Journal of
Marine Science and Engineering

MDPI

Article

SGDC-UIE: A Semantic Guidance Network with Degradation Consistency for Underwater Image Enhancement

Rui Ming, Jianshan Zhang, Taotao Lai, Haibo Luo * and Jiancheng Yang *

Fuzhou Institute of Oceanography, School of Computer Science and Big Data, Minjiang University, Fuzhou 350108, China; rming@mju.edu.cn (R.M.)
* Correspondence: robhappy@mju.edu.cn (H.L.); 1188@mju.edu.cn (J.Y.)

Abstract

Underwater images often suffer from color distortion, low contrast, and structural blurring caused by wavelength-dependent absorption and scattering, which degrade both visual observation and downstream perception. Existing underwater image enhancement methods usually learn image-level restoration mappings, while the relationships among semantic regions, degradation patterns, and restoration responses are not fully exploited. In this paper, we propose a Semantic Guidance Network with Degradation Consistency for Underwater Image Enhancement (SGDC-UIE). Specifically, SGDC-UIE first extracts dense semantic responses from a frozen DINOv3 prior and converts them into foreground, boundary, and background region gates. These gates are then used to guide pseudo-physical degradation estimation, producing attenuation-like, transmission-like, illumination, structure, and background-light priors for region-aware restoration. These pseudo-physical priors are learned, bounded conditioning variables rather than calibrated estimates of underwater optical parameters. Based on these degradation conditions, a dual-branch restoration network corrects low-frequency color and illumination degradation while recovering high-frequency structural details through semantic-aware wavelet restoration. The color-restored and structure-restored outputs are further integrated by a degradation-consistent fusion gate, which adaptively balances visual fidelity and task-relevant structure preservation. In addition, grouped supervision with quality-anchor replay stabilizes task-aware fine-tuning and reduces visual-quality drift. Extensive experiments on paired and no-reference underwater enhancement benchmarks, semantic segmentation, and underwater object detection show that SGDC-UIE achieves competitive restoration quality and improves the usability of enhanced images for downstream perception.

Keywords: underwater image enhancement; semantic guidance; degradation consistency; DINOv3; wavelet restoration

Check for updates

Academic Editors: Yan-Tsung Peng, Wenqi Ren, Jingchun Zhou and Qiuping Jiang

Received: 25 June 2026
Revised: 18 July 2026
Accepted: 22 July 2026
Published: 25 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Underwater imaging is degraded by wavelength-dependent absorption, suspended-particle scattering, and nonuniform illumination, which jointly produce color cast, veiling light, low contrast, and blurred boundaries [1]. Related studies of underwater visible-light propagation also show that wavelength-dependent behavior and path loss vary across water environments, further illustrating the environment-dependent nature of underwater optical degradation [2,3]. These effects impair both visual observation and the semantic and geometric cues needed by underwater segmentation and object detection [4,5]. Underwater

image enhancement (UIE) is therefore important not only for improving appearance but also for supporting reliable marine perception.

Existing UIE techniques include model-free, model-based, and learning-based approaches. Model-free methods efficiently adjust contrast, brightness, saturation, or color balance but often lack robustness to complex color casts and spatially varying illumination. Model-based methods compensate for absorption and scattering by estimating transmission, background light, or color attenuation [6,7], although their accuracy depends on the assumed imaging model. Learning-based methods, supported by paired benchmarks such as UIEB [8], have improved restoration through Transformer and physics-guided adversarial designs [9,10], comparative learning and diffusion modeling [11,12], and frequency-selective or pyramid processing [13,14]. Nevertheless, most of these methods optimize an image-level mapping without explicitly coordinating restoration behavior with the spatial organization of underwater scenes.

Underwater degradation is strongly region dependent: background water is dominated by haze, color cast, and scattered light; object interiors require faithful color and texture recovery; and boundaries or small targets are especially vulnerable to contrast and detail loss. Uniform restoration can therefore improve global quality while over-smoothing contours, amplifying background interference, or altering task-relevant details. Recent methods have introduced semantic or task cues into UIE: SGUIE-Net uses semantic attention [15], HUPE employs semantic collaborative learning [16], and downstream-task-inspired UIE incorporates perception-oriented supervision [17]. The remaining gap is not the absence of semantics or degradation modeling in isolation but their weak coupling: semantic cues commonly act as attention maps, auxiliary losses, task constraints, or concatenated features, while degradation estimation and restoration-strength selection remain largely separate or depend mainly on low-level evidence.

Table 1 summarizes this distinction. SGDC-UIE converts the same dense semantic responses into foreground, boundary, and background gates that condition pseudo-physical degradation estimation, frequency-selective restoration, and spatial fusion. DINOv3 is used as a frozen region-partition prior rather than a trainable task branch, providing transferable region and boundary cues without co-adapting to the enhancement objective [18]. Thus, the contribution lies in the explicit coupling among region gates, degradation conditions, restoration branches, and spatial fusion, rather than in the first use of DINOv3, semantic guidance, physical cues, or frequency processing in UIE.

Table 1. Architectural comparison of representative UIE methods.

Method	Semantic Source	Primary Guidance Mechanism	Semantic-Conditioned Degradation Estimation	Frequency-Selective Restoration	Task-Oriented Validation
SGUIE-Net [15]	Task-specific semantic features	Region-aware attention	No	No	Limited
HUPE [16]	Semantic collaborative features	Perceptual feature collaboration	No	No	Yes
DTIUIE [17]	Downstream task networks	Task-inspired data and loss design	No	No	Yes
DeepAqua Enhance [19]	Implicit learned contextual features	Hierarchical attention and adaptive feature fusion	No	No	No
WaterMamba [20]	None	State-space feature modeling	No	No	No
SGDC-UIE (ours)	Frozen DINOv3 dense prior	Restoration-oriented region gates	Yes	Yes	Yes

Based on this linkage, we propose SGDC-UIE, a Semantic Guidance Network with Degradation Consistency for Underwater Image Enhancement. Frozen DINOv3 responses are first transformed into region gates and used to condition a pseudo-physical estimator of attenuation-like, transmission-like, illumination, structure, and background-light priors. A dual-branch network then performs low-frequency color and illumination correction and semantic-aware wavelet restoration of high-frequency structure. A degradation-consistent

fusion gate spatially combines the two outputs, while grouped supervision and quality-anchor replay stabilize task-aware fine-tuning and limit visual-quality drift.

The main contributions are summarized as follows:

1. We introduce SGDC-UIE, which transforms frozen foundation-model responses into foreground, boundary, and background gates and establishes an explicit coupling among semantic regions, degradation conditions, and restoration behavior. This formulation replaces uniform image-level correction with region-sensitive enhancement.
2. We develop a semantic-modulated pseudo-physical estimator and a degradation-conditioned dual-branch restoration network. Low-frequency color and illumination correction is complemented by semantic-aware wavelet recovery of textures and edges, and an adaptive fusion gate integrates the resulting outputs.
3. We evaluate SGDC-UIE on paired and no-reference enhancement benchmarks and under semantic-segmentation and object-detection protocols. Quantitative, visual, ablation, and complexity results examine restoration quality, downstream utility, component contributions, and efficiency.

2. Related Work

2.1. Underwater Image Enhancement

Existing UIE methods can be roughly divided into prior-based methods and learning-based methods. Prior-based methods usually rely on image formation models, haze-line assumptions, histogram redistribution, Retinex-like decomposition, or local contrast adjustment to compensate for absorption and scattering. For example, Berman et al. [21] studied haze-line priors for underwater restoration, and Li et al. [8] constructed UIEB to support reference-based enhancement evaluation. Unsupervised or weakly supervised approaches further reduce the dependence on paired data by introducing medium-transmission cues, homology constraints, or uncertainty-aware restoration objectives [22,23]. These methods are usually interpretable and lightweight, but their performance is sensitive to hand-crafted assumptions and degradation estimation accuracy.

Deep learning has substantially advanced UIE by learning nonlinear mappings between degraded and high-quality images. CNN-based and fusion-based methods improve color correction and visibility through multi-scale feature extraction and adaptive color-gamma adjustment [24]. Domain adaptation and semi-supervised learning reduce the cost of collecting paired underwater data [25,26]. Transformer-based networks improve long-range color and illumination modeling [9], while comparative learning alleviates the bias caused by a single subjective reference [11]. More recent contrastive and diffusion-based methods further improve perceptual restoration and degradation modeling [12,27]. Collectively, these studies cover diverse supervision regimes and representation strategies for visual restoration.

2.2. Degradation-Aware and Frequency-Selective Restoration

Physical and degradation-aware approaches incorporate image-formation knowledge or learned degradation cues into UIE. Cong et al. [10] introduced physical-model guidance into adversarial enhancement, Li et al. [28] considered nonuniform artificial illumination, and Zheng et al. [29] addressed high-turbidity degradation. Transmission-aware Transformers and color-transfer learning further employ degradation-dependent feature modulation [30,31]. Diffusion-based UIE models more complex degradation distributions, although it typically requires heavier inference and additional constraints to limit structure distortion [32,33].

Frequency-selective restoration separates global color and illumination correction from local texture and structure recovery. Wavelet visual perception fusion processes frequency

bands separately [13], whereas wavelet-domain Fourier interaction exchanges complementary frequency information [34]. Low-light wavelet decomposition and pyramid-based UIE demonstrate coarse-to-fine restoration [14,35], and frequency-aware state-space modeling extends this direction [36]. Efficient channel-wise correction has also been explored for real-time enhancement [37]. These studies establish physical modeling and frequency decomposition as complementary tools for underwater restoration.

2.3. Semantic Guidance and Task-Aware Enhancement

Downstream perception provides an additional perspective for evaluating enhanced underwater images. Dense prediction models demonstrate the importance of global context and mask-level attention for semantic understanding [38]. Underwater detection studies quantify how degraded visibility affects target localization and recognition [5,39], while recent task-oriented detectors further address challenging underwater conditions [40].

Semantic and task-aware UIE methods use such information in several ways. SGUIE-Net employs semantic attention for region-aware enhancement [15]; DTIUIE studies the discrepancy between human-preferred and task-preferred enhancement through task-inspired data and supervision [17]; and HUPE introduces semantic collaborative learning for perceptual enhancement [16]. Semantic-prior guidance has also been investigated in image fusion [41]. In these approaches, semantics primarily serve as attention, collaboration, supervision, or evaluation signals.

Table 1 provides a focused architectural comparison. SGUIE-Net uses task-specific semantics for region-aware attention, HUPE performs semantic feature collaboration, and DTIUIE introduces downstream task networks through task-inspired data and losses. DeepAquaEnhance combines hierarchical attention with adaptive feature fusion [19], whereas WaterMamba employs a visual state-space model for long-range representation [20]. In contrast, SGDC-UIE uses an external frozen semantic source to jointly condition degradation estimation and frequency-selective recovery through a shared interface, thereby establishing an explicit semantic–degradation–restoration coupling.

Foundation visual models provide transferable semantic representations that can support this interface. CLIP demonstrates broad visual semantics [42], while DINO and DINOv2 produce dense self-supervised features [43,44]. Segment Anything exhibits strong region-level perception [4,45], and DINOv3 further improves dense feature quality and cross-domain robustness [18]. These developments motivate our selection of frozen DINOv3 as the dense semantic source in SGDC-UIE.

3. Proposed Method

Figure 1 presents the proposed SGDC-UIE. The framework comprises semantic region partition, degradation-condition construction, degradation-consistent dual-branch restoration, adaptive fusion, and grouped optimization with quality-anchor replay. The following subsections detail how frozen DINOv3 responses are converted into region gates, how these gates modulate pseudo-physical priors, and how the resulting conditions guide restoration and fusion.

3.1. Framework Design

Let $I \in \mathbb{R}^{H\times W\times 3}$ denote a degraded underwater image and J denote its reference image when paired supervision is available. Conventional UIE models usually learn an image-level mapping from I to an enhanced image. However, underwater degradation is spatially and semantically nonuniform: background water regions mainly contain haze-like scattering, color cast, and illumination imbalance, whereas foreground objects and boundaries require more careful texture and edge preservation. Therefore, we formulate

UIE as a semantic-guided, degradation-consistent restoration problem. The objective is to learn an enhancer $\mathcal{E}_\theta$ conditioned on a dense semantic prior $\mathcal{S}$:

$$\hat{J} = \mathcal{E}_\theta(I; \mathcal{S}), \tag{1}$$

where $\hat{J}$ is the enhanced image. The key is not to simply concatenate semantic features with restoration features but to transform semantic responses into degradation-aware restoration conditions.

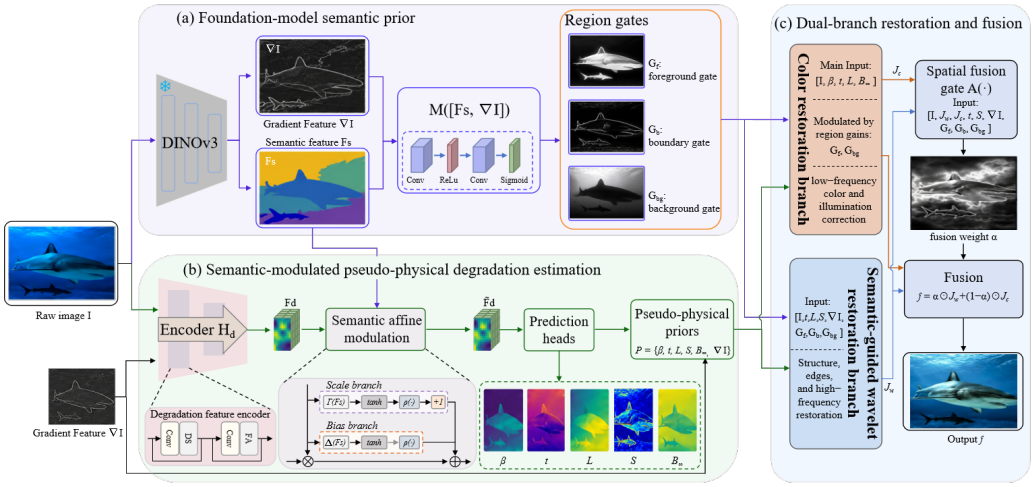


Figure 1. Overall framework of SGDC-UIE.

In this paper, degradation consistency refers to the spatial agreement among semantic region gates, pseudo-physical degradation priors, restoration responses, and fusion weights. A restoration process is considered degradation-consistent when regions dominated by background water, color cast, or scattering receive stronger color/illumination correction, while foreground objects and boundary regions receive stronger structure-preserving restoration. In other words, the enhancement behavior should follow the estimated local degradation state rather than applying a uniform correction to all pixels.

As shown in Figure 1, frozen DINOv3 responses are first converted into foreground, boundary, and background gates. These gates guide a pseudo-physical estimator to construct degradation conditions related to attenuation, transmission, illumination, structure, and background light. The resulting conditions drive two complementary restoration branches: the color branch corrects low-frequency color and illumination degradation, while the wavelet branch recovers high-frequency structure and boundary details. A degradation-consistent fusion gate then integrates the two outputs according to semantic regions and degradation states.

The overall forward process is summarized as

$$\begin{aligned} (P_s, F_s) &= \mathcal{T}_\phi(I), \\ \mathcal{G}_s &= \mathcal{G}(F_s, \nabla I), \\ \mathcal{P} &= \mathcal{D}(I, \nabla I, F_s, \mathcal{G}_s), \\ J_c &= \mathcal{C}(I, \mathcal{P}, \mathcal{G}_s), \\ J_w &= \mathcal{W}(I, \mathcal{P}, \mathcal{G}_s), \\ \hat{J} &= \mathcal{F}(I, J_c, J_w, \mathcal{P}, \mathcal{G}_s), \end{aligned} \tag{2}$$

where $\mathcal{T}_\phi$ is the frozen semantic-prior extractor, P_s and F_s denote semantic responses and intermediate semantic features, $\mathcal{G}_s$ denotes semantic region gates, $\mathcal{P}$ denotes pseudo-

physical degradation priors, J_c and J_w are the color-restored and wavelet-restored results, respectively, and $\mathcal{F}$ is the degradation-consistent fusion module.

Semantic region partition converts dense semantic responses into soft restoration-aware gates. Unlike hard segmentation masks, these gates describe where different enhancement behaviors should be emphasized: the foreground gate protects object appearance, the boundary gate preserves edge and texture structures, and the background gate guides haze and color correction in water regions. Degradation condition construction then transforms these semantic regions into pseudo-physical priors. Instead of predicting exact physical parameters, SGDC-UIE estimates interpretable degradation conditions related to attenuation-like color degradation, transmission-like visibility, illumination distribution, structural degradation, and background light. These conditions guide the color branch toward low-frequency correction and the wavelet branch toward high-frequency structure recovery. Their outputs are fused by a degradation-consistent gate conditioned on semantic regions and degradation priors.

3.2. Semantic-Guided Degradation Condition Construction

3.2.1. Foundation Semantic Region Prior

The semantic-to-degradation process starts from a stable semantic prior. Directly training a lightweight segmentation model together with the enhancer may produce unstable responses, because underwater images often contain severe color cast, turbidity, and weak boundaries. To avoid a moving semantic target, we employ a DINOv3-based semantic-prior extractor and keep it frozen during enhancer training:

$$(P_s, F_s) = \mathcal{T}_\phi(I). \tag{3}$$

where $P_s \in \mathbb{R}^{H\times W\times K}$ denotes the resized dense semantic response over K categories, and $F_s \in \mathbb{R}^{C_s\times H\times W}$ denotes the compact semantic feature used for region gating and degradation modulation. More specifically, the semantic-prior extractor consists of a frozen DINOv3 backbone $\mathcal{B}_\phi$ and a lightweight dense decoder $\mathcal{H}_s$:

$$\begin{aligned} Z_s &= \mathcal{B}_\phi(I), \\ (P_s, F_s) &= \mathrm{Up}(\mathcal{H}_s(Z_s)), \end{aligned} \tag{4}$$

where Z_s denotes the patch-level foundation feature and $\mathrm{Up}(\cdot)$ resizes dense responses to the image resolution. The decoder returns both segmentation logits and the compact feature before the final classifier. The dense decoder is adapted to underwater semantic annotations before enhancer optimization, while the DINOv3 backbone remains frozen. During SGDC-UIE training, the semantic-prior network is used as a fixed condition provider/evaluator, and gradients from the restoration loss are not propagated into it. This setting gives the enhancer a stable semantic coordinate system and avoids semantic drift during enhancement optimization.

Category-level responses alone are insufficient for restoration, because UIE also needs to know where color correction, boundary enhancement, and background haze suppression should be emphasized. Therefore, the semantic feature is resized and combined with the image gradient to generate three soft gates:

$$\mathcal{G}_s = \{G_f, G_b, G_{bg}\} = \sigma(\mathcal{M}([F_s, \nabla I])), \tag{5}$$

where G_f, G_b, and G_{bg} denote foreground, boundary, and background gates, respectively. $\mathcal{M}(\cdot)$ is a lightweight convolutional mapping and $\sigma(\cdot)$ denotes the sigmoid function. The image gradient ∇I enhances boundary awareness. These gates are continuous rather

than binary, which allows the model to handle uncertain object boundaries and turbid transition regions.

The three gates are designed for restoration rather than for semantic labeling. G_f highlights object-like regions whose appearance should be preserved during enhancement; G_b emphasizes object contours and texture transitions that are easily blurred by underwater scattering; G_{bg} describes water-background regions where color cast, veiling light, and nonuniform illumination usually dominate. Compared with directly feeding P_s into the restoration network, the gate representation compresses high-dimensional semantic responses into three restoration-oriented maps. This design avoids making the enhancer depend on exact class labels and makes the semantic prior usable on underwater scenes with category ambiguity.

3.2.2. Semantic-Modulated Pseudo-Physical Priors

Underwater imaging is often described by an image formation model:

$$I^c(x) = J^c(x)t^c(x) + B_\infty^c(1 - t^c(x)), \quad c \in \{r, g, b\}, \tag{6}$$

where $t^c(x)$ is the transmission and B_∞^c is the background light. Estimating exact physical parameters from a single image is ill-posed under nonuniform illumination and mixed water conditions. Therefore, we learn a set of pseudo-physical priors:

$$\mathcal{P} = \{\beta, t, L, S, B_\infty, \nabla I\}. \tag{7}$$

We define a pseudo-physical prior as a learned, bounded degradation condition whose functional role is motivated by an underwater image-formation factor but whose value is not interpreted as a calibrated physical measurement. Specifically, β, t, L, S, and B_∞ describe attenuation-like color degradation, transmission-like visibility, illumination distribution, structure degradation, and background light, respectively. They are optimized as unitless soft conditions without requiring calibrated depth, absorption/scattering coefficients, or inherent optical properties. In implementation, $\beta \in [0,1]^{H\times W\times 3}$, $t, L, S \in [0,1]^{H\times W\times 1}$, and $B_\infty \in [0,1]^{1\times 1\times 3}$ is broadcast to the image resolution. The deterministic gradient ∇I is appended only as an edge cue.

The degradation encoder first extracts low-level degradation features:

$$F_d = \mathcal{H}_d([I, \nabla I]). \tag{8}$$

Since degradation characteristics differ across semantic regions, F_d is modulated by the frozen semantic feature:

$$\tilde{F}_d = F_d \odot (1 + \rho \tanh(\Gamma(F_s))) + \rho \tanh(\Delta(F_s)), \tag{9}$$

where $\Gamma(\cdot)$ and $\Delta(\cdot)$ predict semantic scale and bias terms, and ρ controls the modulation strength. The pseudo-physical priors are then predicted by independent bounded heads:

$$\{\beta, t, L, S, B_\infty\} = \mathcal{H}_p(\tilde{F}_d). \tag{10}$$

Equivalently, each prior is obtained by

$$\begin{aligned} \beta &= \sigma(\mathcal{H}_\beta(\tilde{F}_d)), \quad t = \sigma(\mathcal{H}_t(\tilde{F}_d)), \\ L &= \sigma(\mathcal{H}_L(\tilde{F}_d)), \quad S = \sigma(\mathcal{H}_S(\tilde{F}_d)), \\ B_\infty &= \sigma(\mathcal{H}_B(\mathrm{GAP}(\tilde{F}_d))). \end{aligned} \tag{11}$$

The independent-head design follows the fact that color attenuation, visibility, illumination, and structure degradation have different spatial patterns. Sharing the encoder preserves efficiency, while using separate heads avoids forcing all degradation factors to be represented by a single mixed map. This construction connects semantic regions with degradation priors. For example, a bluish object and a bluish water background may have similar RGB values, but they should receive different correction strengths. The semantic modulation helps the estimator distinguish such cases and produces region-adaptive restoration conditions. In practice, the multiplicative term in Equation (9) changes the response strength of degradation features, while the additive term shifts the feature distribution according to the semantic context. Thus, the estimator can express different degradation states for foreground objects, boundaries, and open-water backgrounds even when their local RGB statistics are close.

The predicted priors are subsequently used with different roles. β, t, L, and B_∞ guide low-frequency color and illumination correction; S, G_f, and G_b guide structure-sensitive restoration; and ∇I provides an explicit edge cue for both wavelet refinement and degradation-consistent fusion. Therefore, the pseudo-physical estimator serves as an intermediate bridge between high-level semantic partition and low-level restoration operators. During optimization, t and B_∞ enter the formation-consistency loss directly, t and L enter the edge-aware smoothness loss, whereas β and S are learned through the restoration, frequency, and task objectives rather than parameter-level targets.

3.3. Degradation-Consistent Dual-Branch Restoration

The estimated degradation priors drive two complementary branches. This design follows the observation that underwater degradation affects low- and high-frequency contents differently: color cast and illumination imbalance mainly appear as low-frequency changes, while boundary blur and texture attenuation mainly appear in high-frequency components.

3.3.1. Region-Guided Color Restoration

The color restoration branch is shown in Figure 2. It receives the raw image and the pseudo-physical priors related to color and illumination. We first form the color-condition tensor

$$X_c = [I, \beta, t, L, B_\infty], \tag{12}$$

where single-channel maps are broadcast or concatenated at the image resolution. The residual predictor is implemented as a compact convolutional residual mapping:

$$\begin{aligned} F_c^0 &= \delta(\mathrm{Conv}_{3\times3}(X_c)), \\ F_c^m &= F_c^{m-1} + \mathcal{B}_c^m(F_c^{m-1}), \quad m = 1, \ldots, M_c, \\ \Delta_c &= \tanh(\mathrm{Conv}_{1\times1}(\mathrm{CA}(F_c^{M_c}))), \end{aligned} \tag{13}$$

where $\delta(\cdot)$ denotes a nonlinear activation, $\mathcal{B}_c^m$ denotes a residual convolutional block, and $\mathrm{CA}(\cdot)$ denotes channel attention. For compactness, Equation (13) is denoted as

$$\Delta_c = \mathcal{R}_c([I, \beta, t, L, B_\infty]), \tag{14}$$

where Δ_c is a bounded three-channel color residual. To avoid over-correcting object textures while still suppressing background color cast, a semantic region gain is predicted by

$$A_c = \mathcal{A}_c(G_{bg}, G_f). \tag{15}$$

The color-restored result is obtained as

$$J_c = \mathrm{clip}(I + \eta_c A_c \odot \Delta_c, 0, 1), \tag{16}$$

where η_c controls the residual strength. The branch predicts a residual image instead of generating the restored image from scratch, which keeps the correction close to the input structure and reduces the risk of producing artificial colors.

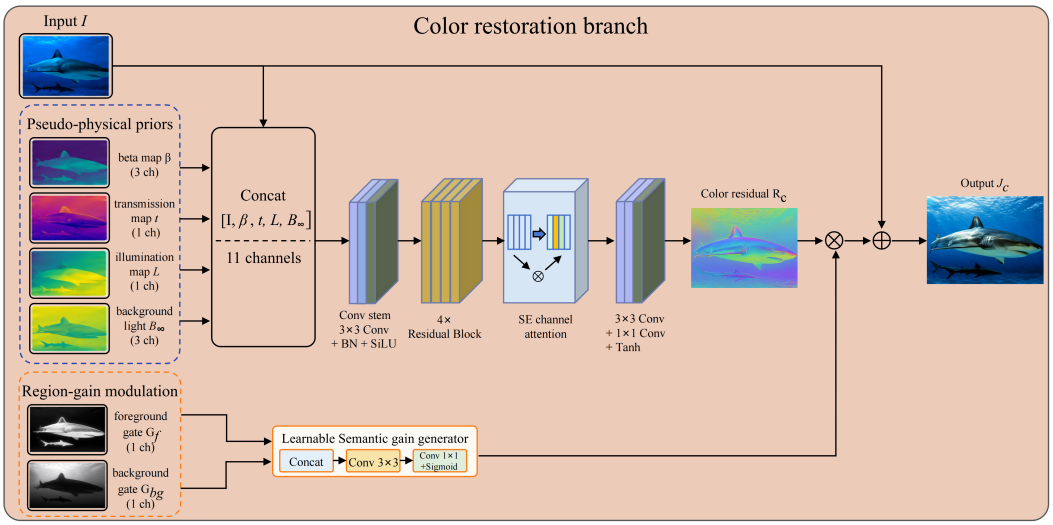


Figure 2. Region-guided color restoration branch.

Conditioning the correction on pseudo-physical priors allows the branch to respond differently to attenuation, transmission, illumination, and background-light variations. The region gain modulates the residual according to foreground and background regions, producing stronger compensation in water-background areas and milder correction on object surfaces. The bounded residual form in Equation (16) further prevents excessive color shifts and keeps the branch computationally lightweight. As a result, J_c provides a stable low-frequency restoration candidate for the later fusion stage.

3.3.2. Semantic-Aware Wavelet Restoration

The semantic-aware wavelet restoration branch is shown in Figure 3. The input image is decomposed into one low-frequency subband and three high-frequency subbands:

$$(I_{LL}, I_{LH}, I_{HL}, I_{HH}) = \mathcal{D}_w(I), \tag{17}$$

where I_{LL} preserves coarse color and illumination information, and I_{LH}, I_{HL}, and I_{HH} contain directional edge and texture responses. We denote $I_{HF} = [I_{LH}, I_{HL}, I_{HH}]$. Compared with processing RGB features only in the spatial domain, the wavelet representation explicitly separates color-illumination correction from detail restoration, which is suitable for underwater images where color cast and texture degradation often appear at different frequency ranges.

For the low-frequency subband, restoration is conditioned on background and illumination-related priors:

$$\begin{aligned} A_{LL} &= \mathcal{A}_{LL}(G_{bg}, t), \\ \hat{I}_{LL} &= I_{LL} + \eta_{LL} A_{LL} \odot \mathcal{R}_{LL}([I_{LL}, t, L, G_{bg}]). \end{aligned} \tag{18}$$

The low-frequency gate A_{LL} is mainly driven by background and visibility cues. It helps the branch correct coarse haze and illumination imbalance without treating the entire

image uniformly. For the high-frequency subbands, restoration is guided by foreground, boundary, and structure degradation cues:

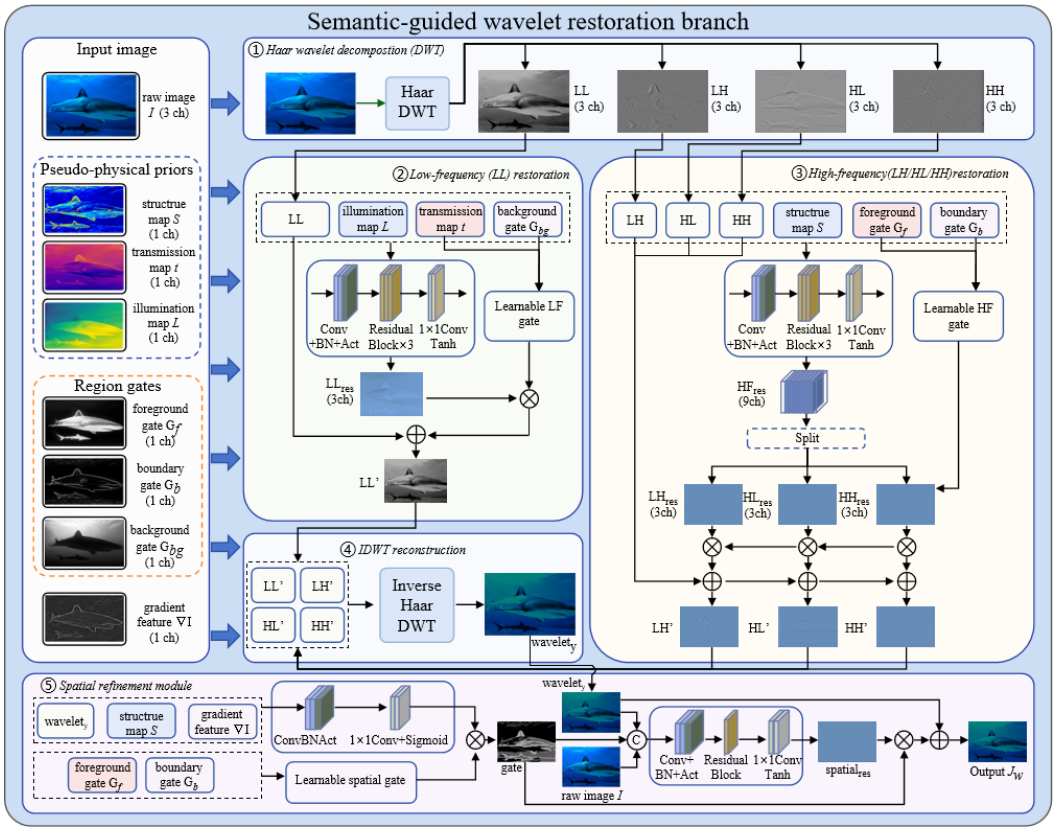


Figure 3. Semantic-aware wavelet restoration branch.

$$\begin{aligned} A_{HF} &= \mathcal{A}_{HF}(G_b, G_f), \\ \hat{I}_{HF} &= I_{HF} + \eta_{HF} A_{HF} \odot \mathcal{R}_{HF}([I_{HF}, G_f, G_b, S]). \end{aligned} \tag{19}$$

The high-frequency gate A_{HF} emphasizes edges and object-related textures. Instead of amplifying all high-frequency coefficients, the branch uses G_b, G_f, and S to focus on structurally meaningful regions and suppress noise-like details in smooth water backgrounds. The restoration mappings $\mathcal{R}_{LL}$ and $\mathcal{R}_{HF}$ are lightweight residual mappings with separate parameters because low-frequency color compensation and high-frequency detail recovery require different filters. The gate generators $\mathcal{A}_{LL}$ and $\mathcal{A}_{HF}$ use sigmoid outputs, so the restoration strength remains spatially bounded. The restored subbands are reconstructed by the inverse Haar transform:

$$J_w^0 = \mathcal{D}_w^{-1}(\hat{I}_{LL}, \hat{I}_{LH}, \hat{I}_{HL}, \hat{I}_{HH}). \tag{20}$$

To suppress subband artifacts, a spatial refinement gate is further predicted:

$$A_w = \mathcal{A}_w([J_w^0, S, \nabla I, G_f, G_b]). \tag{21}$$

The final wavelet-restored result is

$$J_w = \mathrm{clip}\left(J_w^0 + \eta_w A_w \odot \mathcal{R}_s([I, J_w^0, A_w]), 0, 1\right). \tag{22}$$

Thus, the wavelet branch does not uniformly amplify high-frequency responses. Background gates mainly influence low-frequency correction, while foreground and boundary gates guide high-frequency structure recovery. The low- and high-frequency paths cor-

respond to different underwater degradation factors, making the restoration behavior interpretable. Semantic gates prevent the high-frequency path from enhancing background noise as texture, and the spatial refinement step compensates for possible subband reconstruction artifacts to produce a structure-preserving candidate J_w.

3.4. Degradation-Consistent Adaptive Fusion

The color branch and wavelet branch have complementary roles. The former provides stable global correction, whereas the latter recovers task-sensitive structures. Directly averaging them would ignore spatially varying restoration needs. Therefore, SGDC-UIE predicts a degradation-consistent fusion gate as shown in Figure 4. The fusion-condition tensor is

$$X_f = [I, J_c, J_w, t, S, \nabla I, G_f, G_b, G_{bg}], \tag{23}$$

and the fusion gate is produced by a shallow spatial mapping:

$$\alpha = \sigma(\mathrm{Conv}_{1\times1}(\Phi_f(X_f))), \tag{24}$$

where $\Phi_f(\cdot)$ is composed of convolutional blocks and a residual block, and $\alpha \in [0,1]^{H\times W}$ represents the contribution of the wavelet branch. The final enhanced image is

$$\hat{J} = \alpha \odot J_w + (1-\alpha) \odot J_c. \tag{25}$$

Since the fusion gate receives both degradation priors and semantic region gates, it can rely more on J_w near boundaries and texture-rich foreground regions, while assigning larger weights to J_c in smooth background water regions. This degradation-consistent fusion differs from simple summation or channel concatenation because the two restoration candidates are selected locally according to the estimated degradation state. The final output therefore inherits the stable color correction of J_c and the semantic-structure preservation of J_w.

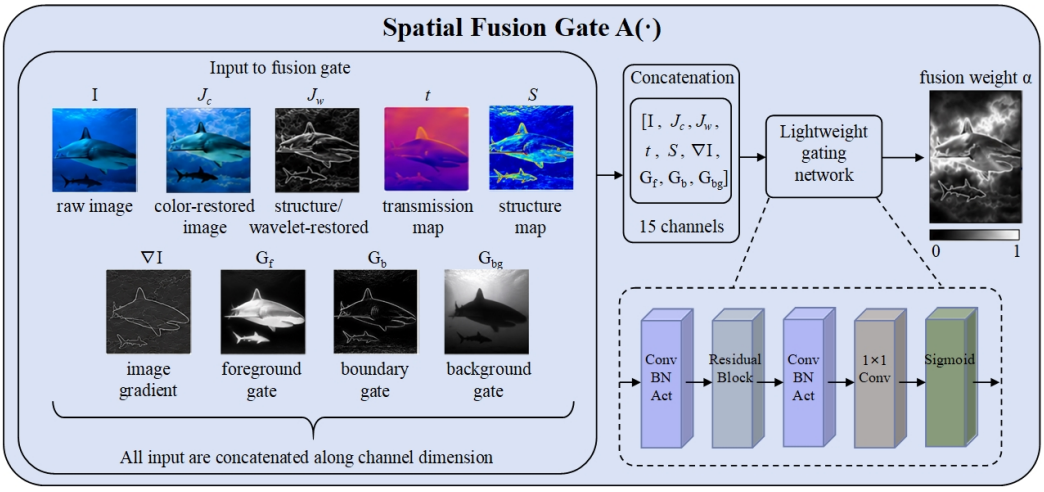


Figure 4. Degradation-consistent fusion gate.

3.5. Joint Optimization and Quality-Anchor Replay

The training objective is organized according to the roles of different supervision signals. Instead of listing many unrelated losses, we group them into restoration fidelity, degradation consistency, frequency consistency, task consistency, lightweight quality regularization, and quality-anchor replay:

$$\mathcal{L}_{total} = \lambda_r \mathcal{L}_{res} + \lambda_p \mathcal{L}_{phys} + \lambda_f \mathcal{L}_{freq} + \lambda_t \mathcal{L}_{task} + \lambda_i \mathcal{L}_{iqa} + \lambda_q \mathcal{L}_{qa}. \tag{26}$$

This grouped form makes the optimization target consistent with the framework design. $\mathcal{L}_{res}$ supervises image-level fidelity, $\mathcal{L}_{phys}$ regularizes the pseudo-physical degradation representation, $\mathcal{L}_{freq}$ constrains wavelet-domain restoration, $\mathcal{L}_{task}$ preserves semantic responses after enhancement, $\mathcal{L}_{iqa}$ provides lightweight no-reference quality guidance, and $\mathcal{L}_{qa}$ prevents quality degradation during task-aware fine-tuning.

The restoration loss supervises the enhanced result with reference images:

$$\mathcal{L}_{res} = \mathcal{L}_1 + \lambda_{ssim}\mathcal{L}_{ssim} + \lambda_{col}\mathcal{L}_{col} + \lambda_{str}\mathcal{L}_{str}. \tag{27}$$

The fidelity terms are

$$\mathcal{L}_1 = \|\hat{J} - J\|_1, \quad \mathcal{L}_{ssim} = 1 - \mathrm{SSIM}(\hat{J}, J). \tag{28}$$

The color constancy term penalizes channel imbalance:

$$\begin{aligned} \mathcal{L}_{col} &= \sum_{(p,q)\in\Omega} \left(\mu_p(\hat{J}) - \mu_q(\hat{J})\right)^2, \\ \Omega &= \{(r,g),(r,b),(g,b)\}, \end{aligned} \tag{29}$$

and the structure loss preserves gradient consistency:

$$\mathcal{L}_{str} = \|\nabla\hat{J} - \nabla J\|_1. \tag{30}$$

The pseudo-physical group encourages the estimated priors to remain consistent with the input formation process:

$$\mathcal{L}_{phys} = \mathcal{L}_{phy} + \lambda_{sm}\mathcal{L}_{sm}, \tag{31}$$

where

$$\mathcal{L}_{phy} = \left\| I - \left(\hat{J} \odot t + B_\infty(1-t)\right)\right\|_1. \tag{32}$$

$\mathcal{L}_{sm}$ denotes edge-aware smoothness regularization for the transmission- and illumination-like maps:

$$\mathcal{L}_{sm} = \sum_{p\in\{t,L\}} \left(|\nabla_x p| e^{-10|\nabla_x I_g|} + |\nabla_y p| e^{-10|\nabla_y I_g|}\right). \tag{33}$$

Here, I_g is the grayscale version of I. This edge-aware form discourages noisy degradation maps in smooth regions while allowing discontinuities near image edges. The frequency-domain group is

$$\mathcal{L}_{freq} = \mathcal{L}_{wav}, \tag{34}$$

where $\mathcal{L}_{wav}$ constrains wavelet subbands to prevent excessive high-frequency amplification:

$$\mathcal{L}_{wav} = \omega_{ll}\|\hat{J}_{LL} - J_{LL}\|_1 + \omega_{hf}\frac{1}{3}\sum_{k\in\{LH,HL,HH\}} \|\hat{J}_k - J_k\|_1 + \omega_e\mathcal{L}_{edge}. \tag{35}$$

Here, $\hat{J}_k$ and J_k are Haar subbands of the enhanced and reference images, respectively, and $\mathcal{L}_{edge}$ is a hinge term that prevents the high-frequency energy of $\hat{J}$ from falling below that of the reference by a preset margin. This loss encourages the branch to recover meaningful structures instead of merely increasing local contrast.

To preserve semantic responses after enhancement, we use the frozen semantic-prior network as a task-consistency evaluator:

$$\mathcal{L}_{task} = \|\Psi(\hat{J}) - \Psi(J)\|_1 + \mathbf{1}_{Y_s}\mathrm{CE}(\mathcal{T}_\phi(\hat{J}), Y_s), \tag{36}$$

where $\Psi(\cdot)$ denotes intermediate semantic features, Y_s is the semantic label when available, and $\mathbf{1}_{Y_s}$ indicates that the cross-entropy term is applied only to samples with semantic annotations. Since $\mathcal{T}_\phi$ is frozen, gradients are propagated only to the enhancer. This loss does not require the enhancer to produce segmentation masks. It only encourages the enhanced image to retain semantic responses that are useful for perception, which is important for underwater scenes where visually pleasing enhancement may still distort object boundaries or task-relevant textures.

The quality regularization $\mathcal{L}_{iqa}$ is implemented with differentiable proxy terms related to underwater colorfulness, contrast, exposure, saturation, and gradient energy. It is used as a weak regularizer rather than as the primary optimization target, preventing the network from pursuing no-reference scores at the cost of paired fidelity.

Task-aware optimization can improve perception but may cause visual-quality drift. We therefore introduce quality-anchor replay during fine-tuning:

$$\mathcal{L}_{qa} = \mathcal{L}_{res}(\hat{J}_q, J_q) + \lambda_{qt}\mathcal{L}_{task}(\hat{J}_q, J_q) + \lambda_{qw}\mathcal{L}_{wav}(\hat{J}_q, J_q) + \lambda_{qi}\mathcal{L}_{iqa}(\hat{J}_q, J_q), \tag{37}$$

where (I_q, J_q) denotes a replayed quality-anchor pair and $\mathcal{L}_{iqa}$ denotes lightweight no-reference quality regularization used during quality replay. This strategy keeps the model anchored to reference-based enhancement quality while optimizing semantic structure preservation. During fine-tuning, paired enhancement samples are periodically replayed with the quality-anchor loss so that semantic supervision does not dominate the restoration objective. The final objective therefore balances visually faithful restoration, interpretable degradation consistency, and downstream perception utility.

4. Experimental Results

In this section, we first describe the experimental settings, including the adopted datasets, implementation details, evaluation metrics, and comparison methods. We then report the enhancement quality comparisons, downstream perception evaluations, and ablation studies. All methods are evaluated in terms of both low-level restoration quality and perception-oriented downstream performance.

4.1. Implementation Details

4.1.1. Datasets

We evaluate SGDC-UIE on paired enhancement benchmarks, real-world unpaired underwater images, and downstream perception datasets. For full-reference enhancement evaluation, UIEB [8], SUIM-E [15], and LSUI [9] are adopted because they provide degraded/reference pairs and cover different underwater appearances, including color cast, turbidity, illumination attenuation, and structure degradation. The three paired datasets are reported separately rather than merged, so that dataset-specific behavior can be observed. To examine generalization on real-world underwater images without reference targets, we further use the RUIE-UCCS and RUIE-UIQS subsets [46]. For perception-oriented evaluation, TIUIED [17] is used for semantic segmentation, while Brackish [47] is used for underwater object detection. We follow the official training and testing splits when they are available; otherwise, the established splits used by the corresponding baselines are kept unchanged. For every dataset, the test image list is fixed before evaluation, and the same test images are used for all compared methods.

4.1.2. Training Details

All images are resized to 384×384 during training. The lightweight dense prediction head of the semantic-prior network is first adapted to underwater semantic categories, while the DINOv3 backbone remains frozen throughout. The decoder is adapted once using

the 1525 pixel-annotated images in the TIUIED training split, whose masks cover water body, fish and vertebrates, reefs and invertebrates, aquatic plants and sea grass, wrecks and ruins, human divers, robots/instruments, and sea floor and rocks. The 110-image test split is held out; only the decoder's 0.591M parameters are optimized, and no semantic labels, water-type labels, or camera metadata are required at inference. The resulting semantic-prior network is then used as a fixed semantic condition provider. The enhancement network is optimized in two stages. In the first stage, SGDC-UIE learns paired image restoration from reference-based enhancement data. In the second stage, the model is fine-tuned with task-aware semantic supervision and quality-anchor replay, so that semantic consistency is improved without drifting away from reference-based enhancement quality. The network is trained with the AdamW optimizer and mixed precision using a batch size of 4. In the task-aware fine-tuning stage, the learning rate is set to 5×10^{-6} and the weight decay is 10^{-3}. Unless otherwise stated, all experiments are implemented in PyTorch 2.8.0 and conducted on an NVIDIA RTX 4090 GPU. For the repeatability analysis, we additionally train the complete DINOv3 configuration with random seeds 2023, 2024, and 2025 under the same settings. The alternative semantic-prior configurations are controlled single-run sensitivity experiments and are therefore interpreted descriptively rather than as repeated-training significance tests.

4.1.3. Evaluation Metrics

We evaluate enhancement using reference-based PSNR, SSIM, PCQI, and LPIPS, together with no-reference UIQM, NIQE, and BRISQUE. UCIQE summarizes colorfulness, saturation, and contrast in the CIELab space, while AG and EI measure detail and edge sharpness. Segmentation is assessed by PixAcc, mIoU, Dice, FG-Dice, and RI-IoU; PixAcc measures pixel-wise accuracy, FG-Dice measures foreground overlap, and RI-IoU measures overlap over task-relevant regions. Detection metrics include precision, recall, mAP50, mAP75, and mAP50:95. All outputs are resized and evaluated using the same scripts. Learned UIQA methods such as PUIQA and EDANet provide complementary perceptual assessment [48,49]; PAUQA-Net similarly learns perceptual quality from subjective annotations [50]. Because their scores depend on dataset-specific annotations, checkpoints, and preprocessing, and no common calibrated model covers all evaluated datasets, we retain established metrics for a reproducible comparison.

4.1.4. Comparison Methods

We compare SGDC-UIE with representative underwater enhancement methods from different technical families. The CNN-based methods include DICAM [51], AdaptiveStd [52], and CLUIE-Net [11]. The physics- or uncertainty-guided methods include PUGAN [10] and PUIE-Net [23]. The Transformer- and frequency-domain methods include UTUIE [9], WF-Diff [34], and UVE-Net [53]. We also include task- or perception-oriented methods, i.e., DTIUIE [17] and HUPE [16]. For methods with released models, we use the official checkpoints and inference settings whenever possible; otherwise, we follow the public implementation and keep the input resolution and evaluation scripts consistent with ours.

4.2. Comparison on Enhancement Quality

4.2.1. Quantitative Comparisons

We first compare SGDC-UIE with representative UIE methods on paired and no-reference underwater datasets. Table 2 reports full-reference quality on UIEB, SUIM-E, and LSUI, while Table 3 evaluates real-world generalization on RUIE-UCCS and RUIE-UIQS. These datasets cover different degradation patterns, including bluish and greenish color casts, low contrast, nonuniform illumination, and weak structural details. Therefore, the

comparison reflects not only pixel fidelity on paired benchmarks but also robustness under unpaired real-world underwater scenes.

Table 2. Full-reference enhancement comparison on UIEB, SUIM-E, and LSUI. Red, green, and blue indicate the first-, second-, and third-best results, respectively. (↑) indicates that higher values are better, (↓) indicates that lower values are better.

Method	UIEB				SUIM-E				LSUI			
	PSNR ↑	SSIM ↑	PCQI ↑	LPIPS ↓	PSNR ↑	SSIM ↑	PCQI ↑	LPIPS ↓	PSNR ↑	SSIM ↑	PCQI ↑	LPIPS ↓
DICAM	22.024	0.786	0.812	0.213	21.193	0.792	0.825	0.209	20.711	0.809	0.843	0.290
AdaptiveStd	10.877	0.453	0.536	0.754	10.723	0.449	0.540	0.714	12.041	0.424	0.487	0.813
CLUIE-Net	9.233	0.264	0.279	0.524	8.917	0.237	0.238	0.535	9.491	0.200	0.218	0.640
PUGAN	23.488	0.761	0.795	0.276	20.663	0.755	0.795	0.271	20.446	0.789	0.824	0.324
PUIE-Net	21.064	0.824	0.864	0.152	22.355	0.875	0.914	0.109	20.550	0.816	0.850	0.270
UTUIE	20.498	0.684	0.713	0.289	21.936	0.762	0.791	0.255	24.940	0.828	0.850	0.244
WF-Diff	21.633	0.871	0.904	0.137	20.111	0.856	0.901	0.156	22.645	0.839	0.869	0.218
UVE-Net	19.613	0.825	0.867	0.188	20.214	0.847	0.889	0.154	20.191	0.815	0.849	0.282
DTIUIE	20.503	0.754	0.797	0.253	20.484	0.782	0.821	0.222	21.156	0.816	0.852	0.296
HUPE	21.605	0.861	0.894	0.138	21.073	0.862	0.899	0.142	20.724	0.840	0.872	0.253
Ours	23.519	0.863	0.908	0.133	22.566	0.866	0.921	0.130	22.499	0.848	0.879	0.224

Table 2 shows that SGDC-UIE provides competitive and well-balanced performance on UIEB. Its PSNR of 23.519 dB is numerically the highest, but it exceeds PUGAN's 23.488 dB by only 0.031 dB; this narrow margin should not be interpreted by itself as a statistically established PSNR advantage. The stronger evidence lies in the joint metric profile: SGDC-UIE obtains a PCQI of 0.908 and an LPIPS of 0.133, compared with 0.795 and 0.276 for PUGAN, while its SSIM is also higher (0.863 versus 0.761). WF-Diff achieves the best SSIM, but its PSNR is lower and, as shown later in Table 4, it requires substantially heavier computation. On SUIM-E, SGDC-UIE ranks first in PSNR and PCQI with 22.566 dB and 0.921, respectively. Compared with perception-oriented baselines such as DTIUIE and HUPE, our method provides stronger paired fidelity while maintaining competitive perceptual similarity. On the larger LSUI benchmark, UTUIE obtains the highest PSNR, whereas SGDC-UIE achieves the best SSIM and PCQI and the second-best LPIPS. We therefore interpret the results as evidence of a favorable cross-metric balance rather than categorical superiority based on a single small numerical gap.

Table 3. No-reference enhancement comparison on RUIE-UCCS and RUIE-UIQS. Red, green, and blue indicate the first-, second-, and third-best results, respectively. (↑) indicates that higher values are better.

Method	RUIE-UCCS					RUIE-UIQS				
	UIQM↑	NIQE ↑	BRISQUE ↑	AG↑	EI↑	UIQM ↑	NIQE ↑	BRISQUE ↑	AG↑	EI↑
DICAM	3.037	5.617	39.226	0.018	0.163	3.029	5.713	40.782	0.020	0.184
AdaptiveStd	1.738	8.815	61.008	0.004	0.038	1.691	8.590	61.278	0.004	0.038
CLUIE-Net	0.755	7.955	59.066	0.012	0.089	0.656	7.650	61.980	0.011	0.086
PUGAN	3.138	5.761	39.174	0.020	0.186	3.101	5.798	40.160	0.022	0.205
PUIE-Net	3.073	5.260	34.146	0.021	0.174	3.065	5.209	35.569	0.022	0.191
UTUIE	3.123	5.419	39.124	0.020	0.180	3.029	5.643	40.882	0.021	0.192
WF-Diff	2.825	5.370	34.938	0.020	0.169	2.873	5.337	36.071	0.021	0.182
UVE-Net	2.800	5.597	36.860	0.016	0.137	2.813	5.626	38.713	0.018	0.153
DTIUIE	3.197	5.446	40.418	0.020	0.184	3.206	5.539	41.416	0.022	0.200
HUPE	3.123	4.994	32.691	0.023	0.191	3.093	5.016	34.170	0.025	0.213
Ours	3.231	5.233	32.718	0.024	0.203	3.221	5.296	34.167	0.025	0.217

Table 3 further evaluates real-world unpaired scenes, where reference images are unavailable and no-reference metrics are more suitable. On RUIE-UCCS, SGDC-UIE

achieves the best UIQM, AG, and EI scores, and ranks second in both NIQE and BRISQUE. On RUIE-UIQS, it again achieves the best UIQM, BRISQUE, AG, and EI scores, while remaining among the top three in NIQE. HUPE obtains strong NIQE and BRISQUE values, which indicates good natural image statistics, but SGDC-UIE provides stronger underwater quality and edge/detail measurements. These results suggest that the proposed method improves visibility and preserves local structural details without relying on excessive contrast amplification, which is important for real underwater images whose references are unavailable.

Table 4. Model complexity and inference efficiency. Red, green, and blue indicate the first-, second-, and third-best results, respectively. (↑) indicates that higher values are better.

Method	Size	Params ↑ (M)	FLOPs ↑ (G)	Latency ↑ (ms)	FPS ↑	Memory ↑ (MB)
DICAM	384	0.944	119.561	14.301	69.924	2677.445
AdaptiveStd	384	0.755	110.189	30.995	32.263	230.955
CLUIE-Net	384	13.395	70.037	4.268	234.311	358.304
PUGAN †	256	189.154	73.999	14.066	71.092	808.792
PUIE-Net	384	16.116	52.796	4.873	205.218	380.470
UTUIE †	256	31.590	2.983	18.272	54.729	219.302
WF-Diff	384	100.552	2671.833	266.450	3.753	1048.694
UVE-Net	384	4.525	105.316	12.719	78.620	147.866
DTIUIE	384	50.821	959.165	109.828	9.105	1260.620
HUPE	384	12.180	86.409	73.297	13.643	261.809
Ours-R	384	0.394	41.919	10.028	99.719	141.876
Ours-Full	384	22.572	54.767	11.539	86.666	237.296

† PUGAN and UTUIE are measured at their native 256 × 256 wrapper size due to constraints in the released implementations. Ours-R denotes the restoration module and Ours-Full the complete end-to-end pipeline.

4.2.2. Visual Comparisons

We then provide visual comparisons to examine whether the quantitative gains are consistent with visual appearance. Figure 5 compares enhanced images and their edge-density distributions on paired UIEB and SUIM-E samples. The red and blue curves denote the reference and enhanced images, respectively. A closer overlap between the two curves indicates that the enhanced result better follows the structural statistics of the reference. DICAM tends to leave residual haze in low-contrast regions, while PUGAN may introduce over-saturated colors and stronger contrast than the reference. WF-Diff and HUPE recover certain local details, but their color or edge-density distributions can still deviate from the reference. In contrast, SGDC-UIE maintains more stable color correction and produces edge distributions closer to the reference, suggesting that semantic-guided degradation modeling helps preserve structure while correcting color.

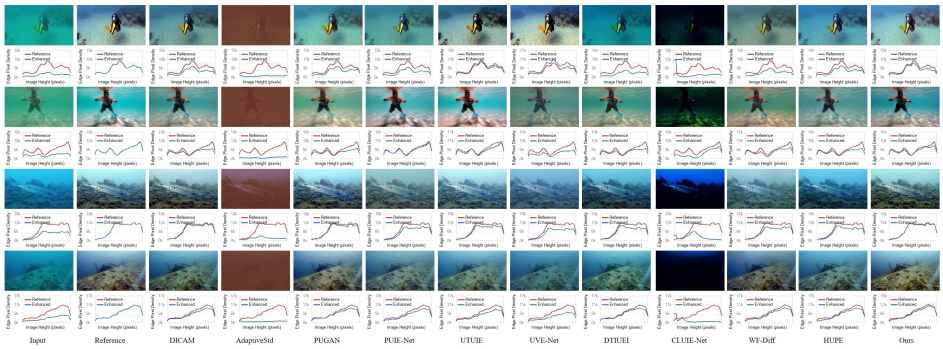


Figure 5. Visual enhancement comparison on UIEB and SUIM-E.

Figure 6 further shows local magnification comparisons on UIEB. Some baselines can improve global brightness, but they either leave visible color bias or blur object boundaries in local regions. For example, AdaptiveStd and CLUIE-Net often fail to recover natural contrast under severe degradation, whereas PUIE-Net and WF-Diff may produce visually smooth regions with weakened fine textures. In contrast, SGDC-UIE restores clearer boundaries and more natural local contrast, which is consistent with its strong PSNR, PCQI, and LPIPS results on UIEB. Figure 7 presents additional examples from LSUI. Although UTUIE obtains the highest PSNR on LSUI, the visual comparison shows that SGDC-UIE preserves local contours and produces more balanced color appearance, matching its leading SSIM and PCQI scores.

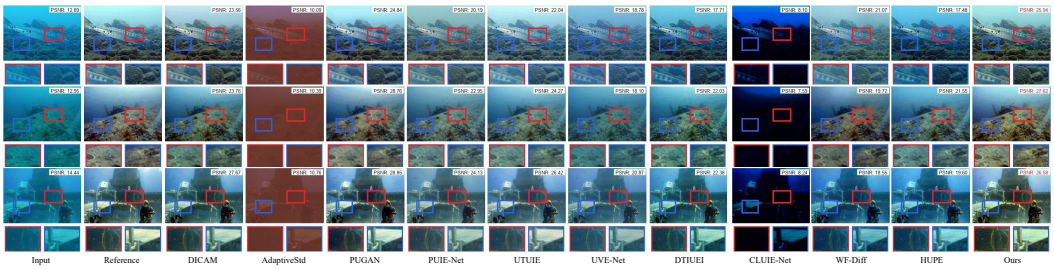


Figure 6. Local visual comparison on UIEB. The red and blue boxes denote the local regions selected for enlarged detail comparison.

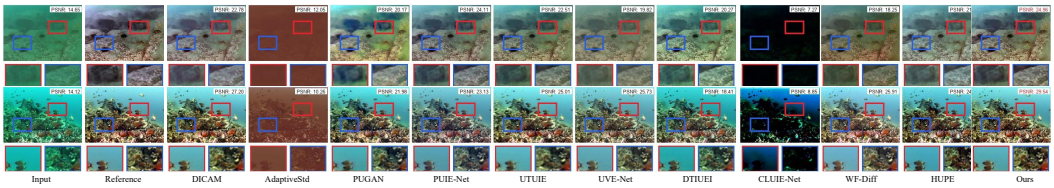


Figure 7. Local visual comparison on LSUI. The red and blue boxes denote the local regions selected for enlarged detail comparison.

For real-world no-reference scenes, Figure 8 shows challenging RUIE-UCCS samples with severe color cast and low visibility. Several competing methods either retain a bluish/greenish tone or introduce local over-enhancement after contrast stretching. In particular, methods with strong global correction may brighten the background water while amplifying noise or unnatural edges. SGDC-UIE produces clearer foreground structures and more natural background water regions, while avoiding obvious artifacts around local edges. This observation is consistent with the best UIQM, AG, and EI values in Table 3. Overall, the visual results indicate that SGDC-UIE improves color, contrast, and structural details in a coordinated manner rather than enhancing one aspect at the expense of another.

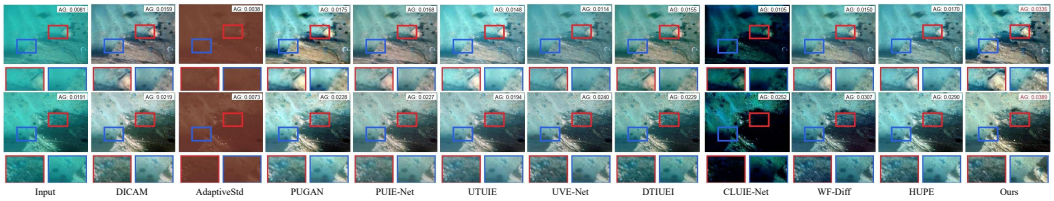


Figure 8. Local visual comparison on RUIE-UCCS. The red and blue boxes denote the local regions selected for enlarged detail comparison.

SGDC-UIE supports cross-dataset generalization through input-adaptive restoration conditioning. At inference, its soft region gates require neither water-type labels nor camera metadata; instead, they modulate input-derived attenuation-like, transmission-like, illumination, structure, and background-light conditions. This design allows the

restoration strength to respond to unseen color casts, scattering, visibility, and illumination changes. The spatial illumination prior and low-frequency path correct local brightness, while bounded residual prediction and adaptive fusion limit over-enhancement. Because the pipeline does not require camera calibration or device-specific optical parameters, its inference procedure is not tied to a particular device model. Results across UIEB, SUIM-E, LSUI, RUIE-UCCS, and RUIE-UIQS suggest generalization across the tested acquisition and degradation distributions, especially on the unpaired RUIE subsets. However, these evaluations are not a controlled cross-device benchmark. Large sensor-response shifts, severe noise, or water conditions far outside the training distribution may reduce semantic-prior reliability. Controlled multi-device evaluation and device-aware adaptation remain future work.

4.2.3. Challenging Cases and Limitations

To avoid presenting only favorable examples, we rank the paired test images using per-image PSNR and SSIM and inspect the lower-performing tail. Figure 9 shows four representative cases selected from UIEB and SUIM-E. Every panel is a real output of the complete SGDC-UIE model, and the raw and reference images are resized only to match the network input geometry for consistent display.

In the first row of Figure 9, the method preserves the shark contours but remains conservative in the dark background, producing substantially lower exposure than the bright reference. In the second row, SGDC-UIE removes part of the dominant green cast and reveals the turtle boundary, yet residual green coloration and weak local contrast remain. The densely textured SUIM-E scenes in the third and fourth rows retain recognizable object structures, but their illumination and high-frequency contrast are under-corrected relative to the references. These examples indicate that bounded residual restoration, while useful for suppressing over-enhancement, can become conservative under extreme exposure shifts, saturated color casts, or dense texture. They also expose the ambiguity of single-reference UIE evaluation, because a large score reduction may partly reflect a substantial exposure difference between two visually plausible renderings. Thus, the aggregate results reported above should be interpreted together with this lower-tail behavior.

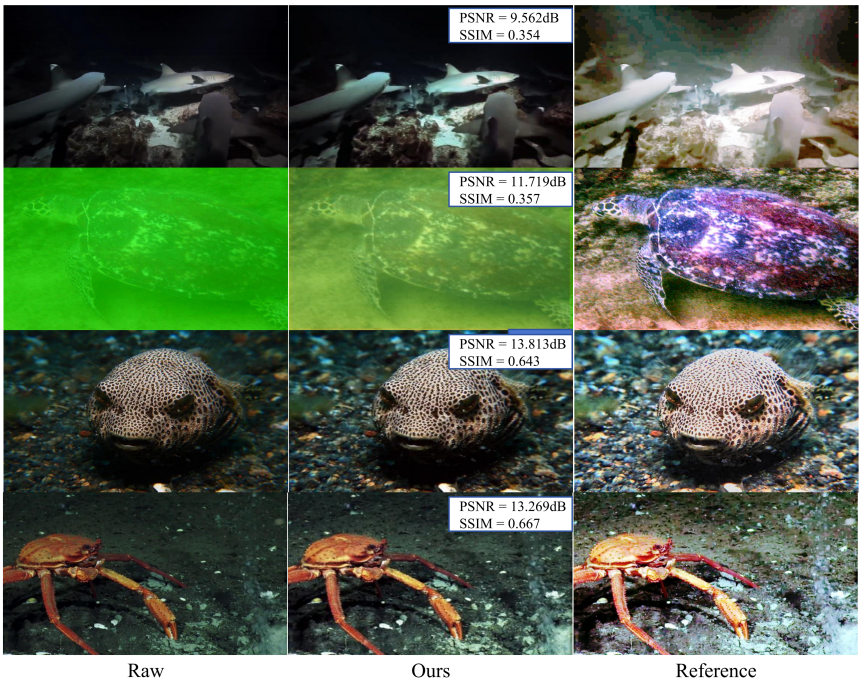


Figure 9. Representative challenging cases from UIEB and SUIM-E.

4.2.4. Computational Complexity and Efficiency Analysis

We also evaluate computational complexity and inference efficiency as summarized in Table 4. The table reports both the lightweight restoration backbone and the complete end-to-end pipeline. Following common restoration-network comparisons, we report parameter count, FLOPs, latency, FPS, and GPU memory under the same input size whenever the released implementation allows it. The restoration module of SGDC-UIE, denoted as Ours-R, contains only 0.394 M parameters, which is the smallest among all compared methods. It also requires 41.919 G FLOPs, ranking second, and achieves 99.719 FPS with 10.028 ms latency. Compared with heavy models such as DTIUIE and WF-Diff, SGDC-UIE is substantially more efficient while still maintaining competitive enhancement quality in Tables 2 and 3.

The efficiency mainly comes from the design of the lightweight dual-branch restoration module. The pseudo-physical priors and semantic gates provide compact restoration conditions, allowing the color and wavelet branches to avoid using a large generic restoration backbone. CLUIE-Net and PUIE-Net achieve faster latency on the tested platform, but their enhancement quality and downstream perception results are less stable across datasets. The complete SGDC-UIE pipeline, denoted as Ours-Full, includes frozen DINOv3 semantic-prior extraction and all subsequent conditioning, degradation estimation, restoration, and fusion operations. It contains 22.572 M parameters and requires 54.767 G FLOPs, 11.539 ms latency, and 237.296 MB GPU memory, corresponding to 86.666 FPS at 384×384 resolution. Relative to Ours-R, the complete pipeline adds 22.178 M parameters, 12.848 G FLOPs, 1.511 ms latency, and 95.420 MB GPU memory. Thus, the semantic-prior stage introduces measurable overhead, but the end-to-end system retains real-time throughput. Reporting Ours-R and Ours-Full separately distinguishes the efficiency of the trainable restoration backbone from the practical deployment cost of the complete framework.

Table 5 further separates the cost of the principal components. The frozen DINOv3 extractor accounts for most parameters, whereas all trainable conditioning and restoration blocks remain lightweight. Among the trainable blocks, the pseudo-physical degradation estimator has the largest isolated FLOP count (17.359 G), while the semantic gates, FiLM conditioning, and adaptive fusion together introduce limited parameter overhead. These costs should be interpreted with the corresponding ablations: removing or corrupting the DINOv3 prior weakens both enhancement and segmentation, suppressing the wavelet branch reduces paired restoration quality, and independently removing the five pseudo-physical priors degrades the overall enhancement–perception balance. The quality-anchor replay mechanism is used only during training and therefore adds no inference-time parameters, FLOPs, latency, or memory. The architecture is thus retained because its components provide complementary functions at measured computational cost, rather than because additional blocks are assumed to be beneficial by default.

Table 5. Component-wise complexity of SGDC-UIE.

Component	Size	Params (M)	FLOPs (G)	Latency (ms)	FPS	Memory (MB)
Frozen DINOv3 semantic prior	384	22.178	12.848	8.595	116.351	11.331
Semantic region gates and FiLM	384	0.062	5.883	0.993	1006.978	108.562
Pseudo-physical degradation estimator	384	0.141	17.359	3.064	326.353	129.134
Color restoration branch	384	0.087	12.905	1.675	597.170	96.188
Wavelet restoration branch	384	0.139	7.552	3.446	290.181	86.306
Adaptive fusion module	384	0.028	4.103	0.680	1470.270	80.438

4.3. Downstream Perception Evaluation

The purpose of underwater image enhancement is not limited to producing visually pleasing images. In practical underwater vision systems, enhanced images are often used by downstream perception models. Therefore, beyond enhancement-quality evaluation, we further assess whether SGDC-UIE benefits dense semantic segmentation and object detection. Following a controlled plug-in evaluation protocol, the downstream models are not re-trained for each enhancement method during testing. For semantic segmentation, enhanced TIUIED images are fed into fixed segmentation models. For object detection, each detector is trained on the raw Brackish training images with the original annotations, and then tested on raw or enhanced Brackish test images. This raw-train/enhanced-test setting evaluates whether an enhancement method improves the visibility of existing target cues rather than benefiting from detector re-training on method-specific outputs. This protocol isolates enhancement as a fixed-model preprocessor; it does not represent jointly retrained or co-optimized enhancement–perception deployment, for which absolute gains and method rankings may differ.

4.3.1. Semantic Segmentation

Table 6 reports segmentation results on TIUIED using three segmentation architectures: DINOv3, U-Net, and DeepLabv3+. The DINOv3 evaluator measures whether the enhanced images remain compatible with foundation-model semantic responses, while U-Net and DeepLabv3+ serve as independent segmentation networks with different architectural biases. This setting is important because a perception-aware enhancer should improve semantic parsing generally rather than overfitting to one particular evaluator.

Table 6. Semantic segmentation comparison on TIUIED. Red, green, and blue indicate the first-, second-, and third-best results, respectively. (↑) indicates that higher values are better.

Method	DINOv3					U-Net					DeepLabv3+				
	PixAcc ↑	mIoU ↑	Dice ↑	FG-Dice ↑	RI-IoU ↑	PixAcc ↑	mIoU ↑	Dice ↑	FG-Dice ↑	RI-IoU ↑	PixAcc ↑	mIoU ↑	Dice ↑	FG-Dice ↑	RI-IoU ↑
Raw	87.778	72.549	82.792	81.041	69.317	70.185	33.579	45.578	39.926	48.622	70.368	37.163	51.576	46.630	46.859
DICAM	87.561	72.590	82.776	81.081	70.077	74.994	38.763	50.362	44.575	52.327	75.252	41.886	55.232	50.137	52.154
PUGAN	86.603	70.791	81.675	80.048	70.227	73.957	38.378	50.108	44.288	50.311	74.741	40.787	53.928	48.653	52.706
WF-Diff	87.312	72.173	82.431	80.719	69.262	73.665	37.488	49.371	43.581	50.693	73.847	40.538	54.379	49.344	53.620
UVE-Net	87.928	72.729	82.818	81.083	69.861	73.746	36.979	48.706	42.884	51.079	74.349	40.902	54.262	49.222	53.795
DTIUIE	87.539	72.349	82.404	80.684	69.677	74.657	38.443	50.162	44.452	52.468	75.184	41.691	55.421	50.456	55.001
HUPE	87.254	71.900	82.229	80.507	69.315	74.937	38.956	50.616	44.953	52.074	75.791	42.951	56.677	51.837	53.927
AdaptiveStd	82.005	62.707	74.576	72.366	59.172	43.392	13.259	20.584	15.015	29.467	45.358	13.937	21.404	15.115	23.347
CLUIE-Net	79.554	62.316	75.425	73.893	53.588	52.947	20.555	30.747	24.823	34.261	52.694	18.030	27.739	21.743	29.509
PUIE-Net	87.847	72.750	82.805	81.045	69.391	74.556	38.882	50.641	44.941	51.088	75.428	42.437	55.586	50.644	53.867
UTUIE	86.477	70.958	82.108	80.477	68.346	74.287	37.869	49.613	43.831	51.168	74.112	41.246	55.302	50.402	52.648
Ours	87.994	73.592	84.043	81.263	70.698	76.417	40.786	52.601	46.943	53.011	76.836	44.573	58.175	53.377	55.941

As shown in Table 6, SGDC-UIE achieves the best results on all reported DINOv3 metrics, including 87.994% PixAcc, 73.592% mIoU, 84.043% Dice, 81.263% FG-Dice, and 70.698% RI-IoU. Compared with DTIUIE and HUPE, which are also perception-oriented methods, SGDC-UIE provides stronger foreground and region-intersection scores, indicating better preservation of task-relevant object structures. Several quality-oriented methods, such as WF-Diff and PUIE-Net, can produce visually smooth images, but their segmentation scores are lower, suggesting that perceptual quality alone does not always preserve semantic boundaries.

The same trend is observed with the independent segmentation models. With U-Net, SGDC-UIE ranks first on all metrics, improving mIoU to 40.786% and Dice to 52.601%. With DeepLabv3+, it again achieves the best performance, reaching 44.573% mIoU and 58.175% Dice. These results indicate that the observed benefit is not limited to the DINOv3 evaluator and that the enhanced images remain useful across different segmentation architectures. The qualitative comparisons in Figure 10 support this observation: SGDC-UIE produces

clearer foreground masks and fewer class-confusion regions, especially around object boundaries and low-contrast underwater structures.

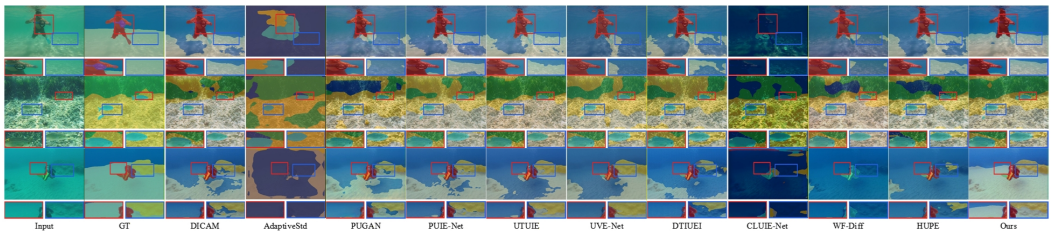


Figure 10. Semantic segmentation results on TIUIED. The red and blue boxes denote the local regions selected for enlarged detail comparison.

4.3.2. Object Detection

Table 7 reports object detection results on Brackish using YOLOv8n, RT-DETR, and Faster R-CNN. These detectors represent one-stage convolutional detection, transformer-based detection, and two-stage region-based detection, respectively. Using multiple detectors helps evaluate whether enhancement improves target visibility in a detector-agnostic manner.

Table 7. Object detection comparison on Brackish. Red, green, and blue indicate the first-, second-, and third-best results, respectively. (↑) indicates that higher values are better.

Method	YOLOv8n					RT-DETR					Faster R-CNN				
	P ↑	R ↑	mAP_{50} ↑	mAP_{75} ↑	$mAP_{50:95}$ ↑	P ↑	R ↑	mAP_{50} ↑	mAP_{75} ↑	$mAP_{50:95}$ ↑	P ↑	R ↑	mAP_{50} ↑	mAP_{75} ↑	$mAP_{50:95}$ ↑
Raw	88.441	68.873	78.862	63.620	55.361	96.984	93.689	96.836	86.236	73.775	70.411	40.508	48.083	25.139	25.957
DICAM	93.907	84.607	92.988	80.496	69.308	97.971	95.333	98.250	89.311	78.504	62.934	76.370	74.991	55.917	48.943
PUGAN	80.485	63.824	71.125	55.578	48.887	92.942	77.383	82.930	71.538	62.814	70.033	74.235	68.575	47.335	43.114
WF-Diff	80.456	63.704	69.579	59.465	49.647	94.070	79.685	87.905	77.310	66.696	60.276	57.877	63.479	45.210	39.714
UVE-Net	94.055	79.440	88.507	74.634	64.621	97.774	92.321	96.581	85.579	74.937	64.257	74.899	67.949	51.571	45.327
DTIUIE	91.888	80.481	86.968	73.026	62.995	97.090	92.660	94.928	88.221	75.565	53.075	58.021	60.069	40.083	36.634
HUPE	97.053	91.100	96.309	83.809	72.427	97.903	97.420	98.497	90.097	78.094	62.419	72.072	70.938	50.404	44.888
AdaptiveStd	0.253	1.236	0.036	0.000	0.005	50.000	1.806	2.606	1.548	1.407	0.149	0.289	0.085	0.011	0.025
CLUIE-Net	53.249	22.647	20.651	10.590	10.857	77.191	41.790	44.583	31.216	28.716	3.057	8.021	2.547	0.891	1.136
PUIE-Net	95.791	88.938	94.703	83.177	70.398	97.424	97.103	98.587	87.658	76.276	29.987	71.091	66.869	48.928	41.935
UTUIE	76.735	52.441	60.977	45.421	39.676	80.155	69.043	74.033	58.602	51.005	17.850	58.684	45.938	26.316	25.649
Ours	97.321	95.329	97.764	88.638	77.146	98.854	97.650	98.873	92.381	82.222	88.118	97.778	95.813	82.091	70.620

SGDC-UIE achieves the best precision, recall, mAP50, mAP75, and mAP50:95 for all three detectors. For YOLOv8n, our method reaches 97.321% precision, 95.329% recall, and 77.146% mAP50:95, outperforming strong baselines such as HUPE and PUIE-Net. For RT-DETR, SGDC-UIE obtains 98.873% mAP50 and 82.222% mAP50:95, showing that transformer-based detectors also benefit from the enhanced images. The improvement is particularly large for Faster R-CNN, where SGDC-UIE achieves 95.813% mAP50 and 70.620% mAP50:95, while the raw input obtains only 48.083% and 25.957%, respectively. Compared with the raw input, the most pronounced change occurs in recall, which rises from 40.508% to 97.778%, whereas precision increases from 70.411% to 88.118%; this pattern is consistent with the recovery of low-contrast targets missed on raw inputs. Because this is a fixed-detector, single-run comparison, the result should be interpreted as protocol-specific evidence rather than proof that the same margin will persist across retraining seeds or jointly optimized protocols.

Compared with methods that primarily optimize visual quality, SGDC-UIE produces images with clearer object boundaries and more discriminative local contrast, which are important for underwater target detection. Some baselines improve one detector but fail to generalize to another, while SGDC-UIE remains consistently strong across one-stage, transformer-based, and two-stage detectors. The detection results in Figure 11 further show fewer missed detections and more reliable bounding boxes under color cast and low-contrast conditions. These downstream results indicate that degradation-consistent

restoration improves not only low-level enhancement metrics but also downstream perception under the controlled fixed-model protocol.

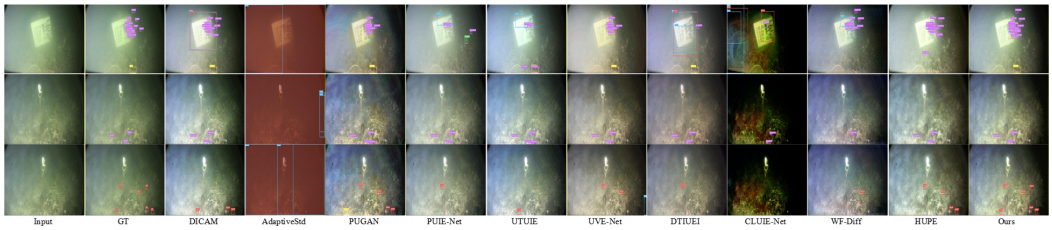


Figure 11. Object detection results on Brackish.

4.4. Ablation Study

We conduct ablation studies to assess whether the proposed components form a coherent coupling among semantic priors, degradation conditions, and restoration behavior. All variants are trained and evaluated under the same protocol, and Table 8 reports paired enhancement quality on SUIM-E and UIEB together with semantic segmentation performance on TIUIED. The analysis focuses on five questions: whether foundation-model semantic responses provide meaningful restoration guidance, how sensitive the framework is to the source and quality of these responses, how each pseudo-physical prior contributes to restoration, whether frequency-aware restoration and semantic gates are necessary, and whether quality-anchor replay and grouped losses stabilize task-aware optimization.

Table 8. Ablation study on SUIM-E, UIEB, and TIUIED. Red, green, and blue indicate the first-, second-, and third-best results, respectively.

Category	Method	SUIM-E				UIEB				TIUIED		
		PSNR	SSIM	UIQM	UCIQE	PSNR	SSIM	UIQM	UCIQE	PixAcc	mIoU	Dice
Complete model	Ours	22.566	0.866	2.562	48.833	23.519	0.863	2.799	47.986	87.994	73.592	84.043
DINOv3 semantic prior	w/o DINOv3 prior	20.741	0.840	2.414	46.929	22.339	0.805	2.565	44.520	87.435	73.588	82.889
	random semantic prior	22.382	0.856	2.409	47.655	22.541	0.815	2.550	45.677	87.910	73.043	82.960
	zero semantic prior	22.562	0.858	2.441	47.551	22.466	0.812	2.552	45.579	87.802	72.625	82.562
Wavelet module	w/o wavelet branch	21.815	0.851	2.420	47.145	21.287	0.847	2.420	47.581	87.437	73.604	83.924
	w/o wavelet loss	22.496	0.859	2.417	48.584	21.862	0.853	2.384	47.551	87.729	72.909	82.731
	w/o wavelet LF	22.364	0.860	2.386	48.761	22.317	0.856	2.357	47.901	87.609	72.726	82.584
	w/o wavelet HF	22.543	0.860	2.430	48.503	21.964	0.853	2.381	47.694	87.697	72.890	82.721
	w/o semantic gates	22.547	0.860	2.437	48.526	21.947	0.852	2.398	47.599	87.728	72.936	82.799
Quality anchor	w/o quality anchor	22.044	0.847	2.472	47.404	20.401	0.844	2.419	46.751	87.990	72.434	83.736
Loss terms	w/o physics loss	22.654	0.859	2.418	48.366	21.906	0.853	2.395	47.634	87.678	72.895	82.735
	w/o structure loss	22.498	0.859	2.430	48.723	21.846	0.851	2.388	47.612	87.703	72.922	82.755
	w/o color loss	22.128	0.858	2.418	48.748	21.999	0.858	2.340	47.947	87.671	72.786	82.623

4.4.1. Break-Down Ablations

We first analyze the complete SGDC-UIE against variants that remove or corrupt its major components. The complete model achieves the best or near-best results across most enhancement and perception metrics, including 22.566 dB PSNR and 0.866 SSIM on SUIM-E, 23.519 dB PSNR and 0.863 SSIM on UIEB, and 84.043% Dice on TIUIED. This balanced behavior is important because underwater enhancement should not only maximize a single fidelity score, but also preserve visual quality and task-relevant structures.

Removing the DINOv3 semantic prior causes a clear decrease in paired enhancement quality. On SUIM-E, PSNR drops from 22.566 dB to 20.741 dB and SSIM decreases from 0.866 to 0.840. On UIEB, PSNR decreases from 23.519 dB to 22.339 dB and SSIM decreases from 0.863 to 0.805. Replacing the prior with random or zero responses also weakens the overall balance, especially on UIEB and TIUIED. Although these corrupted-prior variants may retain some scores on individual metrics, they no longer provide stable semantic region

cues for degradation modulation. This supports the interpretation that the DINOv3 prior is not merely an additional input tensor; it provides a meaningful region-level coordinate system for distinguishing foreground objects, boundaries, and background water regions.

The wavelet branch is another key component. Without the whole wavelet branch, UIEB PSNR decreases from 23.519 dB to 21.287 dB, and SUIM-E PSNR decreases from 22.566 dB to 21.815 dB. This indicates that direct spatial restoration alone is insufficient to recover both color appearance and structural detail. Removing the wavelet loss, low-frequency path, high-frequency path, or semantic gates also reduces the cross-metric stability. These results support the design motivation of the branch: low-frequency restoration mainly stabilizes color and illumination, while high-frequency restoration preserves edges and textures. Semantic gates further prevent the high-frequency path from indiscriminately enhancing background noise as useful structure.

The frequency-specific variants further show why the wavelet design should be semantic-aware rather than a generic detail enhancer. Removing the LF path weakens global appearance consistency, and removing the HF path affects structural restoration; their UIEB PSNR values decrease to 22.317 dB and 21.964 dB, respectively. Disabling semantic gates in the wavelet module also reduces the overall balance among SUIM-E, UIEB, and TIUIED metrics, with UIEB PSNR decreasing to 21.947 dB. These observations indicate that the proposed wavelet branch is not used only to sharpen images. Instead, it separates low-frequency color/illumination correction from high-frequency structure recovery, and then uses semantic gates to decide where each type of restoration should be emphasized. This design is consistent with the underwater degradation process, where background water regions and object boundaries often require different enhancement behaviors.

Table 9 evaluates one-at-a-time removal of the five pseudo-physical conditions while keeping the remaining network, training settings, and evaluation protocol unchanged. Removing any one of them reduces the joint enhancement–perception performance. The transmission and illumination priors are particularly relevant to scattering and spatial brightness correction: without them, SUIM-E PSNR decreases by 0.693 and 0.721 dB, respectively, and both variants obtain lower SSIM, UIQM, and UCIQE on the two paired datasets. Suppressing the structure prior produces the largest SUIM-E UIQM decrease among these three variants, from 2.562 to 2.369, which is consistent with its role in preserving task-relevant boundaries and local details.

Table 9. Pseudo-physical-prior ablation on SUIM-E, UIEB, and TIUIED. Red, green, and blue indicate the first-, second-, and third-best results, respectively.

Variant	SUIM-E				UIEB				TIUIED		
	PSNR	SSIM	UIQM	UCIQE	PSNR	SSIM	UIQM	UCIQE	PixAcc	mIoU	Dice
Full model	22.566	0.866	2.562	48.833	23.519	0.863	2.799	47.986	87.994	73.592	84.043
w/o attenuation prior	22.668	0.825	2.483	46.252	22.280	0.830	2.611	43.671	87.069	73.408	83.425
w/o transmission prior	21.873	0.837	2.476	46.285	23.395	0.852	2.607	43.742	87.013	73.357	83.414
w/o illumination prior	21.845	0.827	2.466	46.463	22.361	0.852	2.603	43.821	87.060	73.473	83.542
w/o structure prior	21.858	0.827	2.369	46.482	23.383	0.852	2.603	43.808	87.022	73.394	83.461
w/o background-light prior	20.917	0.859	2.491	44.840	19.057	0.808	2.516	40.073	86.815	73.051	83.154

The attenuation-prior variant yields a small 0.102 dB increase in SUIM-E PSNR, but this isolated gain is accompanied by lower SUIM-E SSIM, UIQM, and UCIQE, a 1.239 dB PSNR decrease on UIEB, and reductions in all three TIUIED metrics. The background-light prior has the largest overall effect: removing it decreases UIEB PSNR from 23.519 to 19.057 dB and UCIQE from 47.986 to 40.073, while Dice decreases from 84.043% to 83.154%. Thus, no single prior is sufficient by itself; attenuation, transmission, illumination, structure,

and background light provide complementary conditions for balancing color recovery, visibility, structural preservation, and downstream perception.

4.4.2. Semantic Prior Sensitivity

We first examine whether the semantic guidance interface is sensitive to prior quality. The random-prior and zero-prior variants show that simply adding a gating interface is not enough; the gates must carry meaningful semantic responses. For example, random and zero priors achieve lower UIEB SSIM than the full model, and their TIUIED Dice scores drop from 84.043% to 82.960% and 82.562%, respectively. This suggests that degraded underwater images require semantic guidance that is stable under color cast and turbidity. The frozen DINOv3 prior provides meaningful cues under these degradations and helps the restoration module assign different correction strengths to objects, boundaries, and water background.

To quantify sensitivity to semantic-prior quality, we replace DINOv3 with DINOv2, ConvNeXt, and SegFormer while retaining the remaining enhancement framework. As shown in Table 10, DINOv2 is the closest alternative: its TIUIED mIoU and Dice are 73.058% and 83.395%, only 0.534 and 0.648 percentage points below the DINOv3 configuration, although paired enhancement quality also decreases. ConvNeXt nearly matches the DINOv3 UIQM values on SUIM-E and UIEB, but its mIoU and Dice decrease by 13.763 and 10.463 percentage points. SegFormer exhibits a larger task-oriented decline, reaching 55.027% mIoU and 68.963% Dice. These controlled replacements indicate that matching low-level appearance statistics does not guarantee the preservation of region-level semantics. Because each alternative-prior configuration is evaluated from one controlled training run, these results are treated as a descriptive sensitivity analysis; we do not claim that DINOv3 is statistically superior to every alternative. Rather, it provides the strongest overall enhancement–perception balance among the tested priors under the common interface.

Table 10. Semantic-prior comparison on SUIM-E, UIEB, and TIUIED. Red, green, and blue indicate the first-, second-, and third-best results, respectively.

Semantic Prior	SUIM-E				UIEB				TIUIED		
	PSNR	SSIM	UIQM	UCIQE	PSNR	SSIM	UIQM	UCIQE	PixAcc	mIoU	Dice
DINOv3	22.566	0.866	2.562	48.833	23.519	0.863	2.799	47.986	87.994	73.592	84.043
DINOv2	22.394	0.824	2.559	46.352	22.185	0.860	2.790	43.110	87.211	73.058	83.395
ConvNeXt	22.543	0.824	2.561	46.311	22.170	0.849	2.797	43.006	81.742	59.829	73.580
SegFormer	22.227	0.823	2.552	46.332	22.027	0.859	2.700	42.971	79.821	55.027	68.963

Table 11 reports the fixed manuscript results together with the run-to-run dispersion of the complete DINOv3 configuration over seeds 2023, 2024, and 2025. The standard deviations are 0.100 dB for SUIM-E PSNR, 0.075 dB for UIEB PSNR, 0.064 percentage points for TIUIED mIoU, and 0.068 percentage points for Dice; both SSIM standard deviations are below 0.001. These values indicate limited variation under the controlled retraining protocol, but they do not establish statistical significance for the 0.031 dB UIEB PSNR margin over PUGAN. Together with the missing, random, and zero-prior variants in Table 8, the results show that meaningful dense semantic conditions contribute reproducibly under the common protocol.

Table 11. Controlled-repeat stability on SUIM-E, UIEB, and TIUIED.

Dataset	Metric	Main Result	Repeat-Run SD	Runs
SUIM-E	PSNR	22.566	0.100	3
SUIM-E	SSIM	0.866	0.00071	3
UIEB	PSNR	23.519	0.075	3
UIEB	SSIM	0.863	0.00034	3
TIUIED	mIoU	73.592	0.064	3
TIUIED	Dice	84.043	0.068	3

4.4.3. Physical Interpretation of the Pseudo-Physical Priors

We further examine whether the estimated conditions preserve degradation-related spatial organization rather than acting as unconstrained latent channels. Two complementary protocols are used. First, clean underwater images are degraded by a controlled forward process with known wavelength-dependent attenuation, spatially varying transmission and illumination, structural blur, and background light; the resulting simulator fields provide ground-truth degradation maps. Second, because calibrated pixel-wise physical parameters are unavailable for real images, we construct non-learned physics-inspired references from relative channel attenuation, dark-channel-style transmission, Gaussian-smoothed luminance, inverse normalized gradient magnitude, and the brightest dark-channel candidate for background light. These references are evaluated on UIEB and the blue, blue–green, and green subsets of RUIE-UCCS. Spearman correlation is used because the predicted conditions are soft restoration controls rather than estimates in calibrated physical units. For the four spatial conditions, we report the mean absolute within-image correlations; for background light, we report the signed correlation pooled over images and RGB channels.

Table 12 shows that the learned conditions retain measurable degradation-related organization under both protocols. On the controlled data, the transmission condition has a mean absolute spatial correlation of 0.398 with the true field, while the estimated background-light vector reaches a signed correlation of 0.868 with the simulator parameter. On real images, the strongest associations depend on water type: the attenuation-condition correlation reaches 0.745 on blue–green water, the transmission-condition correlation reaches 0.563 on green water, and the illumination- and structure-condition correlations reach 0.463 and 0.428, respectively. This variation is expected because a hand-crafted estimator provides only an approximate reference for real scenes, and different water types emphasize different degradation mechanisms. In Figure 12, each scalar map is independently normalized and color-mapped for visualization, while each background-light patch is rendered from the corresponding RGB vector. All displayed maps are generated from model outputs or reference fields rather than manually drawn.

Table 12. Pseudo-physical-condition correlations on controlled synthetic data, RUIE-UCCS, and UIEB.

Dataset	N	ρ_β	ρ_t	ρ_L	ρ_S	ρ_{B_∞}
Controlled synthetic	120	0.261	0.398	0.376	0.270	0.868
RUIE-UCCS blue	100	0.613	0.490	0.191	0.320	0.542
RUIE-UCCS blue–green	100	0.745	0.271	0.307	0.428	0.592
RUIE-UCCS green	100	0.625	0.563	0.463	0.259	0.440
UIEB	160	0.530	0.453	0.461	0.424	0.394

Figure 12 complements the correlation analysis. The predicted fields follow broad foreground–background transitions and object/terrain boundaries in the simulator fields or physics-inspired references, while remaining smoother and more region-selective because

they are optimized jointly with restoration. These variables are soft conditions rather than calibrated optical measurements, and subsequent affine modulation can absorb their signs and scales. Thus, the absolute correlations measure degradation-related spatial organization rather than parameter identity; the mean signed correlations on the controlled set are −0.182, 0.392, −0.032, and 0.177 for attenuation, transmission, illumination, and structure, respectively. Combined with the individual-prior ablation in Table 9, these results show that the learned conditions provide complementary degradation-oriented controls for restoration.

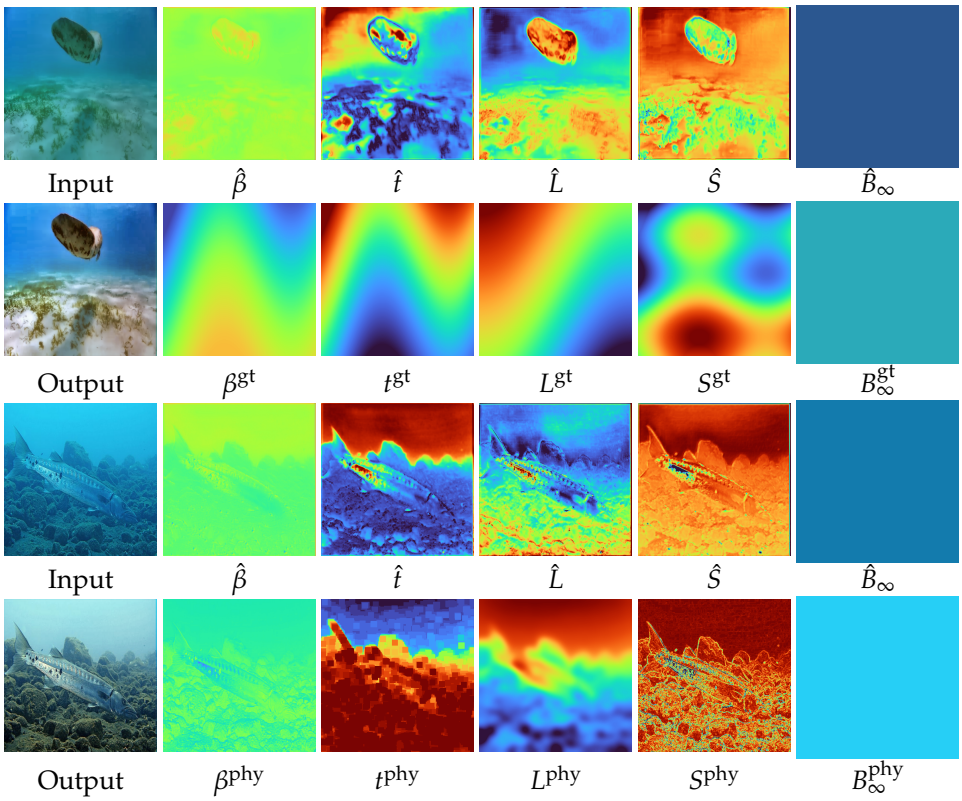


Figure 12. Pseudo-physical-condition visualization on controlled synthetic data and UIEB.

4.4.4. Quality-Anchor Replay and Loss Regularization

Finally, we study the training constraints. Quality-anchor replay is introduced to prevent visual-quality drift during task-aware fine-tuning. Without quality anchors, the model still obtains competitive task-oriented scores, but paired restoration quality becomes much less stable. On SUIM-E, removing quality anchors reduces PSNR from 22.566 dB to 22.044 dB and SSIM from 0.866 to 0.847. The degradation is more evident on UIEB, where PSNR decreases from 23.519 dB to 20.401 dB. This supports the interpretation that task-aware optimization alone may bias the network toward semantic responses while sacrificing reference-based visual fidelity. Quality-anchor replay works as a stabilizing mechanism by repeatedly constraining the model with reliable paired enhancement samples.

We also remove the physics, structure, and color losses to evaluate the grouped objective. These losses regularize different parts of the proposed degradation-consistent process: the physics loss constrains pseudo-physical prior estimation, the structure loss encourages boundary consistency, and the color loss stabilizes chromatic correction. Some variants may improve a single metric, such as the slightly higher SUIM-E PSNR obtained without the physics loss, but they do not consistently improve UIEB quality or TIUIED segmentation. The full model provides the most balanced overall trade-off across paired fidelity, underwater quality, and downstream perception. The ablations support the coupled

design of semantic region partition, degradation prior construction, frequency-aware restoration, and quality-preserving task-aware optimization.

5. Conclusions

In this paper, we proposed SGDC-UIE, a Semantic Guidance Network with Degradation Consistency for Underwater Image Enhancement. The framework uses a frozen DINOv3 semantic prior to construct stable foreground, boundary, and background region gates, and then converts these gates into pseudo-physical degradation conditions. Based on these conditions, a dual-branch restoration module corrects low-frequency color and illumination degradation while recovering high-frequency structural details. The two restoration candidates are finally integrated by a degradation-consistent fusion gate, enabling region-dependent underwater enhancement guided by both semantic regions and local degradation states. Experiments on paired benchmarks, real-world no-reference images, semantic segmentation, and underwater object detection show that SGDC-UIE achieves a favorable balance among visual fidelity, underwater image quality, and downstream perception utility. The ablation studies further support the contributions of the semantic prior, semantic-aware wavelet restoration, quality-anchor replay, and grouped degradation-consistent supervision.

Despite its effectiveness, the current framework still relies on a frozen semantic prior and task-aware annotations to build stable semantic guidance. Moreover, the current evaluation is not a controlled cross-device benchmark; large sensor-response shifts and severe sensor noise may reduce semantic-prior reliability, motivating future multi-device evaluation and device-aware adaptation. The failure-case analysis further shows that bounded correction can be conservative under extreme exposure shifts, saturated color casts, and dense high-frequency textures. The semantic-to-degradation conditioning principle may extend to related restoration tasks, but each task requires its own degradation variables, prediction heads, and supervision. Evaluating this transfer in dehazing, adverse-weather restoration, and low-light enhancement remains future work. In future work, we will extend SGDC-UIE to video-based underwater enhancement with temporal consistency, and explore weaker-supervision strategies and stronger imaging constraints for extreme turbidity, low-light, and long-range underwater scenes.

Author Contributions: Conceptualization, R.M. and J.Z.; methodology, R.M.; software, R.M. and T.L.; validation, R.M., J.Z., H.L. and J.Y.; formal analysis, R.M. and J.Y.; investigation, R.M. and J.Y.; resources, J.Z. and H.L.; data curation, R.M. and T.L.; writing—original draft preparation, R.M.; writing—review and editing, R.M., J.Z., H.L. and J.Y.; visualization, R.M. and J.Y.; supervision, J.Z. and T.L.; project administration, H.L.; funding acquisition, R.M. and J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Natural Science Foundation of Fujian Province (No. 2026J0011478), in part by the Fuzhou City Science and Technology Talent Cultivation Program Project (No. 2025-R-006), in part by the the Special Fund Project for Promoting High-Quality Development of Marine and Fishery Industries in Fujian Province (No. FJHYF-L-2025-07-003), and in part by the Science and Technology Project of the Fuzhou Institute of Oceanography (No. 2025F17).

Data Availability Statement: The public datasets used in this study are available from their respective dataset providers. The source code and trained models will be made available upon acceptance of the paper.

Acknowledgments: The authors would like to thank the providers of the public underwater image datasets used in this study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zhang, W.; Wang, M.; Zhuang, P.; Liu, D. Underwater Image Enhancement via Advantage Feature Weighted Fusion. *IEEE Trans. Circuits Syst. Video Technol.* **2026**, *36*, 2006–2018. [CrossRef]
2. Lillo Sánchez, F.; Palacios Játiva, P.; Sánchez, I.; Saavedra Arancibia, C.; Ali Amjad, A.; Azurdia Meza, C.; Zabala-Blanco, D.; Soto, I. Simulation-Based Performance Evaluation of Underwater Visible Light Communication (UVLC) with RGB LEDs: Deterministic Versus Stochastic Modeling. *IEEE Access* **2026**, *14*, 81555–81572. [CrossRef]
3. Almonacid, L.; Palacios Játiva, P.; Azurdia-Meza, C.A.; Dujovne, D.; Soto, I.; Firoozabadi, A.D.; Gutiérrez Gaitán, M. On the Path Loss Performance of Underwater Visible Light Communication Schemes Evaluated in Several Water Environments. In Proceedings of the 2023 South American Conference on Visible Light Communications (SACVLC), IEEE, Santiago, Chile, 8–10 November 2023; pp. 12–16.
4. Sun, Y.; Hong, Y.; Zhou, X.; Dong, J. Water-AutoSAM: Dual-Domain Enhanced Auto-Prompting SAM for Underwater Segmentation. *J. Mar. Sci. Eng.* **2026**, *14*, 953. [CrossRef]
5. Tian, W.; Wu, G. FEAPN: Feature Enhancement and Alignment Pyramid Network for Underwater Object Detection. *J. Mar. Sci. Eng.* **2026**, *14*, 671. [CrossRef]
6. Du, D.; Si, L.; Xu, F.; Niu, J.; Sun, F. A Physical Model-Guided Framework for Underwater Image Enhancement and Depth Estimation. *IEEE Trans. Circuits Syst. Video Technol.* **2026**, *36*, 10244–10259. [CrossRef]
7. Ma, Y.; Yang, Q.; Zhu, R.; Li, Y.; Wahab, M.M.A. Underwater Image Enhancement Via Dual-Path Frequency-Domain Fusion Network. *IEEE Trans. Circuits Syst. Video Technol.* 2026, *Early Access*. [CrossRef]
8. Li, C.; Guo, C.; Ren, W.; Cong, R.; Hou, J.; Kwong, S.; Tao, D. An Underwater Image Enhancement Benchmark Dataset and Beyond. *IEEE Trans. Image Process.* **2020**, *29*, 4376–4389. [CrossRef]
9. Peng, L.; Zhu, C.; Bian, L. U-Shape Transformer for Underwater Image Enhancement. *IEEE Trans. Image Process.* **2023**, *32*, 3066–3079. [CrossRef] [PubMed]
10. Cong, R.; Yang, W.; Zhang, W.; Li, C.; Guo, C.L.; Huang, Q.; Kwong, S. PUGAN: Physical Model-Guided Underwater Image Enhancement Using GAN with Dual-Discriminators. *IEEE Trans. Image Process.* **2023**, *32*, 4472–4485. [CrossRef] [PubMed]
11. Li, K.; Wu, L.; Qi, Q.; Liu, W.; Gao, X.; Zhou, L.; Song, D. Beyond Single Reference for Training: Underwater Image Enhancement via Comparative Learning. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *33*, 2561–2576. [CrossRef]
12. Qing, Y.; Liu, S.; Wang, H.; Wang, Y. DiffUIE: Learning Latent Global Priors in Diffusion Models for Underwater Image Enhancement. *IEEE Trans. Multimed.* **2025**, *27*, 2516–2529. [CrossRef]
13. Zhang, W.; Zhou, L.; Zhuang, P.; Li, G.; Pan, X.; Zhao, W.; Li, C. Underwater Image Enhancement via Weighted Wavelet Visual Perception Fusion. *IEEE Trans. Circuits Syst. Video Technol.* **2024**, *34*, 2469–2483. [CrossRef]
14. Jiang, W.; Tan, Y.; Qiu, Z.; Wang, Z.; Yu, Y.; Jiang, Q. PyUIE: A Coarse-to-Fine Deep Pyramid Network for Underwater Image Enhancement. *IEEE Trans. Multimed.* **2026**, *28*, 3054–3067. [CrossRef]
15. Qi, Q.; Li, K.; Zheng, H.; Gao, X.; Hou, G.; Sun, K. SGUIE-Net: Semantic Attention Guided Underwater Image Enhancement with Multi-Scale Perception. *IEEE Trans. Image Process.* **2022**, *31*, 6816–6830. [CrossRef]
16. Zhang, Z.; Jiang, Z.; Ma, L.; Liu, J.; Fan, X.; Liu, R. HUPE: Heuristic Underwater Perceptual Enhancement with Semantic Collaborative Learning. *Int. J. Comput. Vis.* **2025**, *133*, 3259–3277. [CrossRef]
17. Lin, B.; Gao, F.; Yu, Y.; Dong, J.; Du, Q. Downstream Task-Inspired Underwater Image Enhancement: A Perception-Aware Study From Dataset Construction to Network Design. *IEEE Trans. Image Process.* **2026**, *35*, 2873–2888. [CrossRef] [PubMed]
18. Siméoni, O.; Vo, H.V.; Seitzer, M.; Baldassarre, F.; Oquab, M.; Jose, C.; Khalidov, V.; Szafraniec, M.; Yi, S.; Ramamonjisoa, M.; et al. DINOv3. *arXiv* **2025**, arXiv:2508.10104. [CrossRef]
19. Rehman, M.U.; Bakht, A.B.; Hussain, I. DeepAquaEnhance: Hierarchical Multi-Head Attention and Adaptive Feature Fusion for Underwater Image Enhancement. *Ecol. Inform.* **2026**, *95*, 103771. [CrossRef]
20. Guan, M.; Xu, H.; Jiang, G.; Yu, M.; Chen, Y.; Luo, T.; Song, Y. WaterMamba: Visual State Space Model for Underwater Image Enhancement. *arXiv* **2024**, arXiv:2405.08419.
21. Berman, D.; Levy, D.; Avidan, S.; Treibitz, T. Underwater Single Image Color Restoration Using Haze-Lines and a New Quantitative Dataset. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 2822–2837. [PubMed]
22. Fu, Z.; Lin, H.; Yang, Y.; Chai, S.; Sun, L.; Huang, Y.; Ding, X. Unsupervised Underwater Image Restoration: From a Homology Perspective. In Proceedings of the AAAI Conference on Artificial Intelligence, Online, 22 February–1 March 2022; Volume 36, pp. 643–651.
23. Fu, Z.; Wang, W.; Huang, Y.; Ding, X.; Ma, K.K. Uncertainty Inspired Underwater Image Enhancement. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; pp. 465–482.
24. Zhang, D.; He, Z.; Zhang, X.; Wang, Z.; Ge, W.; Shi, T.; Lin, Y. Underwater Image Enhancement via Multi-Scale Fusion and Adaptive Color-Gamma Correction in Low-Light Conditions. *Eng. Appl. Artif. Intell.* **2023**, *126*, 106972. [CrossRef]
25. Jiang, Q.; Zhang, Y.; Bao, F.; Zhao, X.; Zhang, C.; Liu, P. Two-Step Domain Adaptation for Underwater Image Enhancement. *Pattern Recognit.* **2022**, *122*, 108324. [CrossRef]

26. Huang, S.; Wang, K.; Liu, H.; Chen, J.; Li, Y. Contrastive Semi-Supervised Learning for Underwater Image Restoration via Reliable Bank. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 18145–18155.
27. Liu, Y.; Jiang, Q.; Wang, X.; Luo, T.; Zhou, J. Underwater Image Enhancement with Cascaded Contrastive Learning. *IEEE Trans. Multimed.* **2025**, *27*, 1512–1525. [CrossRef]
28. Li, M.; Wang, K.; Shen, L.; Lin, Y.; Wang, Z.; Zhao, Q. UIALN: Enhancement for Underwater Image with Artificial Light. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *33*, 3622–3637. [CrossRef]
29. Zheng, Y.; Lu, H.; Wang, J.; Zhang, W.; Guizani, M. High-Turbidity Underwater Image Enhancement via Turbidity Suppression Fusion. *IEEE Trans. Circuits Syst. Video Technol.* **2025**, *35*, 3527–3540. [CrossRef]
30. Li, Y.; Mi, Z.; Wang, Y.; Jiang, S.; Fu, X. TAFormer: A Transmission-Aware Transformer for Underwater Image Enhancement. *IEEE Trans. Circuits Syst. Video Technol.* **2025**, *35*, 601–616. [CrossRef]
31. Li, K.; Fan, H.; Qi, Q.; Yan, C.; Sun, K.; Wu, Q.M.J. TCTL-Net: Template-Free Color Transfer Learning for Self-Attention Driven Underwater Image Enhancement. *IEEE Trans. Circuits Syst. Video Technol.* **2024**, *34*, 4682–4697. [CrossRef]
32. Zhang, Y.; Yuan, J.; Cai, Z. DCGF: Diffusion-Color-Guided Framework for Underwater Image Enhancement. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 1–12. [CrossRef]
33. Zhao, C.; Dong, C.; Cai, W.; Wang, Y. Learning a Physical-Aware Diffusion Model Based on Transformer for Underwater Image Enhancement. *IEEE Trans. Geosci. Remote Sens.* **2026**, *64*, 1–14. [CrossRef]
34. Zhao, C.; Cai, W.; Dong, C.; Zeng, C. Wavelet-Based Fourier Information Interaction with Frequency Diffusion Adjustment for Underwater Image Restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 16–22 June 2024; pp. 8281–8291.
35. Zhou, L.; Tan, H.; Liu, T.; Zhu, Y.; Liang, J.; Tao, Y. WDFN: Wavelet-Based Decomposition-Fusion Network for Low-Light Underwater Image Enhancement. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 1–17. [CrossRef]
36. Ye, B.; Wang, H.; Wang, W.; Wang, T. FreqMambaGAN: A Frequency-Decoupled Mamba-Enhanced CycleGAN for Underwater Image Enhancement. *J. Mar. Sci. Eng.* **2026**, *14*, 1050. [CrossRef]
37. Chen, Y.; Zhang, J.; Li, Y.; Li, Y.; Tang, H.; Wang, H.; Fu, X. Fusion-Based Channel-Wise Isotropic Convergent Real-Time Underwater Image Enhancement. *IEEE Trans. Circuits Syst. Video Technol.* **2025**, *35*, 9763–9774. [CrossRef]
38. Cheng, B.; Misra, I.; Schwing, A.G.; Kirillov, A.; Girdhar, R. Masked-Attention Mask Transformer for Universal Image Segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 1290–1299.
39. Zhou, J.; Jiang, Q.; Zhang, R.; Liu, Z.; Liu, W.; Gao, X. AMSP-UOD: When Vortex Convolution and Stochastic Perturbation Meet Underwater Object Detection. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 26–27 February 2024; Volume 38, pp. 7659–7667.
40. Zhou, W.; Zeng, J.; Li, S.; Zhang, Y. E2E-AUD: An End-to-End Adaptive Underwater Detection Framework Integrating Physical Priors and Frequency-Adaptive Learning. *J. Mar. Sci. Eng.* **2026**, *14*, 1067. [CrossRef]
41. Xiao, B.; Zhang, X.; Li, J.; Yang, X. SPDFusion: A Semantic Prior Knowledge-Driven Method for Infrared and Visible Image Fusion. *IEEE Trans. Multimed.* **2025**, *27*, 1689–1702. [CrossRef]
42. Radford, A.; Kim, J.W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. Learning Transferable Visual Models From Natural Language Supervision. In Proceedings of the International Conference on Machine Learning, Virtual, 18–24 July 2021; pp. 8748–8763.
43. Caron, M.; Touvron, H.; Misra, I.; Jégou, H.; Mairal, J.; Bojanowski, P.; Joulin, A. Emerging Properties in Self-Supervised Vision Transformers. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 9650–9660.
44. Oquab, M.; Darcet, T.; Moutakanni, T.; Vo, H.; Szafraniec, M.; Khalidov, V.; Fernandez, P.; Haziza, D.; Massa, F.; El-Nouby, A.; et al. DINOv2: Learning Robust Visual Features without Supervision. *arXiv* **2024**, arXiv:2304.07193.
45. Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.C.; Lo, W.Y.; et al. Segment Anything. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2–6 October 2023; pp. 4015–4026.
46. Liu, R.; Fan, X.; Zhu, M.; Hou, M.; Luo, Z. Real-World Underwater Enhancement: Challenges, Benchmarks, and Solutions Under Natural Light. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *30*, 4861–4875. [CrossRef]
47. Pedersen, M.; Bruslund Haurum, J.; Gade, R.; Moeslund, T.B. Detection of Marine Animals in a New Underwater Dataset with Varying Visibility. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Long Beach, CA, USA, 15–20 June 2019; pp. 18–26.
48. Rehman, M.U.; Abbas, Z.; Nasir, M.F.; Hussain, I. A Multiscale Physics-Informed Framework for Robust No-Reference Underwater Image Quality Evaluation. *Alex. Eng. J.* **2026**, *135*, 114–125. [CrossRef]

49. Jiang, Q.; Gu, Y.; Wu, Z.; Li, C.; Xiong, H.; Shao, F.; Wang, Z. Deep Underwater Image Quality Assessment with Explicit Degradation Awareness Embedding. *IEEE Trans. Image Process.* **2025**, *34*, 1297–1310. [CrossRef] [PubMed]
50. Lin, B.; Dong, J.; Dong, X.H. Perception-Aware Underwater Image Quality Assessment: Dataset, Perceptual Quality Scores, and Assessment Network. *IEEE Trans. Circuits Syst. Video Technol.* **2025**, *35*, 11113–11128. [CrossRef]
51. Tolie, H.F.; Ren, J.; Elyan, E. DICAM: Deep Inception and Channel-Wise Attention Modules for Underwater Image Enhancement. *Neurocomputing* **2024**, *584*, 127585. [CrossRef]
52. Park, C.W.; Eom, I.K. Underwater Image Enhancement Using Adaptive Standardization and Normalization Networks. *Eng. Appl. Artif. Intell.* **2024**, *127*, 107445. [CrossRef]
53. Xie, Y.; Kong, L.; Chen, K.; Zheng, Z.; Yu, X.; Yu, Z.; Zheng, B. UVEB: A Large-Scale Benchmark and Baseline Towards Real-World Underwater Video Enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 22358–22367.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.