统计量、相位图、频域信噪比图或关键区域掩码。

在应用层面, 外部先验可以通过多种方式注入 Transformer。Histoformer 利用输入图像的直方图统计调制中间特征, 实现颜色分布的自适应校正 (Peng 等, 2025)。Phaseformer 通过约束增强结果与参考图像的频域相位一致性, 抑制边缘漂移并保持几何轮廓稳定 (Khan 等, 2025)。SFormer 根据频域信噪比生成权重掩码, 优先保留结构频带并抑制噪声主导频带 (Tian 等, 2025b)。GS-Transformer 将浑浊区域掩码注入注意力计算, 引导模型重点恢复关键区域结构 (Zhuang 等, 2025)。

效率方面, Spectroformer 通过空间—频域级联查询提升细节恢复稳定性(Khan 等, 2024)。USW-Former 利用稀疏小波表示降低高分辨率计算成本(Mishra 等, 2025)。WaterMamba 将注意力替换为线性复杂度的状态空间更新(Guan 等, 2024a)。O-Mamba 则进一步探索 O-shape 状态空间结构对水下退化的适配能力(Dong 等, 2024)。

在标注受限条件下, UWFormer 利用伪标签与多尺度训练维持增强性能(Chen 等, 2024), 表明注意力主干可以通过半监督或弱监督方式降低对大规模精确配对数据的依赖。Phaseformer 报告的轻量参数量与毫秒级推理进一步表明, 通过频域先验和结构设计, Transformer 也能以较低成本获得全局建模能力(Khan 等, 2025)。

综上, Transformer 的核心优势是全局一致性。依靠主干中的全局交互, Transformer 在大面积色偏、非均匀照明或需要保留结构轮廓的任务中通常更稳定; 但高分辨率输入会带来较高计算和显存开销, 实际部署往往需要窗口化注意力、蒸馏、混合卷积-注意力结构或状态空间替代项控制成本。在强非均匀照明条件下, Transformer 也可能放大噪声并出现边缘振铃, 即边缘附近出现不应有的明暗波纹伪影, 表现为轮廓两侧交替过冲与欠冲。

2.4 基于 Diffusion 的水下图像增强方法

2.4.1 Diffusion 理论基线

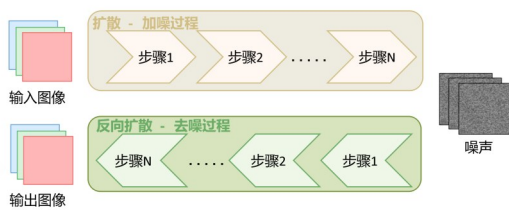


图6 DDPM 前向加噪与反向去噪过程示意图

Fig. 6 Schematic of the DDPM Forward Noising and Reverse Denoising Processes

如图6所示, DDPM 通过逐步向样本加入高斯噪声, 并学习相应的反向去噪过程; 其训练通常以噪声预测误差为目标, 使模型学习从加噪状态恢复清晰信号(Ho 等, 2020)。

Song 等(2021)提出的 DDIM 引入隐式确定性采样, 大幅减少推理步数:

$$\mathbf{x}_{t-1} = \sqrt{\bar{\alpha}_{t-1}} \hat{\mathbf{x}}_0 + \sqrt{1 - \bar{\alpha}_{t-1}} \hat{\epsilon} \quad (7)$$

式中, $\hat{\mathbf{x}}_0$ 为当前时刻 t 对原始干净样本的估计; $\hat{\epsilon}$ 为预测的噪声分量; $\bar{\alpha}_{t-1}$ 为前一时刻的累积噪声系数。该更新允许跨步采样(即 t 的间隔可大于1), 在维持生成质量的同时压缩反向链路长度, 是扩散模型走向近实时部署的重要基础。

Rombach 等(2022)提出的潜在扩散模型(latent diffusion model, LDM)进一步将扩散过程迁移至潜空间, 有效降低计算复杂度:

$$\mathbf{z} = E(\mathbf{x}), \hat{\mathbf{x}} = D(\mathbf{z}), p_{\theta}(\mathbf{z}_{t-1}|\mathbf{z}_t) \quad (8)$$

式中, $E(\cdot)$ 为编码器, 将输入图像 $\mathbf{x}$ 映射为低维潜变量 $\mathbf{z}$; $D(\cdot)$ 为解码器, 将潜变量重建为图像 $\hat{\mathbf{x}}$; $p_{\theta}(\mathbf{z}_{t-1}|\mathbf{z}_t)$ 为潜空间上的反向去噪分布。其核心思想是在低维潜变量上完成扩散与采样, 再解码回图像空间, 从而显著降低高分辨率 UIE 的内存与算力开销。在图像复原方向, SR3 展示了条件扩散可用于从低质量输入逐步恢复高分辨率图像, 去噪扩散复原模型(denoising diffusion restoration models, DDRM)则说明预训练扩散模型可作为逆问题复原的生成先验, 二者共同构成扩散模型从无条件生成走向条件复原的重要桥梁(Rombach 等, 2022; Saharia 等, 2023)。循此思路, 后续工作分别从跳步采样、潜空间建模、多尺度推理与少步/单步蒸馏等方向持续加速, 具体代表方法在下文逐项展开。

基于上述进展, 近年扩散加速路线可归纳为四类。

1) 跳步采样(non-uniform skip sampling), 代表如 Tang 等人(2023)的 DM-water。其核心是在反向链路中选取非均匀时间子序列。具体形式为:

$$S\{t_K > \dots > t_1 > 0\}, K \ll T \quad (9)$$

式中, T 为扩散模型完整反向过程的总步数, K 为实际采样的子序列长度且 $K \ll T$, S 为从 T 中非均匀选取的递减时间子序列。该机制仅在选取的 K 个关键时间步上执行反向更新, 降低推理成本的同时保留对生成质量影响较大的步数以维持恢复保真度。

2) 潜空间扩散, 代表如 Yao 等人的 UWLDM (Song 等, 2021a) 在编码后的低维潜变量上完成加噪和去噪, 再通过解码器重建图像, 从而降低显存占用并利用潜变量中的高层语义信息维持重建能力。

3) 多尺度推理, 代表如 Fang 等人(2025)通过多尺度损失联合约束, 使粗尺度恢复整体结构、细尺度修正边缘细节, 从而兼顾全局布局与局部精度。

4)少步/单步蒸馏(Wu 等, 2025),将长链知识压缩为浅链映射,以降低推理时延,但质量上限受教师模型精度与蒸馏目标设计影响(Wu 等, 2025)。

2.4.2 Diffusion 水下图像增强应用

扩散模型在 UIE 中的核心价值在于可控恢复:相较于单次映射模型,其更易在同一生成链中同时注入物理、语义与任务约束。其应用目标通常可形式化为:

$$L_d = L_b + \lambda_c L_c + \lambda_{ph} L_{ph} + \lambda_{se} L_{se} + \lambda_t L_t \quad (10)$$

式中, L_b 为基础扩散损失,用于驱动模型学习噪声到清晰图像的去噪映射; L_c 、 L_{ph} 、 L_{se} 、 L_t 分别为条件对齐损失、物理一致性损失、语义一致性损失和任务驱动损失; λ_c 、 λ_{ph} 、 λ_{se} 、 λ_t 为对应权重系数,用于平衡各约束的重要性。

在应用层面,当前扩散 UIE 大致分为三类。第一类侧重高保真恢复,代表方法包括 DiffUIE(Qing 等, 2025)和 UW-DiffPhys(Bach 等, 2024),主要提高复杂退化下的稳定细节重建。第二类侧重近实时推理,DM-water 通过采样链压缩降低推理成本(Tang 等, 2023),UWLDM 通过潜空间推理降低计算开销(Song 等, 2021a)。第三类侧重任务协同,ALDiffUIE 通过扩散与对抗学习耦合增强视觉真实感(Ding 等, 2025),CLIP-UIE 通过语义一致性约束抑制语义漂移(Liu 等, 2024),PA-Diff 通过物理感知单元注入相关物理信息(Zhao 等, 2024b),DATDM 则在条件建模与计算效率之间寻求折中(Hu 等, 2025)。

该范式的边界同样清晰:当采样步数被进一步压缩、输入显著偏离训练分布或条件约束噪声过大时,扩散模型容易出现欠恢复、细节漂移与语义偏差。因此在工程部署中,扩散方法更宜作为质量上限与可控性后端,与前端轻量模型形成分层协同。

2.5 水下图像增强监督方式

监督方式直接决定了数据成本、训练可行性与性能上限。相关综述指出,配对数据、无配对真实数据与伪标签构建方式直接影响 UIE 方法的可比性和泛化性能。在 UIE 中,按训练信号分为三类,其统一形式可表示为(Anwar 和 Li, 2020; Cong 等, 2024):

$$L = \lambda_s L_s + \lambda_u L_u + \lambda_{sm} L_{sm} \quad (11)$$

式中, L_s 为来自配对样本的有监督损失; L_u 为来自无配对分布对齐的无监督损失; L_{sm} 为由少量真值与大

量伪标签共同构成的半监督损失; λ_s 、 λ_u 、 λ_{sm} 为对应权重系数,用于平衡三类训练信号在总目标中的相对贡献。

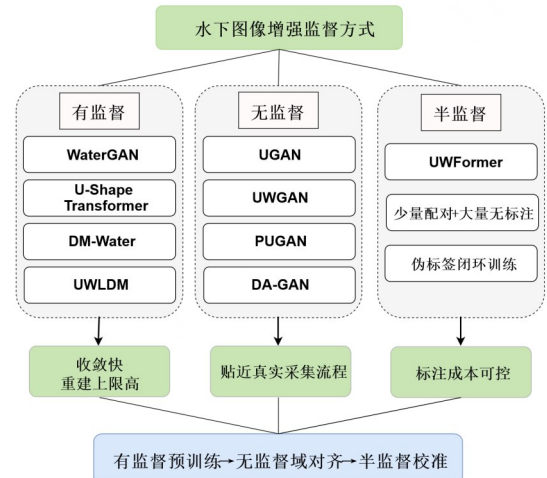


图 7 UIE 监督方式层次框架图

Fig. 7 Hierarchical Framework of UIE Supervision Methods

如图 7 所示按监督方法归类可得到清晰边界:

1)有监督或合成监督: WaterGAN(用于构造合成配对训练数据)(Li 等, 2018)、U-Shape Transformer(Peng 等, 2023)、DM-water(Tang 等, 2023)、UWLDM(Song 等, 2021a)。该类方法优势在于收敛速度快、重建性能上限高,但显著依赖于高质量配对数据。

2)无监督: UGAN(Fabbri 等, 2018)、UWGAN(Wang 等, 2019)、PUGAN(Cong 等, 2023)、DA-GAN(Liu 等, 2026)。该类方法更贴近实际数据获取场景,但训练稳定性与跨域一致性难以保证。

3)半监督: UWFormer(Chen 等, 2024)为代表,利用少量配对数据与大量无标注数据构建伪标签闭环,在标注成本受限时提供有效折中方案。

在工程部署中,更常见的训练流程是阶段化组合:先以有监督训练建立质量上限,再用无监督域对齐提升跨场景稳健性,最后以半监督校准压缩标注成本与性能损失。

2.6 水下图像增强生成流程

从生成流程看,UIE 可分为单次端到端生成与多步迭代生成两类,其推理时延可近似表示为(Anwar 和 Li, 2020; Cong 等, 2024):

$$T_o \approx T_b, T_i \approx \sum_{k=1}^K T_s^{(k)} \quad (12)$$

式中, T_o 为单次端到端生成的推理总时延,近似等

于主干网络单次前向传播耗时 T_b ; T_i 为多步迭代生成的总时延, 为各迭代步耗时 $T_s^{(k)}$ 之和, 其中 K 为迭代总步数。

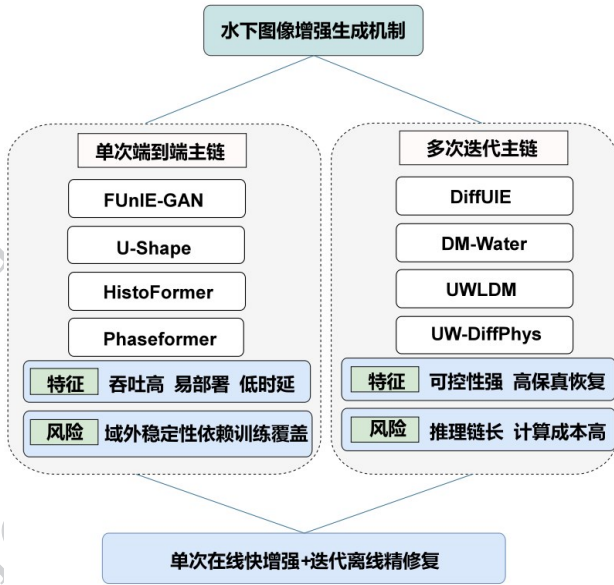


图 8 UIE生成机制层次框架图

Fig. 8 Hierarchical Framework of UIE Generation Mechanisms

从生成路径看, 单次端到端方法通过一次前向传播将退化图像直接映射为增强结果。FUnIE-GAN 强调高吞吐推理 (Islam 等, 2020b); U-shape 通过编码器—解码器和跳跃连接完成图像复原 (Peng 等, 2023); Histoformer 利用直方图统计进行颜色校正 (Peng 等, 2025); Phaseformer 通过频域先验保持结构轮廓 (Khan 等, 2025)。多步迭代方法则通过连续更新中间状态逐步得到增强结果。DiffUIE 强调复杂退化下的高保真恢复 (Qing 等, 2025); DM-water 通过采样链压缩降低推理成本 (Tang 等, 2023); UWLDM 在潜空间中完成扩散推理 (Song 等, 2021a); UW-DiffPhys 进一步引入物理一致性约束 (Bach 等, 2024)。

如图 8 所示两类机制适用于不同的部署目标。在严格时延约束下, 边缘实时场景通常采用单次端到端路径; 离线精细处理与高精度分析场景则更适用迭代路径。因此, 在统一系统中, 可结合两类路径构建分层推理架构: 前端采用单次端到端路径满足实时性需求, 后端采用迭代路径进行精细修复, 此策略更符合实际部署条件。

2.7 水下图像增强失效案例

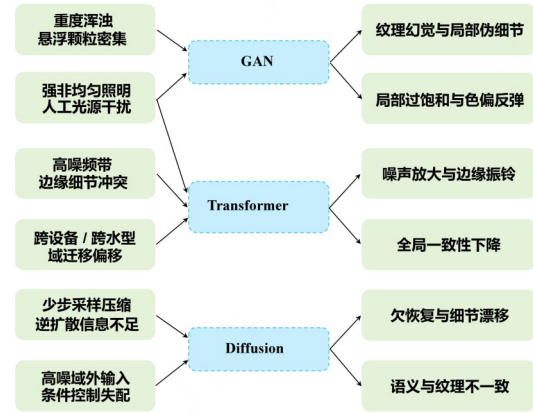


图 9 UIE 三类生成范式在典型退化场景下的失效模式示意图

Fig. 9 Schematic of Failure Modes of Three UIE Generation Paradigms under Typical Degradation Scenarios

从近期文献的失败案例看, UIE 在以下三类场景中最容易出现问題 (Anwar 和 Li, 2020; Cong 等, 2024):

1) 重度浑浊与悬浮颗粒密集: Fabbri 等人 (2018) 和 Islam 等人 (2020b) 的实验都指出, GAN 类方法在颗粒密集区域容易出现伪纹理或纹理幻觉; Khan 等人 (2025) 报告 Transformer 在高噪频带下可能放大局部噪声; Tang 等人 (2023) 的加速扩散设置则可能出现欠恢复与细节漂移。

2) 强非均匀照明与人工光源: Li 等人 (2025) 和 Liu 等人 (2026) 的结果显示, 局部强光与阴影并存时容易出现颜色过补偿, 表现为高亮区偏色、暗区细节塌缩或局部过饱和。

3) 跨设备与跨水型迁移: Cong 等人 (2024) 与 Wang 等人 (2023) 的跨域研究表明, 训练分布与测试分布不一致时, 常出现指标提升, 但观感或任务收益不一致的问题, 并可能降低下游任务稳定性。

如图 9 所示三类范式在典型退化场景下的失效分类。该图将失败现象映射到可追溯机理, 如对抗不稳定、全局注意力噪声放大、少步采样欠恢复, 便于统一呈现失败样例与缓解策略。

此外, 基础模型与多模态先验, 如视觉语言、任务文本约束等正在成为新的变量。它们有助于提供可迁移的语义约束与弱监督适配, 但也会带来额外

算力成本、提示敏感性与跨域偏置风险(Cong 等, 2024; Liu 等, 2024)。在本文的分类框架中,这类方法可视为生成式 UIE 与语义约束结合的扩展方向,其有效性仍需在实际水下任务和统一评测协议下进一步验证。

2.8 多范式融合方法

多范式融合的直接动因来自单范式失效边界: GAN 容易在重浑浊区域引入伪纹理, Transformer 在强噪频带可能放大噪声,扩散模型在强压缩采样下可能出现欠恢复。融合结构能够更好结合各个范式的优点,在复原质量和计算效率之间取得平衡(Tian 等, 2026; Tian 等, 2024),其设计的目标是把这些失败模式分解到不同分支中分别处理。

从目标函数看,多范式融合通常联合对抗、全局一致性、扩散、物理和任务约束(Tian 等, 2026)。在此基础上,不同分支还需针对颜色漂移、噪声放大和欠恢复等失效模式进行权衡(Tian 等, 2024)。

现有融合路径可归纳为三类。第一类为扩散主链、物理分支与对抗细化的结合, UW-DiffPhys 强调物理一致性(Bach 等, 2024), ALDiff-UIE 通过对抗细化增强视觉真实感(Ding 等, 2025)。第二类为条件扩散与全局建模主干的结合, PA-Diff 强化物理条件传递(Zhao 等, 2024b), DATDM 则在条件建模与计算效率之间寻求折中(Hu 等, 2025)。第三类为分层推理架构,在线阶段采用轻量 GAN 或 Transformer, 离线阶段调用扩散模型进行精修,以平衡时延约束与质量上限。

水下图像增强中的融合架构设计可从七个维度进行系统拆解:(1)多输入源融合强调以原始图像及其白平衡、对比度增强等预处理变体作为多路输入,使不同输入分别对模糊、低对比度等退化维度形成互补;(2)多尺度特征融合利用不同卷积核尺寸或金字塔结构提取不同感受野下的特征,关键在于尺度间信息交互的设计——注意力引导的自适应加权比简单拼接更能抑制冗余特征;(3)多域融合将空间域与频率域、或不同色彩空间的特征进行跨域桥接,"先解耦再融合"是较为有效的范式;(4)注意力引导的自适应融合通过通道注意力或空间注意力使融合权重随退化区域动态变化,更进一步可将物理光照先验注入注意力计算,实现数据驱动与物理约束的协同;(5)物理模型与数据驱动的融合将物理成像模型的退化参数嵌入网络特征空间进行联合训练,以

退化一致性损失确保物理参数与数据驱动特征在嵌入空间中对齐,兼顾可解释性与泛化能力;(6)全局-局部协同融合通过并行或串行架构同时建模全局性退化(色偏、亮度不均)与局部性退化(纹理模糊、边缘丢失),串行渐进式逐层融合更稳定,并行交互式信息保真度更高;(7)渐进式/级联融合将复合退化问题分解为多阶段子任务,每阶段聚焦特定退化类型,阶段间通过特征空间传递信息以避免误差累积。

从方法论角度看,UIE 的质量—效率—任务收益三维权衡与多目标优化研究高度一致。Ma 等(2022)提出的局部化决策变量分析为大规模多目标优化提供了可扩展的变量分解思路。在此基础上, fuzzy NAS 将不确定性建模引入架构搜索过程,提升了候选结构评估鲁棒性(Ma 等, 2024a)。Pareto-wise 排序进一步把非支配选择前移到搜索阶段,使精度与复杂度权衡更具可解释性(Ma 等, 2024b)。Long 等(2025)在层次融合架构搜索中引入不一致性抑制机制,也验证了层次融合与一致性约束结合可同时提升性能与稳健性。

融合方法表现出更好的综合性能,根本原因在于其将单范式的失败模式转化为多分支可优化目标;但其代价是训练耦合度上升、超参数空间扩大,因此仍需在统一评测协议下报告质量、时延与任务收益三类指标,避免单侧性能结论。

3 实验分析

3.1 评测指标

要让评测结果可复用,至少需要同时回答四个问题:增强是否还原了结构(重建质量),主观上是否自然(感知质量),部署代价(效率成本),以及是否真正提升了下游任务性能(任务收益)。从图像质量评价角度看,无参考质量评价和融合质量评价的相关研究也表明,单项客观指标难以完整刻画人眼感知质量,不同指标之间的相关性和适用场景需要被明确区分(王志明, 2015; 张小利 等, 2014)。对应地,常用指标可以归纳为四类:

1)重建质量: PSNR、SSIM, 用于度量与参考真值的误差与结构相似性(需要配对数据)。

2)感知质量: UIQM、UCIQE、自然图像质量评价器(natural image quality evaluator, NIQE),以及平均主观评分(mean opinion score, MOS)、成对偏好等主

观评价,更适合无真值场景下的视觉感知评价。

3)部署成本:参数量、浮点运算次数(floating-point operations, FLOPs)、推理时延或每秒帧数(frames per second, FPS),并同时说明输入分辨率与硬件平台。

4)任务收益:检测、分割、导航或三维重建等下游任务的指标增益,用于检验增强是否转化为任务价值。

表 1 从质量、效率、任务收益三个维度对三类范式做了整体对比。一般情况下,GAN 在时延受限场景更具优势,Transformer 在全局一致性与先验融合方面更稳健,Diffusion 在细节与高保真恢复方面更有上限,但往往需要采样加速与轻量化设计来降低部署成本。

3.2 评测协议

Anwar 等人(2020)的综述强调了水下成像退化机理与评测口径统一的重要性。Cong 等人(2024)的系统综述进一步指出,不同论文在数据划分与报告规范上的差异会显著影响横向结论。基于以上共识,可将 UIE 评测协议整理为数据、指标、效率、任务、可复现性五项基础要素:

1)数据:同时报告配对数据集(如 UIEB(Li 等, 2020b)、LSUI(Peng 等, 2023))与真实无参考数据集(如 EUVP(Islam 等, 2020b)、RUIE(Liu 等, 2020))的

划分方式;明确是否包含跨域测试;尽量采用统一划分规则(如固定随机种子或 5-fold 设置)。

2)指标:并行给出全参考指标(Wang 等, 2004) (PSNR、SSIM)与无参考指标(Zhang 等, 2018; Yang 和 Sowmya, 2015; Panetta 等, 2015)(学习感知图像块相似度(learned perceptual image patch similarity, LPIPS)、UCIQE、UIQM),必要时补充 MOS、成对偏好或专家打分等主观评价,避免只用单一视角评价;当两类指标冲突时,需要给出合理解释。

3)效率:报告参数量、FLOPs、推理时延或 FPS,并同时注明输入分辨率与硬件口径(如图形处理器(graphics processing unit, GPU)型号与显存)。

4)任务:至少给出一个下游任务增益(检测/分割/导航/重建),并说明任务模型与评测设置。

5)可复现:公开训练细节(分辨率、批大小、学习率、轮数、硬件环境等)以及代码/模型链接,保证结果可核验。

这一框架的意义在于:它将评测的核心要素明确化,减少了因指标不一致所导致的结论偏差问题;同时也强调了评测的多维度性,避免仅使用单一指标得出片面结论。后续章节将在此基础上进一步讨论退化条件细分、跨域测试和下游任务收益等开放问题。

表 1 三类生成范式的质量、效率与任务收益

Table 1 Comparison of Three Generative Paradigms in Terms of Quality, Efficiency, and Task Benefits			
范式	质量表现	部署效率	任务收益特征
GAN	视觉提升快,极端退化下稳定性一般	推理快,边缘部署友好	对实时检测/导航增益明显
Transformer	全局结构与颜色一致性较好	高分辨率成本偏高	对结构敏感任务(分割/重建)更稳定
Diffusion	高保真与细节恢复优势显著	计算开销高,需采样加速	对高精度感知任务潜在收益最高

3.3 数据集

为保证不同生成式网络方法的定量结果具有可解释性,本章首先对实验所涉及的数据集进行统一说明。本文将数据集按评价属性划分为配对全参考数据集和真实无参考数据集两类。前者提供退化图像及参考图像,可用于计算 均方误差(mean squared error, MSE)、PSNR、SSIM 和 LPIPS 等重建或感知指标;后者通常不具备像素级参考图像,更适合通过 UIQM、UCIQE、水下图像清晰度测量(underwater image sharpness measure, UISM)、NIQE 和无参考图

像空间质量评价器(blind/referenceless image spatial quality evaluator, BRISQUE) 等无参考指标评估真实水下场景中的颜色、清晰度和自然度。

其中,UIEB、UFO、LSUI 和 EUVP 主要用于配对评测,U45、UCCS、SQUID 与 UIEB-C60 主要用于无参考或跨域场景评测;UIEB 测试集(test split of UIEB, Test-UIEB)、UIEB-T90 和 UIEB-90 均可视为 UIEB 有参考测试子集在不同论文中的命名变体,LSUI 测试集(test split of LSUI, Test-LSUI) 则用于考察模型由 UIEB 训练后迁移到 LSUI 场景时的跨域

表现。

表 2-表 4 采用的数据划分、输入尺寸和训练策略并不完全一致,以下的定量结果主要用于同一实验设置下的横向比较,跨表结论应结合数据集属性和评价指标共同解释。本章节只将其中的数值作为同一范式或相近实验口径下的相对参考,不将异构

训练设置下的结果直接作为绝对性能排序依据。

为保证表 2-表 4 的比较公平性,表注中统一标明各方法的训练环境、批大小等信息,在相同数据划分与相同分辨率下重新训练后再比较,避免把不同训练口径下的数值直接视为同等可比的绝对结论。

4.4 定量比较

表 2 基于 GAN 的水下图像增强方法结果(输入尺寸为 256×256, batch size 为 64, 训练 200 iterations, 优化器为 Adam, 初始学习率为 1×10^{-4} , 最小学习率为 1×10^{-7} , 硬件为 NVIDIA RTX 3090)

Table 2 Results of GAN-Based Underwater Image Enhancement Methods (input size: 256×256; batch size: 64; training: 200 iterations; optimizer: Adam; initial learning rate: 1×10^{-4} ; minimum learning rate: 1×10^{-7} ; hardware: NVIDIA RTX 3090)

方法	UIEB					UFO					LSUI				
	PSNR	SSIM	LPIPS	UCIQE	UIQM	PSNR	SSIM	LPIPS	UCIQE	UIQM	PSNR	SSIM	LPIPS	UCIQE	UIQM
	↑	↑	↓	↑	↑	↑	↑	↓	↑	↑	↑	↑	↓	↑	↑
UMGAN (Sun 等, 2023)	21.05	0.88	0.09	0.50	1.39	26.31	0.89	0.10	0.62	1.63	26.03	0.90	0.06	0.63	1.59
UWGAN (Wang 等, 2019)	20.68	0.87	0.09	0.51	1.42	24.36	0.87	0.12	0.55	1.51	24.97	0.88	0.08	0.58	1.49
UGAN (Fabbri 等, 2018)	21.13	0.79	/	0.46	/	24.93	0.77	/	0.44	/	25.30	0.84	/	0.45	/
S-StyleGan (Fathy 等, 2023)	20.05	0.87	0.09	0.52	1.40	25.98	0.85	0.10	0.66	1.51	26.05	0.90	0.07	0.68	1.60
WaterGAN (Li 等, 2018)	20.26	0.84	0.10	0.54	1.40	22.38	0.84	0.15	0.63	1.37	23.52	0.84	0.09	0.61	1.37
FUnIE-GAN (Islam 等, 2020b)	21.14	0.89	0.09	0.51	1.42	23.78	0.88	0.12	0.58	1.42	24.74	0.89	0.08	0.57	1.44

注:加粗数据表示同一数据集、同一指标列中的最优结果;“↑”表示数值越大越好,“↓”表示数值越小越好;“/”表示未报告。

表 2 为同一实验设置下 GAN 类方法的定量结果,实验采用 UIEB、UFO 和 LSUI 三个配对数据集,训练环境为 PyTorch 和 NVIDIA RTX 3090 GPU,批大小为 64(You 等, 2025)。UMGAN 在三个数据集上的表现相对更均衡, UFO 和 LSUI 上的 PSNR 分别达到 26.31dB 和 26.03dB, LSUI 上的 SSIM 为 0.90。FUnIE-GAN 在 UIEB 上表现较突出, PSNR 和 SSIM 分别为 21.14dB 和 0.89,略高于 UMGAN。总体来看,剩余 GAN 方法在不同数据集间仍存在一定波动,说明对抗生成模型的增强效果受训练数据分布和模型约束方式影响较明显。

表 3 为同一实验设置下 Transformer 类方法的定量结果,实验在 UIEB 训练集(training split of UIEB, Train-UIEB)上训练,并在 Test-UIEB 测试集、

Test-LSUI 测试集以及 U45、UCCS、SQUID 等数据集上测试,训练环境为 PyTorch 和 NVIDIA GeForce RTX 2080 GPU,输入尺寸为 256×256(Mishra 等, 2025)。其中, USWFormer 在 Test-UIEB 和 Test-LSUI 上整体表现最稳定, PSNR 和 SSIM 均高于 U-shape 与 Spectroformer。以 Test-UIEB 和 Test-LSUI 中的结果为例, USWFormer 的 PSNR 分别达到 25.68 dB 和 25.34 dB,高于其他 Transformer 方法,说明其在同域测试和跨数据集测试中均具有更好的重建质量与结构保持能力。

表 4 为同一实验设置下 Diffusion 类方法的定量结果,实验采用 UIEB-T90、EUVP、UIEB-C60 和 U45 四个数据集,使用 PyTorch 于单张 NVIDIA RTX 4090D GPU 上训练(Chen 等, 2026)。其中, FPG-

表 3 基于 Transformer 的水下图像增强方法结果 (输入尺寸为 256×256, 优化器为 Adam, 初始学习率为 3×10^{-4} , 采用余弦退火学习率策略, 硬件为 NVIDIA RTX 2080)

Table 3 Results of Transformer-Based Underwater Image Enhancement Methods (input size: 256×256; optimizer: Adam; initial learning rate: 3×10^{-4} cosine annealing learning-rate schedule; hardware: NVIDIA RTX 2080)																	
方法	Test-UIEB		Test-LSUI		U45				UCCS				SQUID				
	PSNR	SSIM	PSNR	SSIM	UIQM	UISM	NIQE	BRISQU	UIQM	UISM	NIQE	BRISQU	UIQM	UISM	NIQE	BRISQU	
	↑	↑	↑	↑	↑	↑	↓	E↓	↑	↑	↓	E↓	↑	↑	↓	E↓	
U-shape (Peng 等 , 2023)	22.91	0.91	22.87	0.85	2.92	5.57	4.31	21.57	2.87	5.39	4.40	23.55	2.42	7.00	4.62	29.16	
Spectroformer (Khan 等 , 2024)	24.96	0.92	25.08	0.87	3.24	7.35	3.84	19.96	3.21	6.56	3.98	23.26	3.09	7.37	3.61	21.48	
Phaseformer (Khan 等 , 2025)	25.06	0.92	21.62	0.84	3.16	6.98	3.93	/	2.98	5.86	4.83	19.02	2.44	/	6.21	/	
WaterMamba (Guan 等 , 2024)	25.29	0.91	21.28	0.76	3.15	6.85	4.22	/	3.07	6.16	4.79	14.92	2.49	/	7.73	/	
USWFormer (Mishra 等 , 2025)	25.68	0.95	25.34	0.91	4.25	7.25	3.80	17.40	4.28	7.03	4.22	24.70	2.60	7.45	4.01	22.42	

注:加粗数据表示同一数据集、同一指标列中的最优结果;“↑”表示数值越大越好,“↓”表示数值越小越好;“/”表示未报告。

表 4 基于 Diffusion 的水下图像增强方法结果 (输入尺寸为 256×256, batch size 为 4, 训练 300 epochs, 优化器为 Adam, 初始学习率为 2×10^{-4} , 采用余弦退火学习率策略, 硬件为 NVIDIA RTX 4090D)

Table 4 Results of Diffusion-Based Underwater Image Enhancement Methods (input size: 256×256; batch size: 4; training: 300 epochs; optimizer: Adam; initial learning rate: 2×10^{-4} ; cosine annealing learning-rate schedule; hardware: NVIDIA RTX 4090D)											
方法	UIEB-T90			EUVP			UIEB-C60		U45		
	MSE ↓	PSNR ↑	SSIM ↑	MSE ↓	PSNR ↑	SSIM ↑	UIQM ↑	UCIQE ↑	UIQM ↑	UCIQE ↑	
DM-Water (Tang 等, 2023)	0.82	21.88	0.82	0.46	23.70	0.81	2.65	0.59	3.03	0.59	
DiffWater (Guan 等, 2024b)	0.72	20.97	0.90	0.40	24.56	0.85	2.82	0.58	3.06	0.60	
WF-Diff (Zhao 等, 2024a)	0.45	23.86	0.87	0.24	26.54	0.84	2.93	0.59	3.13	0.62	
UW-DiffPhys (Bach 等, 2024)	/	22.44	0.87	/	26.02	0.87	2.88	0.52	3.08	0.52	
DiffUIE (Qing 等, 2025)	0.47	23.27	0.89	0.21	26.87	0.83	/	/	/	/	
FD-LDM (Song 等, 2025)	0.43	23.52	0.91	/	/	/	2.79	0.58	3.10	0.61	
FPG-Diff (Chen 等, 2026)	0.41	24.11	0.91	0.13	28.57	0.86	2.74	0.57	3.13	0.61	

注:加粗数据表示同一数据集、同一指标列中的最优结果;“↑”表示数值越大越好,“↓”表示数值越小越好;“/”表示未报告。

Diff 在 UIEB-T90 和 EUVP 上整体表现最稳定, MSE、PSNR 和 SSIM 均处于最优或领先地位。以 UIEB-T90 和 EUVP 为例, FPG-Diff 的 PSNR 分别达到 24.11 dB 和 28.57 dB, 高于其他 Diffusion 方法, 说明其在配对数据集上具有更好的重建质量和结构保持能力。

3.5 物理先验约束有效性分析

为分析物理先验在水下图像增强领域的有效性, 本章节整理物理感知单元在不同骨干网络上的消融结果。表 5 中比较了 U 形网络(U-Net, UNet)、DW-water 和 U-shape 在 UIEBD 数据集上移除或加入物理感知单元(physical perception unit, PPU)后的 PSNR 和 SSIM 变化, 用以观察物理先验约束对不同骨干的增益(Zhao 等, 2024b)。

表 5 物理感知单元在 UIEBD 数据集上的消融结果
Table 5 Ablation Results of the Physical Perception Unit on the UIEBD Dataset

设置	PSNR ↑	SSIM ↑
UNet w/o PPU	17.41	0.7769
UNet w PPU	19.64	0.8157
DW-water w/o PPU	22.78	0.894
DW-water w PPU	22.93	0.897
U-shape w/o PPU	18.15	0.833
U-shape w PPU	20.54	0.848

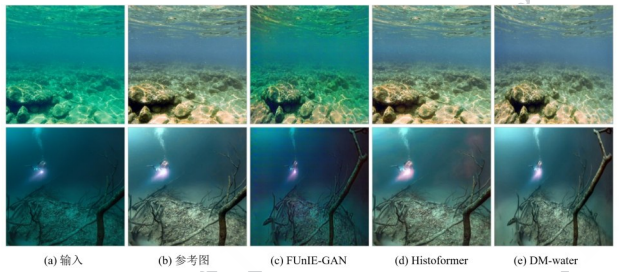
注:“↑”表示数值越大越好。

由表 5 可见, 在 UIEBD 数据集上, UNet、DW-water 和 U-shape 在加入 PPU 后的 PSNR 和 SSIM 均有提升, 说明该物理感知单元具有稳定的可插拔收益。其中, UNet 和 U-shape 的增益更明显, DW-water 的提升相对较小, 表明 PPU 的作用会随骨干结构而变化。总体上, 这说明物理先验信息能够在不改动主干框架的前提下改善重建质量。

3.6 可视化结果

为在定量指标之外补充直观比较, 本文选取 UIEB 基准数据集中具有参考图的两组配对样本进行可视化展示。UIEB 包含 890 张 raw/reference 配对图像和 60 张 challenging 图像, 本文仅使用其中具有参考图的配对子集(Li 等, 2020b)。样本优先选择主体结构清楚、参考图质量自然且能够代表典型水下退化模式的配对图像; 其中一组样本主要体现

蓝绿偏色和低对比问题, 适合观察颜色校正与整体亮度恢复; 另一组样本则以低照度、浑浊和细节模糊为主要特征, 适合观察暗部提升、纹理恢复与结构保持。这样的选图原则有助于在相对可解释的视觉场景中比较不同生成式路线的增强风格差异, 避免由于样本过暗、主体不可辨或参考图异常而影响结论判断。图 9 展示了三类生成式路线的代表方法: FUnIE-GAN(Islam 等, 2020b)作为 GAN 路线代表, Histoformer(Peng 等, 2025)作为直方图建模驱动的 Transformer 路线代表, DM-water/TDM-NS(Tang 等, 2023)作为 Diffusion 路线代表。图像结果主要用于辅助观察颜色校正、对比度恢复和结构保持方面的视觉表现。为保证比较条件一致, 图 9 中样本选自 UIEB 有参考图的配对子集, 结果由对应方法的公开预训练模型或官方推荐配置在相同输入图像上生成, 展示了 GAN、Transformer 与 Diffusion 三类路线在典型样本上的颜色校正、对比度恢复和结构保持效果。



((a)input; (b) ground truth; (c) FUnIE-GAN; (d) Histoformer; (e) DM-water)

图 10 代表性生成式水下图像增强结果
Fig. 10 Representative Generative Underwater Image Enhancement Results

从图 10 可知, 在该可视化样例中, FUnIE-GAN 对整体颜色和亮度的修正较为保守, 能够一定程度增强局部对比度, 但在复杂低照度场景中仍保留较明显的青绿色偏色。相比之下, Histoformer 和 DM-water 在全局颜色校正、亮度恢复和结构保持方面均更接近参考图; 在当前两组样本中, 二者的主观视觉差异主要体现在局部色彩过渡、暗部亮度和边缘细节上。

本文在图 11 所示样本基础上选取相同局部区域进行放大, 并计算增强结果相对于参考图的像素级平均绝对误差。可以看出, Histoformer 与 DM-water 在整体颜色校正和亮度恢复方面较为接近, 但

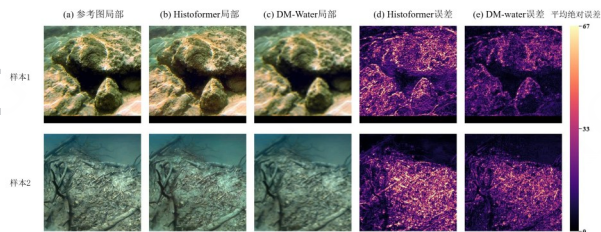


图 11 Transformer 与 Diffusion 方法的局部差异
Fig. 11 Local Differences between Transformer and Diffusion Methods

在局部误差分布上仍存在差别:Histoformer 在部分纹理和边缘区域呈现出更明显的高频误差,而 DM-water 在所选局部区域中的误差分布相对平滑。图中并列展示 Histoformer 与 DM-water 的局部增强结果及其相对于参考图的误差图,颜色越亮表示像素级平均绝对误差越大。

3.7 复杂度分析

为了对现有水下图像增强方法进行性能与复杂度的综合对比,从而为工程部署提供指导,表 6 汇总 TEST-U90 上不同方法的增强质量、参数量、FLOPs、训练时间和测试时间,并在同一表中补充统一服务器环境下的 FPS proxy 和 GPU 板卡功耗;该 FPS 按单张平均时延反推,用于同一服务器上的相

对运行时比较。UDCP 和 UIBLA 为传统方法因此参数量记为 0、FLOPs 不作可比统计。FPS 与 GPU 功耗来自 4×NVIDIA RTX 3090(每张 24 GB)服务器上的统一运行时基准,输入分辨率为 256×256、batch size 为 1、统一计算设备架构(Compute Unified Device Architecture, CUDA)同步计时,功耗为 NVIDIA 管理库(NVIDIA Management Library, NVML)读取的板卡功耗。

由表 6 可以看出,DiffWater 的 SSIM 为 0.8951,是表中最高值,PSNR 为 20.9721 dB,略低于 Ucolor 的 21.0005,说明其在结构保持方面具有优势,但像素级重建指标并非绝对最高。与此同时,DiffWater 的测试时间为 48.517 s,明显高于 Shallow-UWnet 等单次前向方法,反映出反向扩散多步采样带来的推理开销。进一步来看,这种性能提升并不是以更高的计算代价换来的,而是以更高的采样成本和更长的部署时延为代价,因此更适合对质量上限要求较高、对实时性要求相对宽松的离线场景。

MLFcGAN 在 RTX 3090 上达到 179.53 FPS, Ucolor 为 11.22 FPS;对应 GPU 平均功耗约为 292.5 W 和 310.5W,说明质量、吞吐量和能耗之间存在权衡。

表 6 不同方法在 TEST-U90 上的 PSNR、SSIM、参数量、FLOPs、FPS 与功耗
Table 6 PSNR, SSIM, number of parameters, FLOPs, FPS and power consumption of different methods on TEST-U90

方法	质量指标		复杂度		运行时	功耗
	PSNR ↑	SSIM ↑	参数量 (M)	FLOPs(G)	FPS	GPU 功耗(W)平均/峰值
UDCP(Drews 等,2013)	11.3752	0.5156	0.00	/	0.4584	/
UIBLA(Peng 和 Cosman,2017)	15.1351	0.6454	0.00	/	0.2011	/
Shallow-wnet (Naik 等,2021)	15.9740	0.7099	0.22	43.43	399.75	299.7 / 314.7
Ucolor (Li 等,2021)	21.0005	0.8693	108.52	1995.23	11.22	310.5 / 313.3
FUnIE-GAN (Islam 等,2020b)	16.9189	0.7338	7.02	7.20	600.20	261.9 / 277.1
MLFcGAN(Liu 等,2019)	15.1004	0.6573	52.22	31.38	179.53	292.5 / 297.5
DiffWater(Guan 等,2024b)	20.9721	0.8951	155.33	311.00	0.02146	294.2 / 306.9

注:加粗数据表示相应可比指标中的最优结果;PSNR、SSIM 和 FPS 越大越好,参数量、FLOPs 和 GPU 功耗越小越好;传统方法的参数量记为 0,不参与模型复杂度的最优比较;“/”表示未报告或不适用。

3.8 下游任务性能与感知质量分析

客观图像质量指标(如 PSNR、SSIM)与人类主观感知之间存在系统性偏差,且增强算法在视觉上

的改善未必能等价转化为下游视觉任务的性能提升。因此,本文进一步从主观感知质量和下游任务收益两个维度对代表性 UIE 方法进行补充评估。

3.8.1 主观感知质量与客观指标的关联分析

客观指标能否忠实反映人类主观感知,是UIE评测必须正视的前提问题。表7与表8展现了主观

MOS评价实验在UIEBD和EUVP两个数据集上将MOS与主流客观指标进行了相关性分析(Palazari和Dumic,2026)。

表7 UIEBD 准参考子集上 PSNR、SSIM 与 MOS 的比较

Table 7 Comparison of PSNR, SSIM, and MOS on the UIEBD Quasi-Reference Subset

算法模型	PSNR (dB)	SSIM	近似 MOS	p 值
CCL-Net(Liu 等,2024b)	23.269	0.922	3.25	0.006
GuidedHybSensUIR(Guo 等,2025)	22.329	0.936	3.60	0.473
HUPE)(Zhang 等,2025)	20.610	0.858	2.95	0.001
UDNet(Saleh 等,2025)	18.816	0.866	3.70	0.873
原始图像	15.758	0.724	3.85	基准参照

注:加粗数据表示相应指标列中的最优结果。

表8 UIQM、UCIQE 与主观 MOS 的相关性分析

Table 8 Correlation Analysis of UIQM, UCIQE, and Subjective MOS

数据集	指标	PLCC	SRCC
EUVP	UIQM	0.387	0.329
EUVP	UCIQE	0.165	0.137
UIEBD	UIQM	0.262	-0.052
UIEBD	UCIQE	0.349	-0.034

注:加粗数据表示该列相关性绝对值最大的结果;相关系数绝对值越大表示相关性越强。MOS 由原始图像及 CCL-Net(Liu 等,2024b)、GuidedHybSensUIR(Guo 等,2025)、HUPE(Zhang 等,2025)、UDNet(Saleh 等,2025)四种增强算法的结果图共同构成;相关性指标在整个测试集(每数据集 120 张图)上计算得出。

全参考指标的感知错位:在UIEBD准参考子集上,各项算法的PSNR、SSIM与MOS对比如客观指标能否忠实反映人类主观感知,是UIE评测必须正视的前提问题。表7与表8展现了主观MOS评价实验在UIEBD和EUVP两个数据集上将MOS与主流客观指标进行了相关性分析(Palazari和Dumic,2026)。所示。CCL-Net虽取得最高PSNR 23.269 dB,但其MOS显著低于原始图像,而原始图像的PSNR垫底。UDNet的PSNR最低,为18.816 dB,MOS却与原始图像无统计差异。这表明,PSNR/SSIM所奖励的像素级相似度并不等同于人眼感知质量,过度拟合有偏差的准参考图像反而可能损害主观质量。

水下专用指标的感知脱节:UIQM和UCIQE与MOS的相关性同样堪忧,如表8所示,在UIEBD上,

UIQM和UCIQE的SRCC分别仅为-0.052和-0.034,几乎与MOS无关,UIQM甚至呈微弱负相关;在EUVP上分别为0.329和0.137,仍处极低水平。这说明传统水下专用指标与人类主观判断严重脱节,不宜作为算法排序的唯一依据。

因此,UIE评测应根据任务目标划定指标适用边界,不应该只依赖单一指标为依据,主观视觉质量需结合MOS与感知指标,真实场景和下游应用则应提高任务性能及效率指标的权重,避免任何单一指标主导算法排序。

3.8.2 下游任务性能综合评估

为全面评估UIE方法对高层视觉任务的实际增益,本章节在语义分割和实例分割两个任务上,采用多种骨干网络,对比原始图像与多种增强方法作为输入时的性能。

从表9可见,图像增强对下游任务的影响取决于数据集、骨干网络和评价指标,视觉质量提升不必然带来分割性能改善(Lin等,2026)。表9中多个方法在语义分割上的表现差异明显。在SUIM数据集上,UDCP和UIBLA在三个骨干网络下的mIoU均低于Raw,WF-Diff的结果总体接近Raw;DTI-UIE的mIoU在UNet、DeepLabv3和FPN上分别较Raw提高1.09、1.10和1.48个百分点。在LIACi数据集上,部分方法虽能提升特定骨干网络下的Dice,但mIoU并未同步改善;FPN下的mIoU最高值为53.72%,仅较Raw高0.06个百分点,表明增强收益具有明显的任务和结构依赖性。

在实例分割任务中,在Water Mask R-CNN上,
© 中国图象图形学报版权所有

表 9 不同图像增强方法在语义分割与实例分割任务上的性能比较

Table 9 Performance Comparison of Different Image Enhancement Methods on Semantic and Instance Segmentation Tasks

任务	数据集	Backbone	Metric	Raw	UDCP (Drews 等, 2013)	UIBLA (Peng 和 Cosman, 2017)	WF- Diff (Zhao 等, 2024a)	DTI- UIE (Lin 等, 2026)
语 义 分 割	SUIM 数据集 (Islam 等, 2020a)	UNet (Ronneberger 等, 2015)	Dice	72.29	69.58	72.09	72.37	74.48
			mIoU	66.52	62.12	61.67	66.09	67.61
		DeepLabv3 (Chen 等, 2017)	Dice	78.02	77.45	78.1	77.62	78.93
			mIoU	66.51	62.28	61.81	64.84	67.61
		FPN (Seferbekov 等, 2018)	Dice	74.96	73.84	76.27	76.12	76.85
			mIoU	64.74	60.7	62.71	64.58	66.22
	LIACi 数据集 (Waszak 等, 2022)	UNet (Ronneberger 等, 2015)	Dice	75.55	76.8	76.43	77.85	78.02
			mIoU	54.43	52.21	49.18	54.15	55.27
		DeepLabv3 (Chen 等, 2017)	Dice	82.51	80.64	79.75	82.67	83.13
			mIoU	55.35	53.87	50.25	55.22	55.47
		FPN (Seferbekov 等, 2018)	Dice	81.63	80.44	78.74	79.81	81.68
			mIoU	53.66	50.35	45.89	50.95	53.72
实 例 分 割	UIIS 数据集 (Lian 等, 2023)	Water Mask R-CNN (Lian 等, 2023)	mAPs	8.3	6.7	6.3	7.1	7.9
			mAPm	19.0	17.1	17.2	18.6	19.2
			mAPl	34.4	29.6	29.6	32.2	33.8
			mAP50	40.0	36.5	35.9	38.5	40.1
			mAP75	23.9	19.5	20.7	22.4	24.3
			mAP	23.3	20.0	20.1	21.9	23.6

注:加粗数据表示同一任务、数据集、骨干网络和评价指标下的最优结果。

DTI-UIE 的 mAP 达到 23.6%,表现最优。而 UDCP 和 UIBLA 在部分指标上甚至低于 Raw 输入,说明不当增强会破坏目标边缘与结构信息,反而损害实例级识别性能。因此,增强方法应结合不同目标尺度和下游任务进行评价,不能仅依据图像的主观观感判断其有效性。

3.9 跨域泛化性能分析

水下图像增强方法从实验室评测走向真实部署时,跨域泛化是最突出的瓶颈之一。不同数据集在采集设备、水体类型、光照条件和退化程度上存在显著差异,同一模型在源域训练后,在未见过的目标域上性能往往大幅下滑。表 10 在 5 个真实水下基准 (Test-R1000、SQUID、UIEB、EUVP、UFO-120) 上的跨域跨数据集评测结果清晰地揭示了这一问题 (Wang 等, 2023)。

从表 10 可以看出,多数方法在不同数据集上的表现存在明显波动。以 UGAN 为例,其 UCIQE 在五个数据集上介于 0.584 至 0.624 之间,UIQM 从 2.780 到 3.237 不等。这种“一域一表现”的现象说明,现有方法对训练数据分布的统计特征仍较为敏感,尚未学到真正与水体类型无关的退化不变表征。

值得注意的是,不同范式的跨域稳定性存在系统性差异。传统先验方法 (如 HE-Prior) 的 UCIQE 在某些数据集上异常偏高 (如 Test-R1000 上达 0.673),但其对应的感知分数却极低,说明其高分高度依赖特定统计假设、泛化基础薄弱。相比之下,引入域适配机制的 TUDA 在五个数据集上的 UCIQE 为 0.589~0.608,波动范围较小;其 UIQM 平均值为 3.200,为表中最高,但在 SQUID 上的 UIQM 仅为 2.643,说明跨域稳定性仍具有指标和数据集依

表 10 五个真实水下基准上不同方法的跨域泛化结果(UCIQE 和 UIQM)

Table 10 Cross-Domain Generalization Results of Different Methods on Five Real-World Underwater Benchmarks (UCIQE and UIQM)												
方法	UCIQE ↑						UIQM ↑					
	Test-R1000	SQUID	UIEB	EUVP	UFO-120	Avg	Test-R1000	SQUID	UIEB	EUVP	UFO-120	Avg
Fusion-12(Ancuti 等,2012)	0.607	0.561	0.612	0.610	0.619	0.613	2.729	1.507	2.870	2.779	2.769	2.769
Fusion-18(Ancuti 等,2018)	0.583	0.582	0.589	0.588	0.596	0.591	2.879	2.268	3.030	3.153	2.987	3.016
HE-Prior(Li 等,2016)	0.673	0.582	0.678	0.673	0.678	0.675	—	—	2.637	2.656	2.565	2.604
UIBLA(Peng 和 Cosman,2017)	0.581	0.500	0.595	0.594	0.624	0.605	2.030	—	2.262	2.053	2.061	2.183
FUnIE-GAN(Islam 等,2020b)	0.548	0.488	0.560	0.556	0.582	0.566	2.777	1.791	2.949	3.079	2.954	2.952
UGAN(Fabbri 等,2018)	0.611	0.584	0.619	0.604	0.624	0.615	3.130	2.780	3.180	3.237	3.152	3.172
TUDA(Wang 等,2023)	0.596	0.589	0.605	0.596	0.608	0.602	3.125	2.643	3.203	3.313	3.166	3.200

注:加粗数据表示同一数据集、同一指标列中的最优结果;UCIQE 和 UIQM 越大越好。

赖性。

表 11。表中链接直接对应作者或项目公开仓库,读者可据此获取实现与配置文件。

为便于复现实验结果,本文将文中涉及且能够获得公开代码的代表性方法及其代码仓库汇总于

表 11 文中代表性水下图像增强方法及其开源代码仓库

Table 11 Representative Underwater Image Enhancement Methods and Their Open-Source Code Repositories		
范式	方法	开源代码仓库
GAN	FUnIE-GAN(Islam 等,2020b)	https://github.com/xahidbuffon/funie-gan
GAN	WaterGAN(Li 等,2018)	https://github.com/kskin/WaterGAN
GAN	UGAN(Fabbri 等,2018)	https://github.com/cameronfabbri/Underwater-Color-Correction
GAN	UWGAN(Wang 等,2019)	https://github.com/infrontofme/UWGAN_UIE
GAN	PUGAN(Cong 等,2023)	https://github.com/rmcong/PUGAN_TIP2023
GAN	MLFcGAN(Liu 等,2019)	https://github.com/Xiaodong-Bran/MLFcGAN
Transformer	Phaseformer(Khan 等,2025)	https://github.com/Mdraqibkhan/Phaseformer
Transformer	UWFormer(Chen 等,2024)	https://github.com/leiyinting/UWFormer
Transformer	Histoformer(Peng 等,2025)	https://github.com/sunshangquan/Histoformer
Transformer	Spectroformer(Khan 等,2024)	https://github.com/Mdraqibkhan/Spectroformer
Transformer	Uformer(Wang 等,2022)	https://github.com/ZhendongWang6/Uformer
Transformer	Restormer(Zamir 等,2022)	https://github.com/swz30/Restormer
Transformer	WaterMamba(Guan 等,2024a)	https://github.com/Guan-MS/WaterMamba
Diffusion	SLURPP(Wu 等,2025)	https://github.com/kongdai123/SLURPP
Diffusion	UW-DiffPhys(Bach 等,2024)	https://github.com/bachzz/UW-DiffPhys
Diffusion	CLIP-UIE(Liu 等,2024)	https://github.com/OUCVisionGroup/CLIP-UIE
Diffusion	PA-Diff(Zhao 等,2024b)	https://github.com/chenydong/PA-Diff
Diffusion	SeaDiff(Bi 等,2025)	https://github.com/Henry-Bi/SeaDiff
Diffusion	DiffWater(Guan 等,2024b)	https://github.com/Guan-MS/DiffWater

4 挑战与未来研究方向

4.1 挑战

1) 复合退化建模难: 水下图像退化通常由选择性吸收、前向散射、后向散射、悬浮颗粒噪声、运动模糊和非均匀照明共同造成, 不同退化因素之间存在显著耦合关系。现有方法通常将退化过程简化为颜色偏移、对比度下降或雾化效应等单一问题, 但真实水下场景通常同时包含浑浊、低照度、局部强光和动态目标, 使得单一先验或单分支网络难以稳定适配。

2) 跨域泛化难: 跨域泛化是 UIE 走向真实应用时最突出的挑战之一。不同海域、水深、设备、光源和拍摄距离会导致明显的分布差异, 而多数公开数据集仍存在规模有限、场景覆盖不足和配对标注困难等问题。因此, 许多方法在同域测试中表现良好, 在跨水型或跨设备测试中可能出现颜色漂移、局部过增强或任务收益下降。

3) 生成过程可控与可信难: 生成式网络在提升纹理恢复能力与视觉自然性的同时, 也引入可控性与可信性问题。GAN 方法可能在复杂纹理区域生成伪细节, Transformer 方法可能在全局依赖建模中放大噪声或引入颜色偏移, Diffusion 方法虽然具备较高的高保真恢复上限, 但少步或加速采样可能带来欠恢复与细节漂移风险。

4) 轻量化部署受约束: 实际水下应用通常受到算力、功耗、通信带宽和实时性的共同限制。水下机器人、潜航器和边缘摄像设备承载大规模 Transformer 或多步 Diffusion 模型的能力有限, 而离线科研分析又对高保真恢复和细节完整性提出更高要求, 因此如何在恢复质量与部署成本之间取得平衡仍然困难。此外, 从方法层面看, 这一问题并不只意味着缩小网络规模, 更需要将模型压缩、知识蒸馏、量化和低秩/共享权重设计纳入 UIE 的统一优化框架, 使高质量恢复能力迁移到可在边缘设备运行的学生模型。

5) 多模态先验的安全可靠性不足: 随着视觉语言模型、提示学习和大规模生成模型的发展, 多模态先验正在成为 UIE 的潜在新变量。文本提示、语义类别和任务描述可以为增强过程提供更高层约束, 但基础模型也可能带来提示敏感性、跨域偏置和算力成本问题。若过度依赖语义先验, 模型可能在科

学观测场景中生成看似合理但缺乏物理一致性的纹理或结构, 进而影响后续识别与测量结果。

4.2 未来研究方向

1) 面向复合退化的建模优化。后续研究可将水体类型、照明条件、散射强度、噪声分布和成像设备参数显式编码为条件变量, 并与无条件增强模型、仅退化编码模型和物理约束模型进行对照。可检验的假设为: 显式退化建模能够在重浑浊、低照度和混合噪声场景中同步提升 PSNR、SSIM 及下游检测、分割等任务表现, 并降低颜色过补偿与局部结构失真的发生率。

2) 面向跨域泛化的数据构建。建议构建覆盖多水型、多光照、多深度、多设备的真实数据集, 并预留按海域、相机和分辨率划分的跨域测试协议。可检验的方案路线包括单数据集监督训练、跨数据集联合训练和域自适应训练; 若多源训练结合风格不变表征或伪标签约束能够显著提高跨域测试中的多个相关指标, 就可以验证其泛化价值。

3) 面向可控可信的生成约束。后续工作可将增强强度、物理参数或文本提示作为可控输入, 并与标准单次前向生成和少步扩散采样进行对比。可检验的假设为: 加入不确定性估计或置信度图后, 模型可以在增强强度、伪纹理风险和结果稳定性之间形成更可控的权衡, 同时保持可接受的运行时延。

4) 面向部署需求的轻量化设计。建议将全量模型、蒸馏模型、剪枝模型以及“快速增强 + 细修复”的两阶段架构作为基线进行比较, 并在相同输入分辨率下报告参数量、FLOPs、显存占用、推理延迟和能耗。可检验的假设是: 经过蒸馏、剪枝或分层推理后的轻量模型, 能够在较低计算成本下保持与全量模型接近的质量上限, 尤其适用于实时导航和边缘监测场景。进一步地, 建议把量化感知训练、低比特推理和分层蒸馏与剪枝、分层推理一起作为统一基线, 并同时报告多维度结果, 以避免仅凭模型大小判断轻量化效果。

5) 面向多模态先验的可靠融合。后续研究可将语义提示、任务反馈和物理一致性约束统一到可验证的生成流程中, 并与仅使用视觉先验的模型进行对照。可以尝试受物理约束的多模态融合方案, 在不显著增加幻觉风险的前提下, 提高颜色校正、局部对比度和结构保持一致性, 并改善检测、分割等下游任务表现。

6)在数据集构建与基准评测。未来工作应构建包含明确退化等级(如浑浊度、光照非均匀程度)标注的细粒度测试集,并制定统一的跨退化等级评测协议,从而使失效模式分析过渡到可量化的鲁棒性评估。评测时应同时报告 PSNR 等全参考指标,UCIQE 等无参考指标,以及运行时延、显存、FLOPs 和至少一项下游任务结果,以避免单一指标主导结论。

5 结 论

本文围绕生成式网络在水下图像增强中的应用进行了系统性综述,旨在为相关研究人员提供方法分类、实验设计与工程部署方面的理论参考与实践指导。具体为,本文从水下成像退化机理出发,分析了选择性吸收、散射和非均匀照明等因素对图像质量的耦合影响,阐明了水下图像增强问题的不适宜性和场景依赖性。为厘清不同技术路线的演进逻辑,本文围绕 GAN、Transformer 和 Diffusion 三类生成范式,系统梳理了各自的核心框架、生成机制与监督方式,并对比了它们在视觉真实性、全局一致性和高保真恢复方面的优势与局限。在此基础上,本文结合 UIEB、LSUI、U45 等公开数据集,对代表性方法的重建质量、感知质量和模型复杂度进行了定量与定性比较,指出 PSNR、SSIM 与 UIQM、UCIQE 等指标在评价目标上存在显著差异,单一指标难以全面反映水下图像增强的实际效果。最后,本文归纳了生成式 UIE 方法面临的主要挑战,并展望了复合退化建模、跨域泛化、轻量化部署和可信可控生成等未来研究方向。

参考文献(References)

- Ancuti C, Ancuti C O, Haber T and Bekaert P. 2012. Enhancing underwater images and videos by fusion//2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Providence: IEEE: 81-88.
- Ancuti C O, Ancuti C, De Vleeschouwer C and Bekaert P. 2018. Color balance and fusion for underwater image enhancement. *IEEE Transactions on Image Processing*, 27(1): 379-393 [DOI: 10.1109/TIP.2017.2759252].
- Anwar S and Li C. 2020. Diving deeper into underwater image enhancement: A survey. *Signal Processing: Image Communication*, 89: 115978.
- Arjovsky M, Chintala S and Bottou L. 2017. Wasserstein generative adversarial Networks//Proceedings of the 34th International Conference on Machine Learning, vol. 70 (PMLR). Sydney: PMLR: 214-223.
- Bach N G, Tran C M, Kamioka E and Tan P X. 2024. Underwater image Enhancement with physical-based denoising diffusion implicit models[EB/OL]. arXiv preprint arXiv:2409.18476.
- Bi H, Chen L, Cao J, Wang J, Sun J, Rao Y, et al. 2025. SeaDiff: Underwater image enhancement with degradation-aware diffusion model. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(12): 12212-12226.
- Chen H, Wang Z, Qin H and Mu X. 2026. Diffusion-based frequency degradation prior fusion with hierarchical wavelet decompositions for underwater image enhancement. *Information Fusion*, 104394.
- Chen L, Ding D D. 2022. Joint multi-residual learning for underwater image enhancement. *Journal of Image and Graphics*, 27(5): 1577-1588 (陈龙, 丁丹丹. 2022. 多残差联合学习的水下图像增强. *中国图象图形学报*, 27(5): 1577-1588) [DOI: 10.11834/jig.210041].
- Chen L C, Papandreou G, Schroff F and Adam H. 2017. Rethinking atrous convolution for semantic image segmentation[EB/OL]. arXiv preprint arXiv:1706.05587.
- Chen S Z, Ren Z G and Lian Q S. 2016. Single Image Dehazing Algorithm Based on Improved Dark Channel Prior and Guided Filter. *Acta Automatica Sinica*, 42(3): 455-465 (陈书贞, 任占广, 练秋生. 2016. 基于改进暗通道和导向滤波的单幅图像去雾算法. *自动化学报*, 42(3): 455-465) [DOI: 10.16383/j. aas. 2016. c150212].
- Chen W W, Lei Y T, Luo S H, Zhou Z Y, Li M X and Pun C M. 2024. UWFormer: Underwater Image Enhancement via a Semi-Supervised Multi-Scale Transformer//Proceedings of the 2024 International Joint Conference on Neural Networks (IJCNN). Yokohama: IEEE: 1-8 [DOI: 10.1109/IJCNN60899.2024.10651366].
- Cong R, Yang W, Zhang W, Li C, Guo C L, Huang Q, et al. 2023. PUGAN: Physical model-guided underwater image enhancement using GAN with dual-discriminators. *IEEE Transactions on Image Processing*, 32: 4472-4485.
- Cong X, Zhao Y, Gui J, Hou J and Tao D. 2024. A comprehensive survey on underwater image enhancement based on deep learning[EB/OL]. arXiv preprint arXiv:2405.19684.
- Ding X, Chen X, Sui Y, Wang Y and Zhang J. 2025. Underwater image Enhancement using a diffusion model with adversarial learning. *Journal of Imaging*, 11(7): 212.
- Dong C, Zhao C, Cai W and Yang B. 2024. O-mamba: O-shape state-space model For underwater image enhancement [EB/OL]. arXiv preprint arXiv:2408.12816.
- Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. 2021. An image is worth 16x16 words: Transformer

- ers for Image recognition at scale//International Conference on Learning Representations (ICLR). [Online]: OpenReview.net.
- Dreus P L J, Nascimento E R, Moraes F M, Botelho S S C and Campos M F M. 2013. Transmission estimation in underwater single images//2013 IEEE International Conference on Computer Vision Workshops (ICCVW). Sydney: IEEE:825-830.
- Fabbri C, Islam M J and Sattar J. 2018. Enhancing underwater imagery using Generative adversarial networks//2018 IEEE International Conference on Robotics and Automation (ICRA). Brisbane: IEEE: 7159-7165.
- Fang Y, Li Q and Wang K. 2025. Multi-scale diffusion model for underwater image restoration and enhancement. PLOS ONE, 20 (9): e0331465.
- Fathy M E, Mohamed S A, Awad M I and Munim H E A E. 2023. Submergestylegan: Synthetic underwater data generation with style transfer for domain adaptation//2023 International Conference on Digital Image Computing: Techniques and Applications (DICTA). Port Macquarie: IEEE:546-553.
- Fu X, Zhuang P, Huang Y, Liao Y, Zhang X P and Ding X. 2014. A retinex-based enhancing approach for single underwater image//2014 IEEE International Conference on Image Processing (ICIP). Paris: IEEE:4572-4576.
- Fu X Y and Cao X Y. 2020. Underwater image enhancement with global - local networks and compressed-histogram equalization. Signal Processing: Image Communication, 86: 115892 [DOI: 10.1016/j.image.2020.115892].
- Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. 2014. Generative adversarial nets//Advances in Neural Information Processing Systems, vol. 27. Montreal: Curran Associates:2672-2680.
- Guan M, Xu H, Jiang G, Yu M, Chen Y, Luo T, et al. 2024a. DM-water: Visual state space model for underwater image enhancement [EB/OL]. arXiv preprint arXiv:2405.08419.
- Guan M, Xu H, Jiang G, Yu M, Chen Y, Luo T, et al. 2024b. 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.
- Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V and Courville A. 2017. Improved training of wasserstein Gans//Advances in Neural Information Processing Systems, vol. 30. Long Beach: Curran Associates:5769-5779.
- Guo X J, Chen X H, Wang S Q and Pun C M. 2025. Underwater image restoration through a prior guided hybrid sense approach and extensive benchmark analysis. IEEE Transactions on Circuits and Systems for Video Technology, 35(5):4784-4800.
- Guo Y J, Wu Q, Yuan J J, Hou J C and Lv W H. 2021. Research Progress on Underwater Optical Image Processing. Journal of Electronics & Information Technology, 43(2): 426-435 (郭银景, 吴琪, 苑娇娇, 侯佳辰, 吕文红. 2021. 水下光学图像处理研究进展. 电子与信息学报, 43 (2): 426-435) [DOI: 10.11999/JEIT190803].
- He K, Sun J and Tang X. 2011. Single image haze removal using dark channel prior. IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(12): 2341-2353.
- Ho J, Jain A and Abbeel P. 2020. Denoising diffusion probabilistic models//Advances in Neural Information Processing Systems, vol. 33.
- Hu W, Chen S, Luo T, Zhang L, Zhang H, Liu Z, et al. 2025. DATDM: Dynamic attention transformer diffusion model for Underwater image enhancement. Alexandria Engineering Journal, 126: 591-604.
- Huang H, Hu K Y, Guo L, Wang H F and Zhu L Y. 2021. Improved defogging algorithm for sea fog. Journal of Harbin Institute of Technology, 53(8): 81-91 (黄鹤, 胡凯益, 郭璐, 王会峰, 朱礼亚. 2021. 改进的海雾图像去除方法. 哈尔滨工业大学学报, 53(8): 81-91).
- Huang X and Belongie S. 2017. Arbitrary style transfer in real-time with adaptive instance Normalization//2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE:1501-1510.
- Islam M J, Edge C, Xiao Y, Luo P, Mehtaz M, Morse C, et al. 2020a. Semantic segmentation of underwater imagery: Dataset and benchmark//2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Las Vegas: IEEE:1769-1776.
- Islam M J, Xia Y and Sattar J. 2020b. Fast underwater image enhancement for Improved visual perception. IEEE Robotics and Automation Letters, 5 (2): 3227-3234 [DOI: 10.1109/LRA. 2020. 2974710].
- Isola P, Zhu J Y, Zhou T and Efros A A. 2017. Image-to-image translation With conditional adversarial networks//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE:1125-1134.
- Johnson J, Alahi A and Li F F. 2016. Perceptual losses for real-time style transfer and super-resolution//Computer Vision - ECCV 2016. Amsterdam: Springer:694-711.
- Karras T, Aittala M, Laine S, Harkonen E, Hellsten J, Lehtinen J, et al. 2021. Alias-free generative adversarial networks//Advances in Neural Information Processing Systems, vol. 34. [Online]: Curran Associates:852-863.
- Karras T, Laine S, Aittala M, Hellsten J, Lehtinen J and Aila T. 2020. Analyzing and improving the image quality of StyleGAN//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). [Online]: IEEE:8110-8119.
- Karras T, Laine S and Aila T. 2019. A style-based generator architecture for Generative adversarial networks//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE:4401-4410.
- Kawar B, Elad M, Ermon S and Song J. 2022. Denoising diffusion Restoration models//Advances in Neural Information Processing Sys-

- tems, vol. 35. New Orleans: NeurIPS Foundation:23593-23606.
- Khan M R, Mishra P, Mehta N, Phutke S S, Vipparthi S K, Nandi S, et al. 2024. Spectroformer: Multi-domain query Cascaded transformer network for underwater image enhancement//2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa: IEEE:1454-1463.
- Khan M R, Negi A, Kulkarni A, Phutke S S, Vipparthi S K and Murala S. 2025. Phaseformer: Phase-based attention mechanism for Underwater image restoration and beyond//2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Tucson: IEEE:9618-9629.
- Li C Y, Guo J C, Cong R M, Pang Y W and Wang B. 2016. Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior. IEEE Transactions on Image Processing, 25(12): 5664-5677 [DOI: 10.1109/TIP.2016.2612882].
- Li C, Anwar S and Porikli F. 2020a. Underwater scene prior inspired deep underwater image and video enhancement. Pattern Recognition, 98: 107038.
- Li C, Anwar S, Hou J, Cong R, Guo C and Ren W. 2021. Underwater Image Enhancement via Medium Transmission-Guided Multi-Color Space Embedding. IEEE Transactions on Image Processing, 30: 4985-5000.
- Li C, Guo C, Ren W, Cong R, Hou J, Kwong S, et al. 2020b. An underwater image enhancement benchmark dataset and beyond. IEEE Transactions on Image Processing, 29: 4376-4389.
- Li J, Skinner K A, Eustice R M and Johnson-Roberson M. 2018. WaterGAN: Unsupervised generative network to enable real-time color correction of monocular Underwater images. IEEE Robotics and Automation Letters, 3(1): 387-394.
- Li Y, Mi Z, Wang Y, Jiang S and Fu X. 2025. TAFormer: A Transmission-aware transformer for underwater image enhancement. IEEE Transactions on Circuits and Systems for Video Technology, 35(1): 601-616.
- Liang X, Liu X, Liu Z and Yu H. 2026. CGFormer: A color-guided dual-branch network for underwater image enhancement. Journal of Visual Communication and Image Representation, 119: 104866 [DOI: 10.1016/j.jvcir.2026.104866].
- Lian S, Li H, Cong R, Li S, Zhang W and Kwong S. 2023. Watermask: Instance segmentation for underwater imagery//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris: IEEE:1305-1315.
- Lin B S, Gao F, Yu Y W, Dong J Y and Du Q. 2026. Downstream task inspired underwater image enhancement: a perception-aware study from dataset construction to network design. IEEE Transactions on Image Processing.
- Liu H B, Yang J, Wu Z P, Zhang Q N and Deng Y. 2015. A Fast Single Image Dehazing Method Based on Dark Channel Prior and Retinex Theory. Acta Automatica Sinica, 41(7): 1264-1273 (刘海波, 杨杰, 吴正平, 张庆年, 邓勇. 2015. 基于暗通道先验和Retinex理论的快速单幅图像去雾方法. 自动化学报, 41(7): 1264-1273) [DOI: 10.16383/j.aas.2015.c140748].
- Liu R, Fan X, Zhu M, Hou M and Luo Z. 2020. Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light. IEEE Transactions on Circuits and Systems for Video Technology, 30(12): 4861-4875.
- Liu S, Li K, Ding Y and Qi Q. 2024. Underwater image enhancement by Diffusion model with customized CLIP-classifier [EB/OL]. arXiv preprint arXiv:2405.16214.
- Liu X, Gao Z and Chen B M. 2019. MLFCGAN: Multilevel feature fusion-based conditional GAN for underwater image color correction. IEEE Geoscience and Remote Sensing Letters, 17(9): 1488-1492.
- Liu X, Xu H, Ju Y, Wang S, Liu C and Chen L. 2026. DA-GAN: Dual-attention GAN for underwater image enhancement with contrast and color Correction. IEEE Transactions on Geoscience and Remote Sensing.
- Liu Y, Jiang Q, Wang X, Luo T and Zhou J. 2024. Underwater image enhancement with cascaded contrastive learning. IEEE Transactions on Multimedia, 27:1512-1525.
- Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. 2021. Swin-transformer: Hierarchical vision transformer using shifted windows//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE:10012-10022.
- Luo H, Wen J, Li Z, Chen D, Xi M, He J and Yang J. 2026. A review of underwater image enhancement and restoration technology. Journal of Image and Graphics, 31(5): 1324-1349 (罗虎, 温家宝, 李政键, 陈德盛, 奚萌, 何景逸, 杨嘉琛. 2026. 水下图像增强与复原技术综述. 中国图象图形学报, 31(5): 1324-1349) [DOI: 10.11834/jig.250479].
- Long K, Xie G, Ma L, Li Q, Huang M, Lv J, et al. 2025. Enhancing multimodal learning via hierarchical fusion architecture search with Inconsistency mitigation. IEEE Transactions on Image Processing.
- Ma L, Huang M, Yang S, Wang R and Wang X. 2022. An adaptive Localized decision variable analysis approach to large-scale multi-objective and many-Objective optimization. IEEE Transactions on Cybernetics, 52(7): 6684-6696.
- Ma L, Li N, Yu G, Geng X, Cheng S, Wang X, et al. 2024b. Pareto-wise ranking classifier for multiobjective evolutionary neural Architecture search. IEEE Transactions on Evolutionary Computation, 28(3): 570-581.
- Ma L, Li N, Zhu P, Tang K, Khan A, Wang F, et al. 2024a. A novel Fuzzy neural network architecture search framework for defect recognition with Uncertainties. IEEE Transactions on Fuzzy Systems, 32(5): 3274-3285.
- Ma L, Zhou Y, Ma Y, Yu G, Li Q and He Q. 2025. Defying multi-model Forgetting in one-shot neural architecture search using orthogonal gradient learning. IEEE Transactions on Computers, 74(5): 1678-1689.

- Mehta S and Rastegari M. 2022. MobileViT: Light-weight, general-purpose, and Mobile-friendly vision transformer//International Conference on Learning Representations (ICLR). [Online]: OpenReview.net.
- Mirza M and Osindero S. 2014. Conditional generative adversarial nets [EB/OL]. arXiv preprint arXiv:1411.1784.
- Mishra P, Mehta N, Vipparthi S K and Murala S. 2025. USWformer: Efficient sparse wavelet transformer for underwater image enhancement//2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Tucson: IEEE:3372-3382.
- Naik A, Swarnakar A and Mittal K. 2021. Shallow-UWnet: Compressed model for underwater image enhancement [EB/OL]. arXiv preprint arXiv:2101.02073.
- Palazari S and Dumic E. 2026. Towards reliable evaluation of underwater image enhancement using subjective and objective analysis. *Electronics*, 15(11): 2412 [DOI: 10.3390/electronics15112412].
- Panetta K, Gao C and Agaian S. 2015. Human-visual-system-inspired underwater image quality measures. *IEEE Journal of Oceanic Engineering*, 41(3): 541-551.
- Peng L, Zhu C and Bian L. 2023. U-shape transformer for underwater image enhancement. *IEEE Transactions on Image Processing*, 32: 3066-3079.
- Peng Y T and Cosman P C. 2017. Underwater image restoration based on image blurriness and light absorption. *IEEE Transactions on Image Processing*, 26(4): 1579-1594.
- Peng Y T, Chen Y R, Chen G R and Liao C J. 2025. Histoformer: Histogram-based transformer for efficient underwater image enhancement. *IEEE Journal of Oceanic Engineering*, 50 (1): 164-177.
- Qing Y, Liu S, Wang H and Wang Y. 2025. DiffUIE: Learning latent global priors in Diffusion models for underwater image enhancement. *IEEE Transactions on Multimedia*, 27: 2516-2529.
- Rombach R, Blattmann A, Lorenz D, Esser P and Ommer B. 2022. High-resolution image synthesis with latent diffusion models//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE:10684-10695.
- Ronneberger O, Fischer P and Brox T. 2015. U-net: Convolutional networks for biomedical image segmentation//Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015. Munich: Springer:234-241.
- Saharia C, Ho J, Chan W, Salimans T, Fleet D J and Norouzi M. 2023. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4): 4713-4726.
- Saleh A, Sheaves M, Jerry D and Azghadi M R. 2025. Adaptive deep learning framework for robust unsupervised underwater image enhancement. *Expert Systems with Applications*, 268: 126314.
- Saoud L S, Elmezain M, Sultan A, Abdelwahab M H H, Seneviratne L D and Hussain I. 2024. Seeing through the haze: A comprehensive review of underwater image enhancement techniques. *IEEE Access*, 12: 145206-145233.
- Seferbekov S, Iglovikov V, Buslaev A and Shvets A. 2018. Feature pyramid network for multi-class land segmentation//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Salt Lake City: IEEE:272-2723.
- Shang J, Li Y, Hu X and Yuan J. 2025. LGT: Luminance-guided transformer-based multi-feature fusion network for underwater image enhancement. *Information Fusion*, 118: 102977.
- Song J, Meng C and Ermon S. 2021a. Denoising diffusion implicit models//International Conference on Learning Representations (ICLR). [Online]: OpenReview.net.
- Song J, Xu H, Jiang G, Yu M, Chen Y, Luo T, et al. 2025. Frequency domain-based latent diffusion model for underwater image enhancement. *Pattern Recognition*, 160: 111198.
- Song Y, Sohl-Dickstein J, Kingma D P, Kumar A, Ermon S and Poole B. 2021b. Score-based generative modeling through stochastic differential Equations//International Conference on Learning Representations (ICLR). [Online]: OpenReview.net.
- Sun B, Mei Y, Yan N and Chen Y. 2023. UMGAN: Underwater image enhancement network for unpaired image-to-image translation. *Journal of Marine Science and Engineering*, 11(2): 447.
- Tang Y, Iwaguchi T and Kawasaki H. 2023. Underwater image enhancement by Transformer-based diffusion model with non-uniform sampling for skip strategy//Proceedings of the 31st ACM International Conference on Multimedia. Ottawa: ACM:5419-5427.
- Tian C, Cheng T, Ma Y, Zhu Q, Zhang B and Zhang D. 2026. A Heterogeneous CNN With Transformers for Image Denoising. *IEEE Transactions on Consumer Electronics*, 72(2).
- Tian C, Cheng T, Peng Z, et al. 2025a. A survey on deep learning fundamentals. *Artificial Intelligence Review*, 58(12): 381.
- Tian C W and Zuo W M. 2024. Deep Learning and Image Restoration. Publishing House of Electronics Industry (田春伟, 左旺孟. 2024. 深度学习与图像复原. 电子工业出版社).
- Tian C W, Song M J, Zuo W M, et al. 2025. Application of convolutional neural networks in image super-resolution. *CAAI Transactions on Intelligent Systems*, 20(3): 719-749 (田春伟, 宋明键, 左旺孟, 等. 2025. 卷积神经网络在图像超分辨率上的应用. 智能系统学报, 20(3): 719-749 [DOI: 10.11992/tis.202409027]).
- Tian C, Zheng M, Zuo W, Zhang S, Zhang Y and Lin C W. 2024. A cross Transformer for image denoising. *Information Fusion*, 102: 102043.
- Tian X, Lei Y, Zhang X, Li Z, Pun C M and Chen X. 2025b. SFormer: SNR-guided transformer for underwater image enhancement from the frequency domain [EB/OL]. arXiv preprint arXiv:2508.18664.
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, et al. 2017. Attention is all you need//Advances in Neural Information Processing Systems, vol. 30. Long Beach: Curran Associates: 6000-6010.
- Wang N, Zhou Y, Han F, Zhu H and Yao J. 2019. UWGAN: Underwa-

- ter GAN for real-world underwater color restoration and dehazing [EB/OL]. arXiv preprint arXiv:1912.10269.
- Wang Z, Bovik A C, Sheikh H R and Simoncelli E P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600-612.
- Wang Z, Cun X, Bao J, Zhou W, Liu J and Li H. 2022. Uformer: A general u-shaped transformer for image restoration//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE:17683-17693.
- Wang Z M. 2015. Review of No-reference Image Quality Assessment. *Acta Automatica Sinica*, 41(6): 1062-1079 (王志明. 2015. 无参考图像质量评价综述. *自动化学报*, 41(6): 1062-1079) [DOI: 10.16383/j.aas.2015.c140404].
- Wang Z, Shen L, Xu M, Yu M, Wang K and Lin Y. 2023. Domain Adaptation for underwater image enhancement. *IEEE Transactions on Image Processing*, 32: 1442-1457.
- Waszak M, Cardaillac A, Elvæsæter B, Rødølen F and Ludvigsen M. 2022. Semantic segmentation in underwater ship inspections: Benchmark and data set. *IEEE Journal of Oceanic Engineering*, 48(2): 462-473.
- Wu D and Zhu Q S. 2015. The Latest Research Progress of Image Dehazing. *Acta Automatica Sinica*, 41(2): 221-239 (吴迪, 朱青松. 2015. 图像去雾的最新研究进展. *自动化学报*, 41(2): 221-239) [DOI: 10.16383/j.aas.2015.c131137].
- Wu J, Wang T, Siddique M A B, Islam M J, Fermuller C, Aloimonos Y, et al. 2025. Single-step latent diffusion for underwater image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Xiao J, Fan Z H, Li D, Fu X Y and Zha Z J. 2025. Non-causal selective state space model for image restoration. *Journal of Image and Graphics*, 30(10): 3173-3186 (肖杰, 范子豪, 李东, 傅雪阳, 查正军. 2025. 面向图像复原的非因果选择性状态空间模型. *图像与图形学报*, 30(10): 3173-3186).
- Xiao J, Fu X Y, Zhu Y R and Zha Z J. 2026. Bayesian Window Transformer for Image Restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(3): 2431-2444 [DOI: 10.1109/TPAMI.2025.3626947].
- Yang M and Sowmya A. 2015. An underwater color image quality evaluation metric. *IEEE Transactions on Image Processing*, 24(12): 6062-6071.
- Yao L, Yan J, Chen W, Feng H and Chen H. 2025. UWLDLM: Underwater image enhancement based on latent diffusion model//2025 40th Youth Academic Annual Conference of Chinese Association of Automation (YAC). Zhengzhou: IEEE:352-357.
- You K, Li X H, Yi P F, Zhang Y H, Xu J K, Ren J J, Bai H Y and Ma C H. 2025. PIC-GAN: Symmetry-driven underwater image enhancement with partial instance normalisation and colour detail modulation. *Symmetry*, 17(2): 201 [DOI: 10.3390/sym17020201].
- Yu J, Li D P and Liao Q M. 2011. Physics-based Fast Single Image Fog Removal. *Acta Automatica Sinica*, 37(2): 143-149 (禹晶, 李大鹏, 廖庆敏. 2011. 基于物理模型的快速单幅图像去雾方法. *自动化学报*, 37(2): 143-149) [DOI: 10.3724/SP.J.1004.2011.00143].
- Zamir S W, Arora A, Khan S, Hayat M, Khan F S and Yang M H. 2022. Restormer: Efficient transformer for high-resolution image restoration//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE:5728-5739.
- Zhang F, Zhang Z, Zhang F and Tian X. 2026. Physics-Guided Temporal Underwater Image Enhancement Using Implicit Neural Representations and Diffusion Models. *Journal of Marine Science and Engineering*, 14(9): 798 [DOI: 10.3390/jmse14090798].
- Zhang R, Isola P, Efros A A, Shechtman E and Wang O. 2018. The unreasonable effectiveness of deep features as a perceptual metric//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City: IEEE:586-595.
- Zhang X L, Li X F and Li J. 2014. Validation and Correlation Analysis of Metrics for Evaluating Performance of Image Fusion. *Acta Automatica Sinica*, 40(2): 306-315 (张小利, 李雄飞, 李军. 2014. 融合图像质量评价指标的相关性分析及性能评估. *自动化学报*, 40(2): 306-315) [DOI: 10.3724/SP.J.1004.2014.00306].
- Zhang Y, Yuan J and Cai Z. 2025. DCGF: Diffusion-color-guided framework for underwater image enhancement. *IEEE Transactions on Geoscience and Remote Sensing*, 63: 1-12.
- Zhang Z X, Jiang Z Y, Ma L, Liu J Y, Fan X and Liu R S. 2025. HUPE: Heuristic underwater perceptual enhancement with semantic collaborative learning. *International Journal of Computer Vision*, 133(6): 3259-3277.
- Zhao C, Cai W, Dong C and Hu C. 2024a. Wavelet-based fourier Information interaction with frequency diffusion adjustment for underwater image Restoration//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE: 8281-8291.
- Zhao C, Dong C and Cai W. 2024b. Learning a physical-aware diffusion model Based on transformer for underwater image enhancement [EB/OL]. arXiv preprint arXiv:2403.01497.
- Zhou J, Yang T and Zhang W. 2023. Underwater vision enhancement technologies: A comprehensive review, challenges, and recent trends. *Applied Intelligence*, 53(3): 3594-3621.
- Zhu J Y, Park T, Isola P and Efros A A. 2017. Unpaired image-to-image Translation using cycle-consistent adversarial networks//2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE:2223-2232.
- Zhuang J, Zheng Y, Guo B and Yan Y. 2025. Globally deformable information election transformer for underwater image enhancement. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(1): 19-32.
- Zhuang P, Zhang X, Tong J, Wang C, Yuan J and Li C. 2025. Under-

water light field imaging: A survey. CAAI Artificial Intelligence Research, 4: 9150051 [DOI: 10.26599/AIR.2025.9150051].

作者简介

张琦, 女, 副研究员, 主要研究方向为图像复原。E-mail: hit_zq910057@163.com
田春伟(通讯作者), 男, 教授, 博士生导师, 主要研究方向为视频/图像复原和识别。E-mail: chunweitian@hit.edu.cn
吴子昂, 男, 博士研究生, 哈尔滨工业大学计算学部, 主要研

究方向为图像复原。E-mail: ziangwu@stu.hit.edu.cn
薛万利, 男, 教授, 博士生导师, 主要研究方向为无人机视觉感知与理解、手语无障碍沟通。E-mail: xuewanli@email.tjut.edu.cn
朱旗, 男, 教授, 博士生导师, 主要研究方向为脑网络分析、AI交叉应用。E-mail: zhuqi@nuaa.edu.cn
左旺孟, 男, 教授, 博士生导师, 主要研究方向计算机视觉、机器学习。E-mail: cswmzuo@gmail.com