Image Enhancement: A Necessity for Effective Underwater Object Detection?

Anonymous authors
Paper under double-blind review

Abstract

Underwater vision is essential for applications such as marine engineering, aquatic robotics, and environmental monitoring. However, severe image degradation caused by light absorption and scattering often compromises object detection performance. Although underwater image enhancement (UIE) intuitively seems beneficial for restoring visual information and improving detection accuracy, its actual impact remains unclear. This work systematically evaluates state-of-the-art enhancement models and investigates their effects on underwater object detection to answer the key question: **"Is UIE necessary for accurate object detection?"** We conducted a systematic evaluation of 20 representative UIE algorithms—spanning traditional methods, convolutional neural networks (CNNs), generative adversarial networks (GANs), Transformers, and Diffusion models. These methods are applied to two benchmark datasets, RUOD and URPC2020, producing 21 domain variants per dataset (raw + 20 enhanced). To rigorously assess the effect of enhancement on detection, we trained five object detectors on each domain, resulting in 210 unique model configurations (5 detectors $\times$ 21 domains $\times$ 2 datasets). Our findings reveal that, contrary to intuitive expectations, most enhancement techniques actually **degrade** detection accuracy. Only well-designed methods, such as diffusion-based approaches that preserve key low-level features without introducing artificial distortions, can minimize this negative impact. These results provide critical insights into the role of enhancement in underwater vision and highlight important considerations for future research.

1 Introduction

In recent years, underwater vision has emerged as a crucial area of research due to its wide-ranging applications in marine engineering, aquatic robotics, environmental monitoring, and the maintenance of underwater infrastructure (Anwar & Li, 2020; Cui et al., 2020). A fundamental challenge in these applications is object detection (OD), which enables the accurate identification and localization of underwater objects. However, detection performance is often compromised due to the severe degradation of underwater images. Factors such as light absorption, scattering, and backscattering lead to low contrast, color distortion, and blurring, making it difficult for deep learning models to reliably detect objects in underwater environments.

A natural solution to this problem is underwater image enhancement, which aims to restore lost visual information and improve image quality. Intuitively, enhanced images should improve OD accuracy. However, the actual impact of enhancement on detection performance remains unclear. While some studies suggest that enhancement aids detection, others indicate that it may introduce artifacts or distortions that negatively affect model performance (Awad et al., 2024; Chen et al., 2020b). Existing research lacks a systematic evaluation of enhancement techniques in the context of OD, leaving a critical gap in understanding their real-world effectiveness.

This paper addresses the key question: **"Does underwater image enhancement improve object detection performance?"** To investigate this, we conducted a large-scale empirical study analyzing the correlation between enhancement and detection performance. We selected 20 representative underwater image enhancement algorithms, encompassing traditional methods (Hsu & Cheng, 2021a; Fu et al., 2014a; Hummel, 1977; Jr et al., 2013) and SOTA deep learning approaches including CNNs (Li et al., 2020a;b; Wang et al., 2021; Huo et al., 2021) , GANs (Li et al., 2017a; Fabbri et al., 2018a; Wang et al., 2019; Desai et al., 2022; Wang et al., 2023; Jiang et al., 2022) , Transformers (Tang et al., 2022; Khan et al., 2024; Wang et al., 2024) , and Diffusion models (Zhao et al., 2024a; Du et al., 2025; Tang et al., 2023) . These algorithms were individually applied to two benchmark underwater datasets, RUOD (FU2, 2023) and URPC2020 (Liu et al.,

2021), followed by comprehensive qualitative and quantitative analyses of the enhanced images. Following this, we independently trained and tested five widely used OD models on both raw and enhanced image sets—a total of 210 models, including 200 trained on enhanced images and 10 on raw underwater images. This extensive study provides a thorough evaluation of how enhancement techniques influence underwater object detetcion performance.

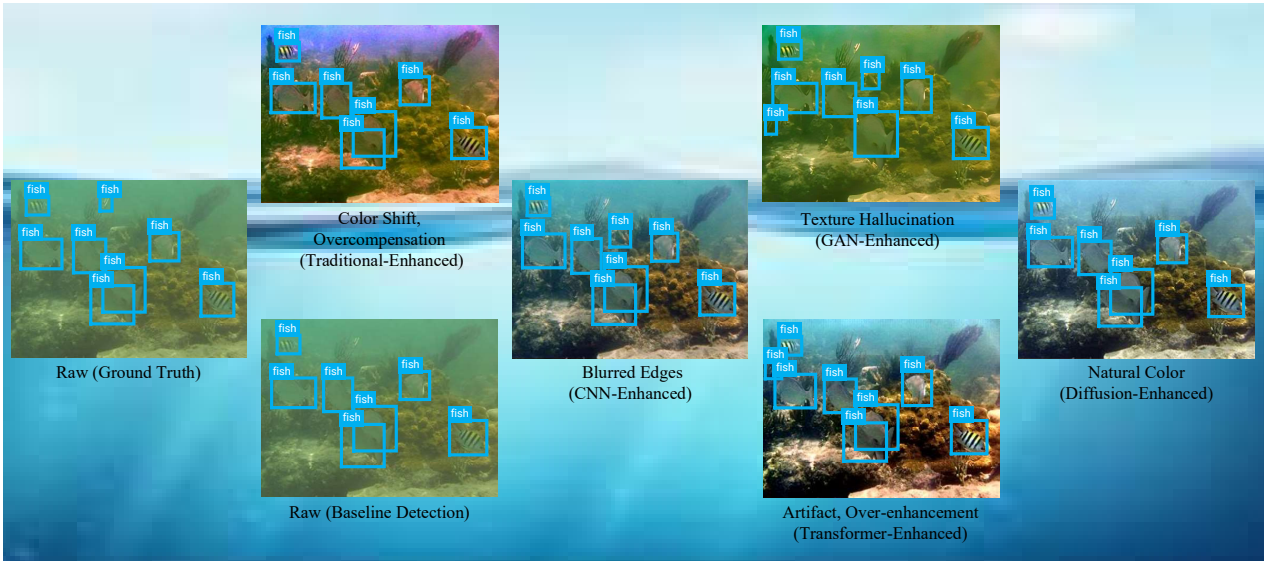

Figure 1: Qualitative comparison of UIE methods and their effects on OD. The raw image and its detection result serve as a baseline. Each enhanced version—Traditional WB (Hsu & Cheng, 2021a), CNN-based PRWNet (Huo et al., 2021), GAN-based WaterGAN (Li et al., 2017a), Transformer-based AutoEnhancer (Tang et al., 2022), and Diffusion-based UIEDP (Du et al., 2025)—introduces distinct visual changes, illustrating how UIE alters image statistics and influences detection performance.

Our empirical observations revealed several key findings. Contrary to intuitive expectations, most enhancement techniques actually reduced OD performance. Although enhanced images may appear visually improved to human observers, they often introduce artifacts and domain shifts, as illustrated in Figure 1, which negatively affect detectors. Preserving edge details is critical for maintaining detection accuracy, while unintended color shifts and additional noise degrade performance. Although contrast changes have little effect, color richness and saturation play an important role in supporting accurate detection. Effective enhancement methods must preserve structural integrity, maintain color consistency, and minimize noise and artifacts. Among all evaluated approaches, diffusion-based methods, such as WF-Diff (Zhao et al., 2024a) and UIEDP (Du et al., 2025), demonstrate clear superiority in balancing color correction and detail preservation, consistently achieving higher detection accuracy across multiple detectors. Building on these findings, the main contributions of this paper are as follows:

- We present a comprehensive study assessing the impact of UIE on OD performance, providing the first large-scale empirical evaluation of enhancement techniques.
- Through extensive experiments and analysis, we identify critical limitations in existing UIE algorithms, particularly their lack of robustness and adaptability to complex underwater environments.
- We reveal that current underwater image quality assessment metrics do not correlate well with OD performance, highlighting the urgent need for task-aware UIE evaluation criteria that better reflect downstream task utility.

This study does not propose new image enhancement techniques but instead evaluates whether preprocessing with enhancement methods improves object detection in underwater environments. We aim to inspire further research into integrated approaches that effectively combine image enhancement with OD, ultimately enhancing both visual quality and detection accuracy in underwater vision systems.

2 Related Work

2.1 Underwater Image Degradation and Enhancement

Underwater images often appear degraded due to light absorption and scattering during propagation through water. This process can be described by the *underwater image formation model*, which accounts for both direct attenuation and backscattering. The observed image $I(x, y)$ can be expressed as $I(x, y) = J(x, y)e^{-\beta d(x,y)} + B(1 - e^{-\beta d(x,y)})$, where $J(x, y)$ is the true scene radiance, β is the attenuation coefficient determined by water properties and wavelength, $d(x, y)$ is the distance between the object and camera, and B is the background light caused by scattering. As the distance increases, backscatter dominates, producing color distortion, reduced contrast, and overall haziness.

Understanding this degradation process enables the development of UIE techniques, which aim to recover $J(x, y)$ from $I(x, y)$, improving visual quality and supporting downstream tasks such as object detection (OD). Enhancement methods are broadly categorized into traditional and deep learning-based approaches.

Traditional methods focus on adjusting intensity distributions or pixel values to improve contrast and color balance. Color correction (Buchsbaum, 1980; Hussain & Akbari, 2016; Hsu & Cheng, 2021b; Finlayson & Trezzi, 2004) and Retinex-based approaches (Fu et al., 2014b) adjust illumination and local contrast, while prior-based methods such as DCP (He et al., 2009) exploit scene priors to remove haze. Contrast enhancement techniques (Agaian et al., 2007; Voronin et al., 2018; Ma et al., 2018; Bai et al., 2020) further improve visibility and detail sharpness.

With deep learning, CNN-based (Zhai et al., 2023; Li et al., 2020a;b) and GAN-based (Li et al., 2017b; Fabbri et al., 2018b; Zhai et al., 2022) methods became dominant, leveraging encoder–decoder and adversarial frameworks to enhance visual realism. Transformer-based models (Chen et al., 2021; Wang et al., 2022; Dosovitskiy et al., 2020) extend these capabilities through self-attention, while diffusion models (Ho et al., 2020; Song et al., 2021) offer stable optimization and strong restoration performance compared to GANs (Goodfellow et al., 2014; Karras et al., 2019) and VAEs (Higgins et al., 2016; Kingma & Welling, 2013; Van Den Oord et al., 2017). Recent works also integrate UIE with OD, and frequency-domain studies (e.g., WF-Diff (Zhao et al., 2024b)) use wavelet–Fourier interactions to refine both high- and low-frequency features, achieving state-of-the-art results.

2.2 Object Detection

Object detection (OD) identifies and localizes objects by predicting category labels and bounding boxes. Modern OD methods are generally categorized as two-stage or one-stage approaches.

Two-stage detectors (**e.g.**, R-CNN, Fast R-CNN, and Faster R-CNN (Girshick et al., 2014; Girshick, 2015; Ren et al., 2015)) use a proposal–refinement pipeline for high accuracy but slower inference. One-stage detectors, such as SSD (Liu et al., 2016), RetinaNet (Lin et al., 2020), and YOLO-NAS (AI, 2023), perform detection in a single pass for real-time efficiency. RetinaNet introduces Focal Loss to address class imbalance, while YOLO-NAS employs Neural Architecture Search (NAS) to optimize the trade-off between speed and accuracy. Both paradigms continue to evolve with transformer-based and anchor-free designs that further push OD performance. In this work, we employ representative detectors to evaluate how UIE affects detection accuracy.

2.3 Joint Image Enhancement and Object Detection

Several studies have explored the impact of image enhancement on high-level vision tasks, including OD. (Pei et al., 2018; 2021) found that image enhancement provides limited improvements in detection performance and, in some cases, may even degrade it. Similarly, (Hahner et al., 2019) reported that dehazing does not improve semantic segmentation performance, while (Li et al., 2019) observed that existing de-raining algorithms can negatively affect OD compared to directly using unprocessed rainy images. In underwater environments, (Zhuang et al., 2022) demonstrated that image enhancement can improve segmentation and saliency detection, though these benefits were observed only in specific cases. Conversely, (Chen et al., 2020a) found that most UIE methods hinder rather than enhance OD, a conclusion further supported by experiments in (Liu et al., 2022b;a). These findings suggest that enhancement techniques often fail to consistently benefit high-level vision tasks, yet the underlying causes remain insufficiently analyzed. Recent

studies (Liu et al., 2022a; Jiang et al., 2021) have attempted to develop end-to-end approaches that jointly optimize UIE and OD. However, the effectiveness of such methods remains uncertain, underscoring the need for further investigation into the interaction between enhancement and detection in underwater vision. In this paper, we address this gap by conducting a comprehensive evaluation of UIE's effect on OD. We assess a diverse set of enhancement algorithms and object detectors, providing a systematic analysis of their interactions and performance impact.

3 Investigating the Effects of UIE on Visual Quality and Structural Fidelity

Before investigating whether and how UIE contributes to OD performance, we first investigate its direct effects on image quality itself. Specifically, we examine how UIE influences key visual and structural aspects such as color consistency, edge and texture preservation, and artifact introduction.

We begin by outlining the experimental setup, including the selected datasets and enhancement methods. We then present an analysis of the enhancement results using both qualitative (visual inspection) and quantitative (reference-free image quality metrics) evaluations. First, we describe the experimental setup on UIE. Next, we analyze the pre-processing results from various UIE algorithms using both qualitative and quantitative methods.

Selected Underwater Image Dataset. We conducted our experiments on two underwater datasets: (1) RUOD (FU2, 2023), the Real-World Underwater OD dataset, is the largest underwater image dataset with object bounding boxes. It contains 14,000 high-resolution underwater images, with 9,800 images used for training and approximately 75,000 annotations spanning 10 categories of aquatic objects such as divers, plants, and various marine animals; and (2) URPC2020 (Liu et al., 2021), the Underwater Robot Professional Contest dataset, consists of 5,543 underwater images categorized into four classes: holothurian, echinus, scallop, and starfish. For our study, we randomly split the dataset into 4,100 training images and 1,443 testing images. The use of both datasets helps us to draw more generalized conclusions. RUOD (FU2, 2023) is a large and diverse dataset from multiple sources, featuring various underwater conditions such as murky waters, low visibility, and strong noise. In contrast, URPC2020 (Liu et al., 2021) includes images from both real and artificial underwater environments, with some offering better lighting and clearer water.

Selected UIE Algorithm. To comprehensively investigate the impact of UIE on underwater OD, we explore both traditional and deep learning-based UIE methods. Traditional UIE techniques enhance obscured details by adjusting intensity distribution or applying pixel-level transformations to improve contrast. We select representative approaches, including color constancy methods (White Balance-based WB(Hsu & Cheng, 2021a), Retinex (Fu et al., 2014a)), contrast enhancement techniques (Histogram Equalization (HE) (Hummel, 1977)), prior knowledge-based methods (UDCP (Jr et al., 2013)), and transform-domain-based approaches (wavelet transform-based PRWNet (Huo et al., 2021)). We also studied top-performing deep learning-based UIE methods leveraging architecture: CNN-based (UWCNN (Li et al., 2020a), Water-Net (Li et al., 2020b), and UIEC^2-Net (Wang et al., 2021)), GAN-based (WaterGAN (Li et al., 2017a), UGAN (Fabbri et al., 2018a), UWGAN (Wang et al., 2019), AquaGAN (Desai et al., 2022), TUDA (Wang et al., 2023), and TOPAL (Jiang et al., 2022)), Transformer-based (AutoEnhancer (Tang et al., 2022), Spectroformer (Khan et al., 2024), and UIE-Convformer (Wang et al., 2024)), and Diffusion-based (WF-Diff (Zhao et al., 2024a), UIEDP (Du et al., 2025), and DM_Underwater (Tang et al., 2023)). To ensure a fair and unbiased comparison of UIE models, we re-trained these models on the UIEB dataset (Li et al., 2020b) and the EUVP dataset (Islam et al., 2020) to generate the corresponding enhanced results.

3.1 Qualitative Analysis of Enhanced Results

Qualitative Evaluation. The results of UIE using different algorithms are shown in Fig. 2. Due to light absorption in water, red light disappears first, followed by green and blue, leading to color distortions, reduced contrast, and poor visibility. This selective attenuation of light wavelengths results in greenish or bluish underwater images, as observed in the raw images in Fig. 2(a). Color deviation significantly impacts the visual quality of underwater images, making it difficult to discern objects and fine details accurately.

(1) Traditional Methods. White Balance (WB) (Hsu & Cheng, 2021a) and Retinex-based approaches (Fu et al., 2014a) partially correct global color shifts but often introduce overcompensation in certain areas, leading to unnatural yellowish or purplish tones. In particular, WB amplifies local illumination inconsistency,

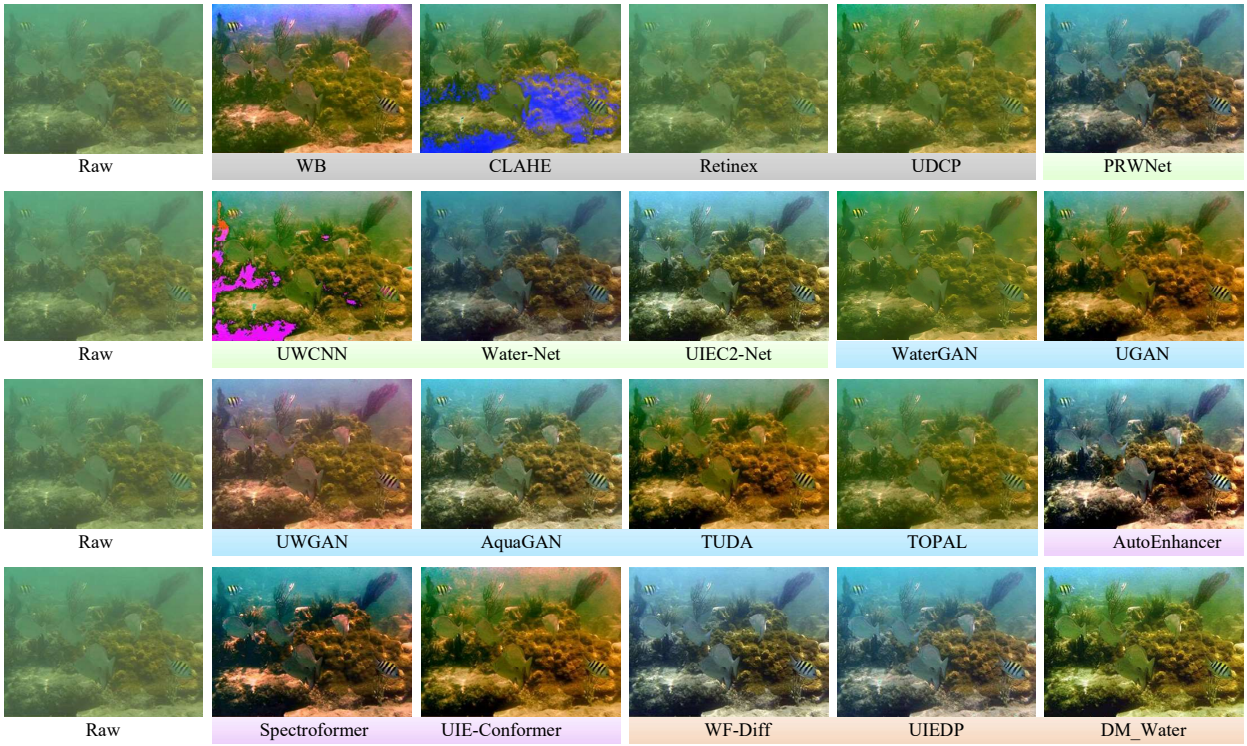

Figure 2: Visual comparison of UIE results. Raw images are shown for reference. Methods are color-coded: Traditional ; CNNs ; GANs ; Transformers ; and Diffusions .

while Retinex enhances global brightness at the expense of suppressing local contrast. CLAHE (Hummel, 1977) improves local contrast but generates visible artifacts in homogeneous regions, resulting in artificial texture patterns. UDCP (Jr et al., 2013), while effective in haze removal, introduces a reddish color bias that disrupts the scene's original color composition.

(2) CNN-Based Deep Learning Methods. PRWNet (Huo et al., 2021) shows notable improvement in preserving fine-grained textures and structural edges through frequency-domain processing, although slight oversharpening is observed under heavy scattering conditions. UWCNN (Li et al., 2020a), Water-Net (Li et al., 2020b), and UIEC2-Net (Wang et al., 2021) enhance color vibrancy and contrast, yet they occasionally over-saturate highlights, causing loss of fine detail in brightly lit regions. Some subtle halo effects are also visible around object boundaries.

(3) GAN-Based Methods. GAN-based models such as UGAN (Fabbri et al., 2018a) and AquaGAN (Desai et al., 2022) exhibit strong color correction capabilities but often introduce blurred artifacts and inconsistent texture reconstruction, especially at object contours and in highly degraded regions. These artifacts manifest as "smearing" effects, weakening object distinction.

(4) Transformer-Based Methods. AutoEnhancer (Tang et al., 2022), Spectroformer (Khan et al., 2024), and UIE-Convformer (Wang et al., 2024) produce smoother contrast gradients and more globally coherent color restoration compared to CNN and GAN models. They effectively suppress halo artifacts and better preserve structural boundaries, although occasional mild over-smoothing can reduce textural richness in detailed regions. **(5) Diffusion-Based Methods.** Diffusion models, notably WF-Diff (Zhao et al., 2024a) and UIEDP (Du et al., 2025), deliver the most balanced performance: restoring natural color tones, enhancing perceptual contrast without exaggeration, and maintaining edge sharpness. Notably, WF-Diff reduces noise amplification and minimizes color bleeding, leading to visually more stable and naturalistic results. DM_Water (Tang et al., 2023), a lightweight diffusion model, offers moderate enhancement but exhibits slight under-correction in highly turbid scenarios.

Key Visual Insights: (1) Color Consistency: Diffusion-based (Tang et al., 2023; Du et al., 2025; Zhao et al., 2024a) and Transformer-based (Khan et al., 2024; Wang et al., 2024) methods provide the most consistent color correction without introducing severe domain shifts. GAN-based methods (Fabbri et al., 2018a; Desai et al., 2022) often generate unnatural hues and inconsistent color patches.

(2) Edge and Texture Preservation: Frequency-domain models (*e.g.*, PRWNet, WF-Diff (Zhao et al., 2024a)) and Transformer-based models (Tang et al., 2022; Khan et al., 2024; Wang et al., 2024)outperform others in preserving object edges and fine textures critical for visual perception and downstream OD tasks.

(3) Artifact Introduction: Traditional methods (*e.g.*, CLAHE (Hummel, 1977) and Retinex (Fu et al., 2014a)) and GAN-based methods (Fabbri et al., 2018a; Li et al., 2017b) often cause visible artifacts, including texture noise, color bleeding, and blurring, which reduce image realism.

(4) Over-Saturation: CNN-based methods (*e.g.*, UWCNN (Li et al., 2020a) and Water-Net (Li et al., 2020b)) tend to over-saturate bright regions, leading to color overshooting and highlight clipping. **Overall Observations:** Diffusion models, especially WF-Diff (Zhao et al., 2024a) and UIEDP (Du et al., 2025), achieve the best qualitative results, effectively balancing enhancement strength and naturalness. Transformer-based models closely follow, offering strong perceptual improvements with minimal artifact introduction. Traditional and GAN-based methods, while enhancing certain aspects, frequently compromise structural integrity and color fidelity.

3.2 Quantitative Analysis of Enhanced Results

To solidify the visual insights described above, we conducted a comprehensive numerical evaluation of the enhancement results using reference-free image quality assessment metrics. These objective metrics provide a complementary perspective to human judgment, enabling quantitative comparisons across UIE algorithms.

Evaluation Metric. For UIE, since no ground truth (clear) images are available for raw inputs, we adopt five widely used reference-free image quality assessment metrics: Average Gradient (AG) (Zhang et al., 2019b), Edge Intensity (EI) (Azmi et al., 2019), Information Entropy (IE) (Zhang et al., 2019a), Underwater Image Quality Measure (UIQM) (Yang & Sowmya, 2015), and Underwater Image Contrast Measure (Yang & Sowmya, 2015). AG measures sharpness and texture by calculating intensity gradients. EI quantifies edge strength, with higher values indicating more texture detail. IE reflects the amount of information in an image. UIQM is a composite score based on colorfulness (UICM), sharpness (UISM), and contrast (UIConM). Higher UIQM scores indicate better overall visual quality. All metrics are used as originally defined, without altering weights or hyperparameters.

Quantitative Comparisons. The images from the selected datasets were enhanced using selected enhancement models and evaluated using the selected enhancement metrics. The results for the RUOD (FU2, 2023) and URPC2020 (Liu et al., 2021) datasets are presented in Table 1, demonstrating the superior performance of WF-Diff (Zhao et al., 2024a) across all metrics, particularly in UIQM (Yang & Sowmya, 2015) and UIConM (Yang & Sowmya, 2015) on both datasets. This indicates its capability to enhance color vibrancy, improve contrast, refine soft edges, and effectively reduce fog effects. AquaGAN (Desai et al., 2022) also performs notably well, especially in terms of UIQM, reflecting its strong ability to enhance colorfulness, sharpness, and contrast. Meanwhile, PRWNet (Huo et al., 2021) achieved the highest AG (Zhang et al., 2019b) score on both datasets, indicating its effectiveness in preserving sharpness and texture details. Similarly, UDCP (Jiang et al., 2022) excels in UCIQE, highlighting its effectiveness in color distribution correction and underwater image clarity enhancement. Additionally, AutoEnhancer demonstrates a notable enhancement capability, further improving underwater image visibility. TUDA (Wang et al., 2023) and TOPAL (Jiang et al., 2022) achieved the highest EI (Azmi et al., 2019) scores on URPC2020 (Liu et al., 2021) dataset, suggesting their effectiveness in global contrast enhancement and edge preservation. Overall, deep learning-based methods significantly outperform traditional approaches, including White Balance (WB) (Hsu & Cheng, 2021a), CLAHE (Hummel, 1977), and Retinex (Fu et al., 2014a), across multiple evaluation metrics. Interestingly, in line with findings from (Liu et al., 2020), our quantitative results reveal that the numerical scores do not always align with human perception of image quality. **This suggests a gap between current underwater image quality assessment metrics and actual visual quality**

Table 1: UIE scores in AG (Zhang et al., 2019b), EI (Azmi et al., 2019), IE (Zhang et al., 2019a), UIQM (Yang & Sowmya, 2015), and UIConM (Yang & Sowmya, 2015) on **RUOD** (FU2, 2023) and **URPC2020** (Liu et al., 2021). Best in **bold**, second-best underlined. Methods are color-coded: Traditional, CNNs, GANs, Transformers, and Diffusions. Trad. – Traditional, Trans. – Transformer, Diff. – Diffusion.

Methods		RUOD (FU2, 2023) Dataset					URPC2020 (Liu et al., 2021) Dataset				
Name	**Type**	**AG ↑**	**EI ↑**	**IE ↑**	**UIQM ↑**	**UIConM ↑**	**AG ↑**	**EI ↑**	**IE ↑**	**UIQM ↑**	**UIConM ↑**
RAW	-	1.31	12.79	6.53	1.59	0.41	2.97	15.46	6.32	1.87	0.47
WB (Hsu & Cheng, 2021a)	Trad.	1.29	13.63	6.18	1.70	0.43	1.15	10.14	7.08	1.53	0.35
CLAHE (Hummel, 1977)	Trad.	4.01	48.95	5.97	2.49	0.38	3.33	35.10	6.44	3.04	0.33
Retinex (Fu et al., 2014a)	Trad.	3.56	37.73	7.09	3.14	0.49	2.58	26.44	6.40	2.51	0.39
UDCP (Jr et al., 2013)	Trad.	2.69	28.62	7.31	2.65	0.54	3.48	24.47	6.69	2.30	0.70
PRWNet (Huo et al., 2021)	CNN	**4.98**	50.18	7.06	3.15	0.57	**6.41**	36.88	6.07	3.60	0.48
UWCNN (Li et al., 2020a)	CNN	3.81	31.59	7.13	3.34	0.78	3.21	36.38	6.69	2.78	0.55
Water-Net (Li et al., 2020b)	CNN	1.79	16.52	7.65	1.95	0.42	5.35	12.52	**7.81**	3.29	0.37
UIEC^2-Net (Wang et al., 2021)	CNN	2.19	27.03	7.36	1.86	0.49	4.60	32.53	6.52	2.08	0.36
WaterGAN (Li et al., 2017a)	GAN	1.70	18.96	7.62	2.75	0.47	5.44	16.93	6.96	2.02	0.36
UGAN (Fabbri et al., 2018a)	GAN	2.09	48.65	7.04	3.47	0.54	2.49	39.42	6.02	3.73	0.66
UWGAN (Wang et al., 2019)	GAN	1.91	27.96	7.40	1.94	0.59	3.55	33.23	6.48	2.30	0.75
AquaGAN (Desai et al., 2022)	GAN	2.72	32.41	**7.85**	3.69	0.47	3.50	37.01	6.36	**4.60**	0.38
TUDA (Wang et al., 2023)	GAN	4.05	43.56	7.53	3.04	0.46	3.63	37.94	7.54	4.07	0.40
TOPAL (Jiang et al., 2022)	GAN	1.69	29.76	6.95	2.11	0.42	3.35	48.35	6.03	2.49	0.50
AutoEnhancer (Tang et al., 2022)	Trans.	1.73	38.96	6.98	1.63	0.54	5.37	**56.85**	7.49	3.89	0.64
Spectroformer (Khan et al., 2024)	Trans.	4.62	18.96	7.02	2.29	0.59	3.36	14.63	6.24	2.68	0.55
UIE-Convformer (Wang et al., 2024)	Trans.	2.08	42.18	7.39	3.35	0.41	2.51	34.09	6.14	2.80	0.31
WF-Diff (Zhao et al., 2024a)	Diff.	3.17	**53.38**	7.79	**3.80**	**0.79**	5.50	38.28	6.57	4.04	**0.77**
UIEDP (Du et al., 2025)	Diff.	3.22	32.64	6.90	3.13	0.56	3.65	25.70	6.33	3.53	0.40
DM_Water (Tang et al., 2023)	Diff.	4.19	48.96	7.02	3.48	0.53	4.14	42.28	6.11	4.29	0.44

as perceived by the human eye, emphasizing the need for improved evaluation methods that incorporate perceptual quality considerations.

Table 2: mAP$_{50:95}$) of **Faster R-CNN** (Ren et al., 2017), **Cascade R-CNN** (Cai & Vasconcelos, 2021), **RetinaNet** (Lin et al., 2020), **SSD** (Liu et al., 2016), and **YOLO-NAS** (AI, 2023). Best in **bold**, second-best underlined. Trad. – Traditional, Trans. – Transformer, Diff. – Diffusion, C-RCNN – Cascade R-CNN.

Methods		RUOD (FU2, 2023) Dataset					URPC2020 (Liu et al., 2021) Dataset				
Name	**Type**	**Faster R-CNN**	**C-RCNN**	**RetinaNet**	**SSD**	**YOLO-NAS**	**Faster R-CNN**	**C-RCNN**	**RetinaNet**	**SSD**	**YOLO-NAS**
RAW	-	**57.91**	59.44	**50.71**	**45.96**	**63.46**	**43.49**	**44.32**	**40.72**	**35.20**	**49.62**
WB (Hsu & Cheng, 2021a)	Trad.	55.33	57.77	48.69	42.72	60.49	41.12	41.93	38.24	32.60	46.86
CLAHE (Hummel, 1977)	Trad.	56.37	57.94	49.14	42.93	62.17	42.79	43.73	39.45	35.03	49.43
Retinex (Fu et al., 2014a)	Trad.	56.02	57.80	48.46	41.15	60.84	40.43	42.31	38.49	33.91	47.65
UDCP (Jr et al., 2013)	Trad.	56.64	58.02	49.31	41.53	61.67	42.60	43.22	39.01	34.54	48.76
PRWNet (Huo et al., 2021)	CNN	54.17	55.63	47.85	39.91	58.53	41.23	42.14	38.52	32.65	47.63
UWCNN (Li et al., 2020a)	CNN	54.01	55.16	47.10	39.52	58.18	41.03	41.95	37.90	32.52	47.14
Water-Net (Li et al., 2020b)	CNN	55.79	57.56	48.88	40.94	58.94	41.89	42.01	38.74	33.65	48.37
UIEC^2-Net (Wang et al., 2021)	CNN	56.04	58.21	49.42	41.82	61.73	42.85	43.47	39.58	34.76	49.09
WaterGAN (Li et al., 2017a)	GAN	54.74	54.01	46.75	39.82	57.32	41.02	41.83	37.42	31.53	45.98
UGAN (Fabbri et al., 2018a)	GAN	52.62	53.68	45.60	38.73	57.08	39.02	39.83	36.32	31.51	44.69
UWGAN (Wang et al., 2019)	GAN	53.15	57.24	48.39	40.68	58.42	40.16	41.72	37.93	33.13	45.47
AquaGAN (Desai et al., 2022)	GAN	53.91	57.68	48.57	40.15	59.03	41.13	42.08	38.06	33.64	46.12
TUDA (Wang et al., 2023)	GAN	53.31	55.98	47.63	39.34	57.79	39.82	40.65	36.72	32.04	45.73
TOPAL (Jiang et al., 2022)	GAN	54.17	57.45	48.86	44.03	59.92	43.05	43.66	40.42	34.52	49.19
AutoEnhancer (Tang et al., 2022)	Trans.	54.39	57.23	48.01	43.65	62.06	40.92	41.72	38.16	33.04	48.93
Spectroformer (Khan et al., 2024)	Trans.	56.72	58.13	49.30	44.15	61.41	41.95	43.47	39.86	34.85	49.13
UIE-Convformer (Wang et al., 2024)	Trans.	57.61	56.94	48.23	44.20	61.52	42.65	43.46	39.68	34.78	49.24
WF-Diff (Zhao et al., 2024a)	Diff.	57.74	**59.52**	49.92	45.37	62.30	43.13	44.09	40.53	35.10	49.57
UIEDP (Du et al., 2025)	Diff.	57.48	58.82	48.03	44.76	60.96	42.42	43.47	39.52	35.12	48.97
DM_Water (Tang et al., 2023)	Diff.	57.85	59.07	48.72	45.16	61.98	42.75	43.63	40.15	35.08	49.69

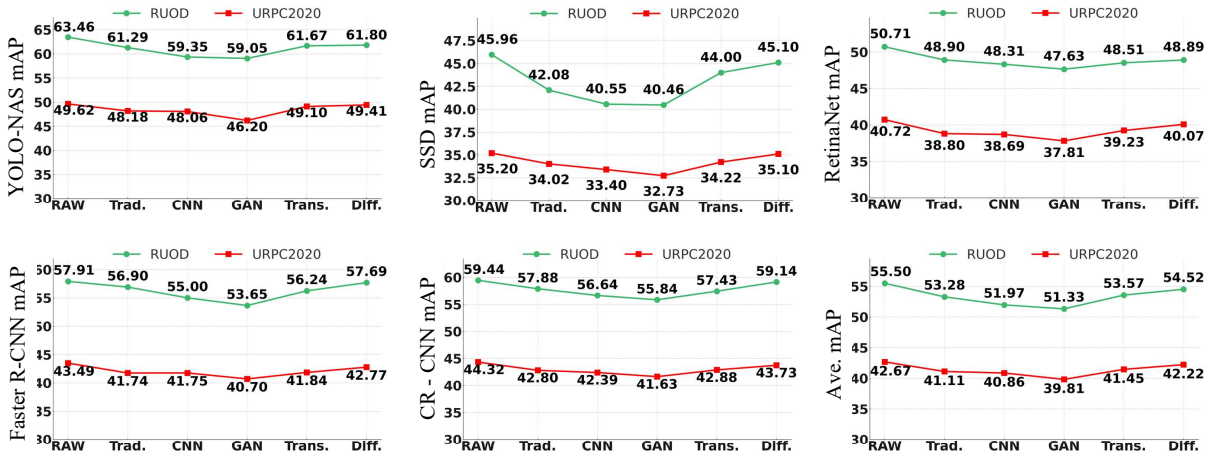

Figure 3: $mAP_{50:95}$ of five detectors on RUOD (FU2, 2023) and URPC2020 (Liu et al., 2021) using RAW and enhanced images. Enhancements cover five UIE types: Trad., CNN, GAN, Trans., and Diff. Overall, detection accuracy generally drops after enhancement.

4 Effect of Underwater Image Enhancement on Object Detection

Image Domains. To evaluate the impact of UIE on OD performance, we applied 20 UIE algorithms to two benchmark datasets, RUOD (FU2, 2023) and URPC2020 (Liu et al., 2021). This resulted in 21 domains per dataset (raw + 20 enhanced). For each domain, we trained five object detectors, yielding a total of 210 model configurations (5 detectors $\times$ 21 domains $\times$ 2 datasets).

Object Detectors. We selected five popular deep learning-based object detectors: three one-stage detectors (SSD (Liu et al., 2016), RetinaNet (Lin et al., 2020), YOLO-NAS (AI, 2023)) and two two-stage detectors (Faster R-CNN (Ren et al., 2017), Cascade R-CNN (Cai & Vasconcelos, 2021)). Each detector was retrained separately on each domain-specific dataset using the same training and testing splits. We used the SGD optimizer with a momentum of 0.95, learning rate of 0.01, and weight decay of 10^{-5}. Input image resolution was set to 400×300 for both datasets. All experiments were conducted on a Linux server with four Nvidia A100 GPUs. Each detector was trained and evaluated within its corresponding domain to ensure consistent performance comparisons. The OD performance is measured using mean Average Precision (mAP), which averages the AP across all object classes. The results are summarized in Table 2. Visual samples are shown in Fig. 4.

4.1 Key Observations and Analysis

We made several key observations and conducted analysis:

(1) Raw images yield the highest detection performance: Detectors trained on raw images consistently achieve higher mAP values than those trained on enhanced images, with one exception: the Cascade R-CNN (Cai & Vasconcelos, 2021) detector, which shows a marginal improvement of 0.08% over its raw-trained counterpart. However, this slight increase does not demonstrate a clear advantage of enhancement-based training. Enhancement methods often introduce distribution shifts in pixel intensities, color statistics, and textures, leading to a mismatch with original training assumptions. Raw images, despite degradation, preserve the natural domain where detectors perform best.

(2) CNN-based UIE performs slightly worse than traditional UIE: To support this observation, we calculate the average detection performance for each enhancement category (Figure 3). CNN-based methods (e.g., PRWNet (Huo et al., 2021), UWCNN (Li et al., 2020a), Water-Net (Li et al., 2020b), UIEC2-Net (Wang et al., 2021)) show slightly lower detection accuracy than traditional approaches like CLAHE (Hummel, 1977), Retinex (Fu et al., 2014a), and TOPAL (Jiang et al., 2022). The average mAP for CNN-based methods is 52.22 (RUOD) and 40.97 (URPC2020), compared to 52.84 and 40.97 for traditional UIE methods. This slight drop may result from CNN models being tuned for visual quality rather than detection accuracy. They can introduce subtle artifacts, domain shifts, or uneven adjustments—especially in low-contrast areas—that

affect object features. In contrast, traditional methods apply simpler, more uniform corrections that tend to preserve these features more reliably. These findings align well with the visual comparisons in Figure 2.

(3) GAN-based UIE has the most severe negative impact on detection performance: UGAN (Fabbri et al., 2018a), UWGAN (Wang et al., 2019) and AquaGAN (Desai et al., 2022) significantly degrade detection accuracy, reducing the mAP by an average of approximately 4% compared to training on raw images. GANs often generate hallucinated textures, unnatural colors, and blurred edges, to improve perceptual appeal. However, these distortions corrupt the fine spatial structures and edge sharpness critical for OD, confusing the detectors.

(4) Diffusion-based UIE has minimal negative effects: Among all enhancement methods, diffusion-based approaches had the least negative impact on detection performance, with results closest to those obtained with raw images. Notably, WF-Diff (Zhao et al., 2024a) achieved the second-highest accuracy across most detectors. WF-Diff applies wavelet transformation in the frequency domain combined with diffusion processes in the wavelet space, leading to the best overall performance. Diffusion models reconstruct images by denoising while preserving statistical structure, thus maintaining key features like object contours, color consistency, and textures necessary for robust detection. Their gradual, probabilistic refinement avoids the over-correction and distortion seen in other methods.

(5) Detection performance of one-stage and two-stage detectors: The performance degradation in one-stage detectors is similar to that in two-stage detectors. Detectors trained on raw images generally achieved higher mAP values than those trained on enhanced images, with Cascade R-CNN (Cai & Vasconcelos, 2021) being the only exception, reaching 59.52% mAP on the enhanced RUOD (Liu et al., 2021) dataset. Since both detector types rely heavily on low-level texture and edge features, any domain shift or feature alteration caused by UIE methods affects them similarly.

(6) Detection performance on two datasets: Detectors trained on RUOD (FU2, 2023) consistently outperform those trained on URPC2020 (Liu et al., 2021). Both datasets show similar performance trends across enhancement methods, with two exceptions: YOLO-NAS (AI, 2023) achieved the second-highest mAP in the TOPAL (Jiang et al., 2022) domain on RUOD, and SSD (Liu et al., 2016) achieved the second-highest mAP in the UIEDP (Du et al., 2025) domain on URPC2020. The better results on RUOD are due to its greater diversity of scenes and object categories, which promote stronger generalization. In contrast, URPC2020's narrower domain makes detectors more sensitive to mismatches introduced by enhancement.

In summary, these findings suggest that, despite the intuitive expectation that UIE should aid detection, most current enhancement methods introduce distributional shifts, artifacts, or domain inconsistencies that degrade detector performance. Only well designed UIE techniques (*e.g.*, diffusion-based) that preserve key low-level features without introducing artificial distortions can minimize this negative impact.

4.2 Study of the Relationship Between Image Enhancement and Detection

To further explore the underlying relationship between underwater image enhancement and OD performance, we conducted a trend analysis using scatter plots to examine the correlation between the enhanced image quality composite index (UIQM (Yang & Sowmya, 2015)) scores and detection accuracy (mAP) across various enhancement methods on the RUOD (FU2, 2023) and URPC2020 (Liu et al., 2021) datasets.

In Fig. 5, each dot shows a detector's mAP versus the image quality score of an enhancement method. The regression line indicates the linear trend between UIQM and mAP, with the shaded area showing the 95% confidence interval—wider bands mean greater uncertainty, while narrower ones indicate higher confidence

Key Observations: (1) Weak Correlation Between UIQM and Detection Performance: The scatter plots reveal that UIQM scores exhibit a weak or negligible correlation with detection accuracy across different models and datasets. Some detectors, such as YOLO-NAS (AI, 2023) and RetinaNet (Lin et al., 2020) on the RUOD (FU2, 2023) dataset, show a slight downward trend, while others display an almost flat regression line, indicating little to no relationship between UIQM (Yang & Sowmya, 2015) and detection performance. This suggests that higher UIQM (Yang & Sowmya, 2015) scores do not necessarily translate to improved OD accuracy. The scattered distribution of points around the regression line further supports

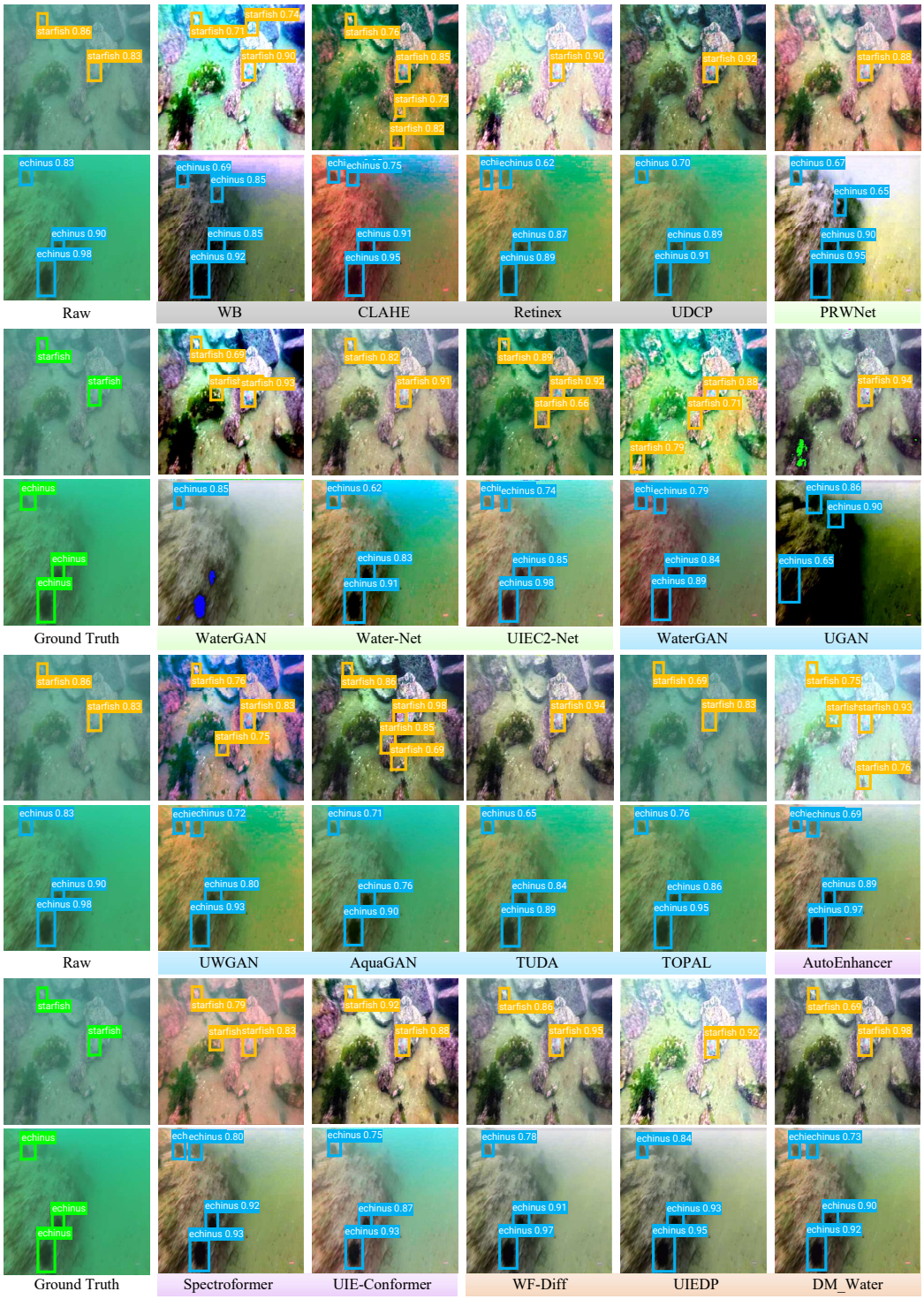

Figure 4: YOLO-NAS (AI, 2023) detection results on raw and enhanced images. Green boxes indicate ground truth. Detection Results are shown for Trad., CNN, GAN, Trans., and Diff. methods. Orange and blue boxes mark echinus and starfish. Raw images yield better detection than most enhanced ones.

the inference that detection accuracy is influenced by other factors beyond UIQM (Yang & Sowmya, 2015). **This implies that conventional image quality metrics may not be reliable indicators of detection performance in underwater environments.**

(2) Limitations of Current Evaluation Metrics: Existing objective underwater image enhancement evaluation metrics (*e.g.*, the composite index UIQM (Yang & Sowmya, 2015)) do not effectively reflect the

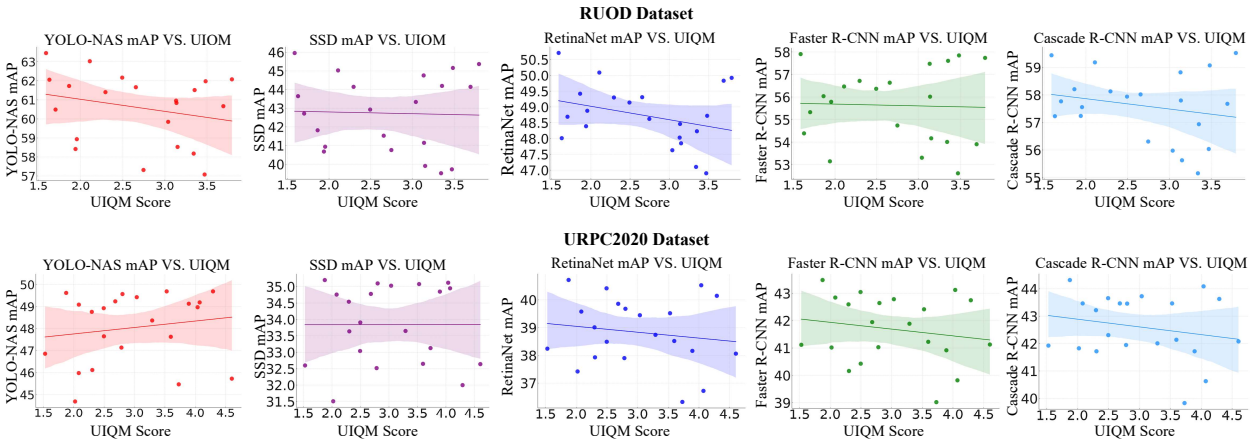

Figure 5: Relationship between detection accuracy (mAP) and image quality scores for different detectors on the RUOD and URPC2020 datasets. Regression lines with 95% confidence intervals show that higher image quality scores have weak or no correlation with detection performance.

impact of enhancement on OD performance. This highlights the need for a more detection-aware enhancement evaluation metric.

(3) Degradation Caused by GAN-based Methods: GAN-based enhancement methods (*e.g.*, UGAN (Fabbri et al., 2018a), UWGAN (Wang et al., 2019), AquaGAN (Desai et al., 2022)) exhibit the worst average detection performance in qualitative comparisons (Fig. 3). As shown in Fig. 2, UWGAN (Wang et al., 2019) introduces noise and alters the original greenish underwater domain into inconsistent colors, such as brown and purple. These transformations introduce artifacts, blur edges, and degrade visual quality, ultimately reducing detection accuracy.

(4) Noise and Color Distortion Introduced: The enhancement processes of WB (Hsu & Cheng, 2021a), Retinex (Fu et al., 2014a), and CLAHE (Hummel, 1977) introduce noise and color distortion, particularly purple noise, leading to performance degradation compared to raw images. Among them, WB (Hsu & Cheng, 2021a) suffers from a low-contrast problem, making object boundaries less distinguishable. In contrast, CLAHE (Hummel, 1977) preserves more edge information than both WB and Retinex (Fu et al., 2014a). As a result, detectors trained on CLAHE-enhanced images outperform those trained on WB- and Retinex-enhanced images.

(5) Domain Shifts by Enhancement Methods: Methods such as Spectroformer, and UIE-Convformer (Wang et al., 2024) exhibit domain shifts, converting the greenish underwater tone into brownish or reddish hues. In contrast, CNN-based enhancement methods like UIEC2-Net (Wang et al., 2021) show minimal deviation from raw images, helping preserve detector performance.

(6) Superiority of Diffusion-Based Enhancement: Diffusion-based enhancement methods (*e.g.*, WF-Diff (Zhao et al., 2024a) and UIEDP (Du et al., 2025)) perform significantly better in color correction and detail preservation. These methods achieve relatively higher detection accuracy within their respective domain detectors, underscoring the advantages of diffusion-based approaches for underwater OD.

Key Inferences: From our observations, we infer that: **(1) Edge preservation is crucial:** Edge degradation significantly impacts detector performance. Enhancement algorithms must preserve edge details to maintain detection accuracy. **(2) Color cast can disrupt detection:** Unintended color shifts introduced by enhancement methods can create domain inconsistencies, degrading detection performance. **(3) Noise decreases detection accuracy:** Enhancement methods should avoid introducing additional noise, as it negatively impacts detection. **(4) Contrast has minimal effect, while color richness and saturation matter:** while contrast changes have little impact on detection, color richness and saturation play a crucial role in influencing detector performance. These findings emphasize the need for enhancement methods that prioritize structural integrity and color consistency while minimizing noise and unwanted artifacts to improve underwater OD.

5 Discussion and Conclusion

In this work, we conducted a comprehensive empirical investigation into the impact of UIE on OD. Our study systematically evaluated 20 diverse UIE algorithms, including traditional, CNN based, GAN based, Transformer based, and Diffusion based methods, and retrained 5 benchmark object detectors across 21 domains, resulting in a total of 210 model configurations. This extensive experimental design allowed us to rigorously examine the interplay between enhancement quality and detection accuracy. By covering a broad spectrum of enhancement techniques and detection architectures, our findings provide robust and generalizable insights into how visual preprocessing affects downstream OD performance in underwater environments. Our key findings are summarized as follows:

- UIE does not universally improve detection performance. Contrary to intuitive expectations, most enhancement techniques reduced OD accuracy. Although enhanced images may appear visually improved to human, they often introduce artifacts and domain shifts that negatively affect detector performance.
- One-stage and two-stage detectors are similarly impacted. Our results indicate that both types of detection architectures suffer comparable performance degradation when trained on enhanced images. This suggests that the negative impact is rooted in the enhancement process itself rather than in the detector design.
- The primary factor affecting detection is not the original underwater color cast, but rather the multiple new color tones introduced by enhancement methods. These artificial color shifts often result in inconsistent feature representations, leading to performance drops.
- Enhancement methods may introduce noise, edge blurring, and texture loss, all of which can significantly degrade object detector performance.
- While color correction is often a goal in enhancement, excessive adjustments to contrast, saturation, and color richness can negatively affect detection. Among these, over-saturation and unnatural color alterationshave the most significant impact.
- The choice of enhancement method is critical. GAN-based techniques tend to produce perceptual artifacts and yield the worst detection performance. In contrast, diffusion-based methods such as WF-Diff and UIEDP offer a better balance between perceptual quality and detection accuracy.
- Current underwater image quality assessment metrics are limited. They often fail to align with both human visual perception and the performance of OD models. This highlights the need for new evaluation metrics that are specifically designed for task-oriented applications and better reflect the effectiveness of enhancement methods in detection tasks.

These findings point to a deeper challenge: underwater enhancement must be task-aware. Enhancing visual appeal alone is not sufficient when the end goal is high-level vision tasks like detection or segmentation. As such, future efforts should shift toward task-oriented enhancement frameworks that co-optimize visual quality and downstream utility. We propose three critical directions for future research:

- Task-aligned enhancement models: Develop UIE approaches that are optimized jointly with object detectors, enabling end-to-end learning that enhances both perceptual quality and task performance.
- New evaluation metrics: Existing quality metrics fail to reflect downstream utility. There is a clear need for task-sensitive metrics that correlate with detection outcomes.
- Robustness and generalization: UIE methods should generalize across diverse underwater scenes while minimizing distortion and maintaining feature consistency for detection.

In conclusion, our large-scale empirical study provides a critical benchmark for understanding how underwater image enhancement affects OD. We hope this work inspires future efforts toward developing enhancement methods that are not only visually compelling but also effective in supporting real-world underwater vision tasks. Moving forward, we aim to continue exploring task-oriented enhancement strategies and novel evaluation metrics that bridge the gap between visual perception and deep learning-based detection.

References

Rethinking general underwater object detection: Datasets, challenges, and solutions. *Neurocomputing*, 517: 243–256, 2023. ISSN 0925-2312. doi: https://doi.org/10.1016/j.neucom.2022.10.039.

S. S. Agaian, B. Silver, and K. A. Panetta. Transform coefficient histogram-based image enhancement algorithms using contrast entropy. *IEEE Transactions on Image Processing*, 16(3):741–758, 2007.

Deci AI. Yolo-nas: Faster and more accurate object detection with nas. *Deci AI Research*, 2023. URL `https://github.com/Deci-AI/YOLO-NAS`.

Saeed Anwar and Chongyi Li. Diving deeper into underwater image enhancement: A survey. *Signal Processing: Image Communication*, 89:115978, 2020.

Ali Awad, Ashraf Saleem, Sidike Paheding, Evan Lucas, Serein Al-Ratrout, and Timothy C. Havens. Evaluating the impact of underwater image enhancement on object detection performance: A comprehensive study, 2024. URL `https://arxiv.org/abs/2411.14626`.

K. Z. M. Azmi, A. S. A. Ghani, Z. M. Yusof, and Z. Ibrahim. Natural-based underwater image color enhancement through fusion of swarm intelligence algorithm. *Applied Soft Computing*, 85:105810, 2019. doi: 10.1016/j.asoc.2019.105810.

L. Bai, W. Zhang, X. Pan, and C. Zhao. Underwater image enhancement based on global and local equalization of histogram and dual-image multi-scale fusion. *IEEE Access*, 8:128973–128990, 2020.

Gershon Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin Institute*, 310(1):1–26, 1980.

Zhaowei Cai and Nuno Vasconcelos. Cascade r-cnn: Delving into high quality object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 43(5):1483–1498, 2021. doi: 10.1109/TPAMI.2019.2956516.

Hanting Chen, Yunhe Wang, Tianyu Guo, Chang Xu, Yiping Deng, Zhenhua Liu, Siwei Ma, Chunjing Xu, Chao Xu, and Wen Gao. Pre-trained image processing transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12299–12310, 2021.

Xingyu Chen, Yue Lu, Zhengxing Wu, Junzhi Yu, and Li Wen. Reveal of domain effect: How visual restoration contributes to object detection in aquatic scenes. *arXiv preprint arXiv:2003.01913*, 2020a.

Xingyu Chen, Yue Lu, Zhengxing Wu, Junzhi Yu, and Li Wen. Reveal of domain effect: How visual restoration contributes to object detection in aquatic scenes, 2020b. URL `https://arxiv.org/abs/2003.01913`.

Suxia Cui, Yu Zhou, Yonghui Wang, and Lujun Zhai. Fish detection using deep learning. *Applied Computational Intelligence and Soft Computing*, 2020(1):3738108, 2020.

Chaitra Desai, Badduri Sai Sudheer Reddy, Ramesh Ashok Tabib, Ujwala Patil, and Uma Mudenagudi. Aquagan: Restoration of underwater images. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 295–303, 2022. doi: 10.1109/CVPRW56347.2022.00044.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, and et al. Gelly, Sylvain. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

Dazhao Du, Enhan Li, Lingyu Si, Wenlong Zhai, Fanjiang Xu, Jianwei Niu, and Fuchun Sun. Uiedp: Boosting underwater image enhancement with diffusion prior. *Expert Systems with Applications*, 259:125271, 2025. ISSN 0957-4174. doi: https://doi.org/10.1016/j.eswa.2024.125271. URL `https://www.sciencedirect.com/science/article/pii/S0957417424021389`.

Cameron Fabbri, Md Jahidul Islam, and Junaed Sattar. Enhancing underwater imagery using generative adversarial networks. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 7159–7165, 2018a. doi: 10.1109/ICRA.2018.8460552.

Cameron Fabbri, Md Jahidul Islam, and Junaed Sattar. Enhancing underwater imagery using generative adversarial networks. In *2018 IEEE international conference on robotics and automation (ICRA)*, pp. 7159–7165. IEEE, 2018b.

G. Finlayson and E. Trezzi. Shades of gray and colour constancy. In *Proceedings of Color Imaging*, pp. 37–41, 2004.

X. Fu, P. Zhang, Y. Huang, et al. A retinex-based enhancing approach for single underwater image. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pp. 4572–4576. IEEE, 2014a. doi: 10.1109/ICIP.2014.7025912.

Xueyang Fu, Peixian Zhuang, Yue Huang, Yinghao Liao, Xiao-Ping Zhang, and Xinghao Ding. A retinex-based enhancing approach for single underwater image. In *2014 IEEE international conference on image processing (ICIP)*, pp. 4572–4576. IEEE, 2014b.

Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448, 2015.

Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 580–587, 2014.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 2672–2680, 2014.

Martin Hahner, Dengxin Dai, Christos Sakaridis, Jan-Nico Zaech, and Luc Van Gool. Semantic understanding of foggy scenes with purely synthetic data. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pp. 3675–3681, 2019. doi: 10.1109/ITSC.2019.8917518.

Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1956–1963. IEEE, 2009.

Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. β-vae: Learning basic visual concepts with a constrained variational framework. *arXiv preprint arXiv:1611.02731*, 2016.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems (NeurIPS)*, Virtual Conference, December 6-12 2020.

Wei-Yen Hsu and Han-Chang Cheng. A novel automatic white balance method for color constancy under different color temperatures. *IEEE Access*, 9:111925–111937, 2021a. doi: 10.1109/ACCESS.2021.3102636.

Wei-Yen Hsu and Han-Chang Cheng. A novel automatic white balance method for color constancy under different color temperatures. *IEEE Access*, 9:111925–111937, 2021b.

R. Hummel. Image enhancement by histogram transformation. *Computer Graphics and Image Processing*, 6(2):184–195, 1977. doi: 10.1016/S0146-664X(77)80009-X.

Fushuo Huo, Bingheng Li, and Xuegui Zhu. Efficient wavelet boost learning-based multi-stage progressive refinement network for underwater image enhancement. In *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pp. 1944–1952, 2021. doi: 10.1109/ICCVW54120.2021.00221.

Md Akmol Hussain and Akbar Sheikh Akbari. Max-rgb based colour constancy using the sub-blocks of the image. In *2016 9th International Conference on Developments in eSystems Engineering (DeSE)*, pp. 289–294. IEEE, 2016.

Md Jahidul Islam, Youya Xia, and Junaed Sattar. Fast underwater image enhancement for improved visual perception. *IEEE Robotics and Automation Letters*, 5(2):3227–3234, 2020. doi: 10.1109/LRA.2020.2974710.

Lihao Jiang, Yi Wang, Qi Jia, Shengwei Xu, Yu Liu, Xin Fan, Haojie Li, Risheng Liu, Xinwei Xue, and Ruili Wang. Underwater species detection using channel sharpening attention. In *Proceedings of the 29th ACM International Conference on Multimedia*, MM '21, pp. 4259–4267, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450386517. doi: 10.1145/3474085.3475563. URL `https://doi.org/10.1145/3474085.3475563`.

Zhiying Jiang, Zhuoxiao Li, Shuzhou Yang, Xin Fan, and Risheng Liu. Target oriented perceptual adversarial fusion network for underwater image enhancement. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(10):6584–6598, 2022. doi: 10.1109/TCSVT.2022.3174817.

P. Drews Jr, E. do Nascimento, F. Moraes, S. Botelho, and M. Campos. Transmission estimation in underwater single images. In *2013 IEEE International Conference on Computer Vision Workshops (ICCVW)*, pp. 825–830. IEEE, 2013. doi: 10.1109/ICCVW.2013.112.

Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4401–4410, 2019.

Md Raqib Khan, Priyanka Mishra, Nancy Mehta, Shruti S. Phutke, Santosh Kumar Vipparthi, Sukumar Nandi, and Subrahmanyam Murala. Spectroformer: Multi-domain query cascaded transformer network for underwater image enhancement. In *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 1443–1452, 2024. doi: 10.1109/WACV57701.2024.00148.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Chongyi Li, Saeed Anwar, and Fatih Porikli. Underwater scene prior inspired deep underwater image and video enhancement. *Pattern Recognition*, 98:107038, 2020a. doi: 10.1016/j.patcog.2019.107038.

Chongyi Li, Chunle Guo, Wenqi Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao. An underwater image enhancement benchmark dataset and beyond. *IEEE Transactions on Image Processing*, 29:4376–4389, 2020b. doi: 10.1109/TIP.2019.2955241.

Jie Li, Katherine A. Skinner, Ryan Eustice, and M. Johnson-Roberson. Watergan: Unsupervised generative network to enable real-time color correction of monocular underwater images. *IEEE Robotics and Automation Letters (RA-L)*, 2017a. accepted.

Jie Li, Katherine A Skinner, Ryan M Eustice, and Matthew Johnson-Roberson. Watergan: Unsupervised generative network to enable real-time color correction of monocular underwater images. *IEEE Robotics and Automation letters*, 3(1):387–394, 2017b.

Siyuan Li, Iago Breno Araujo, Wenqi Ren, Zhangyang Wang, Eric K. Tokuda, Roberto Hirata Junior, Roberto Cesar-Junior, Jiawan Zhang, Xiaojie Guo, and Xiaochun Cao. Single image deraining: A comprehensive benchmark analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 42(2):318–327, 2020. doi: 10.1109/TPAMI.2018.2858826.

Chongwei Liu, Haojie Li, Shuchang Wang, Ming Zhu, Dong Wang, Xin Fan, and Zhihui Wang. A dataset and benchmark of underwater object detection for robot picking. In *2021 IEEE international conference on multimedia & expo workshops (ICMEW)*, pp. 1–6. IEEE, 2021.

Risheng Liu, Xin Fan, Ming Zhu, Minjun Hou, and Zhongxuan Luo. Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light. *IEEE Trans. Cir. and Sys. for Video Technol.*, 30(12):4861–4875, December 2020. ISSN 1051-8215. doi: 10.1109/TCSVT.2019.2963772. URL `https://doi.org/10.1109/TCSVT.2019.2963772`.

Risheng Liu, Zhiying Jiang, Shuzhou Yang, and Xin Fan. Twin adversarial contrastive learning for underwater image enhancement and beyond. *IEEE Transactions on Image Processing*, 31:4922–4936, 2022a.

Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. Ssd: Single shot multibox detector. *European Conference on Computer Vision (ECCV)*, pp. 21–37, 2016. doi: 10.1007/978-3-319-46448-0_2.

Zheng Liu, Yaoming Zhuang, Pengrun Jia, Chengdong Wu, and Hongli Xu ang Zhanlin Liu. A novel underwater image enhancement and improved underwater biological detection pipeline, 2022b. URL `https://arxiv.org/abs/2205.10199`.

J. Ma, X. Fan, S. X. Yang, X. Zhang, and X. Zhu. Contrast limited adaptive histogram equalization-based fusion in yiq and hsi color spaces for underwater image enhancement. *International Journal of Pattern Recognition and Artificial Intelligence*, 32(07):1854018, 2018.

Yanting Pei, Yaping Huang, Qi Zou, Yuhang Lu, and Song Wang. Does haze removal help cnn-based image classification? In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.

Yanting Pei, Yaping Huang, Qi Zou, Xingyuan Zhang, and Song Wang. Effects of image degradation and degradation removal to cnn-based image classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(4):1239–1253, 2021. doi: 10.1109/TPAMI.2019.2950923.

Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pp. 91–99, 2015.

Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 39(6):1137–1149, 2017. doi: 10.1109/TPAMI.2016.2577031.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *9th International Conference on Learning Representations (ICLR)*, Virtual Event, Austria, May 3-7 2021.

Yi Tang, Takafumi Iwaguchi, Hiroshi Kawasaki, Ryusuke Sagawa, and Ryo Furukawa. Autoenhancer: Transformer on u-net architecture search for underwater image enhancement. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, pp. 1403–1420, December 2022.

Yi Tang, Hiroshi Kawasaki, and Takafumi Iwaguchi. Underwater image enhancement by transformer-based diffusion model with non-uniform sampling for skip strategy. In *Proceedings of the 31st ACM International Conference on Multimedia*, MM '23, pp. 5419–5427, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9798400701085. doi: 10.1145/3581783.3612378. URL `https://doi.org/10.1145/3581783.3612378`.

Aaron Van Den Oord, Oriol Vinyals, et al. Neural discrete representation learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 30, 2017.

V. Voronin, E. Semenishchev, S. Tokareva, A. Zelenskiy, and S. Agaian. Underwater image enhancement algorithm based on logarithmic transform histogram matching with spatial equalization. In *2018 14th IEEE International Conference on Signal Processing (ICSP)*, pp. 434–438. IEEE, 2018.

Biao Wang, Haiyong Xu, Gangyi Jiang, Mei Yu, Tingdi Ren, Ting Luo, and Zhongjie Zhu. Uie-convformer: Underwater image enhancement based on convolution and feature fusion transformer. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 8(2):1952–1968, 2024. doi: 10.1109/TETCI.2024.3359061.

Nan Wang, Yabin Zhou, Fenglei Han, Haitao Zhu, and Yaojing Zheng. Uwgan: Underwater gan for real-world underwater color restoration and dehazing, 2019.

Yudong Wang, Jichang Guo, Huan Gao, and Huihui Yue. Uiec^2-net: Cnn-based underwater image enhancement using two color spaces. *Signal Processing: Image Communication*, 96:116250, 2021. ISSN 0923-5965. doi: 10.1016/j.image.2021.116250. URL `https://www.sciencedirect.com/science/article/pii/S0923596521001004`.

Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17683–17693, 2022.

Zhengyong Wang, Liquan Shen, Mai Xu, Mei Yu, Kun Wang, and Yufei Lin. Domain adaptation for underwater image enhancement. *IEEE Transactions on Image Processing*, 32:1442–1457, 2023.

M. Yang and A. Sowmya. An underwater color image quality evaluation metric. *IEEE Transactions on Image Processing*, 24(12):6062–6071, 2015. doi: 10.1109/TIP.2015.2475688.

Lujun Zhai, Yonghui Wang, Suxia Cui, and Yu Zhou. Enhancing underwater image using degradation adaptive adversarial network. In *2022 IEEE International Conference on Image Processing (ICIP)*, pp. 4093–4097, 2022. doi: 10.1109/ICIP46576.2022.9897624.

Lujun Zhai, Yonghui Wang, Suxia Cui, and Yu Zhou. A comprehensive review of deep learning-based real-world image restoration. *IEEE Access*, 11:21049–21067, 2023. doi: 10.1109/ACCESS.2023.3250616.

W. Zhang, L. Dong, X. Pan, J. Zhou, L. Qin, and W. Xu. Single image defogging based on multi-channel convolutional msrcr. *IEEE Access*, 7:72492–72504, 2019a. doi: 10.1109/ACCESS.2019.2919970.

Weidong Zhang, Lili Dong, Xipeng Pan, Jingchun Zhou, Li Qin, and Wenhai Xu. Single image defogging based on multi-channel convolutional msrcr. *IEEE Access*, 7:72492–72504, 2019b. doi: 10.1109/ACCESS.2019.2920403.

Chen Zhao, Weiling Cai, Chenyu Dong, and Chengwei Hu. Wavelet-based fourier information interaction with frequency diffusion adjustment for underwater image restoration. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8281–8291, 2024a. doi: 10.1109/CVPR52733.2024.00791.

Chen Zhao, Weiling Cai, Chenyu Dong, and Chengwei Hu. Wavelet-based fourier information interaction with frequency diffusion adjustment for underwater image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8281–8291, 2024b.

Peixian Zhuang, Jiamin Wu, Fatih Porikli, and Chongyi Li. Underwater image enhancement with hyper-laplacian reflectance priors. *IEEE Transactions on Image Processing*, 31:5442–5455, 2022.