PHYSICAL DEGRADATION-AWARE GENERATIVE PRIOR OPTIMIZATION FOR ZERO-SHOT UNDERWATER IMAGE RESTORATION

Anonymous authors
Paper under double-blind review

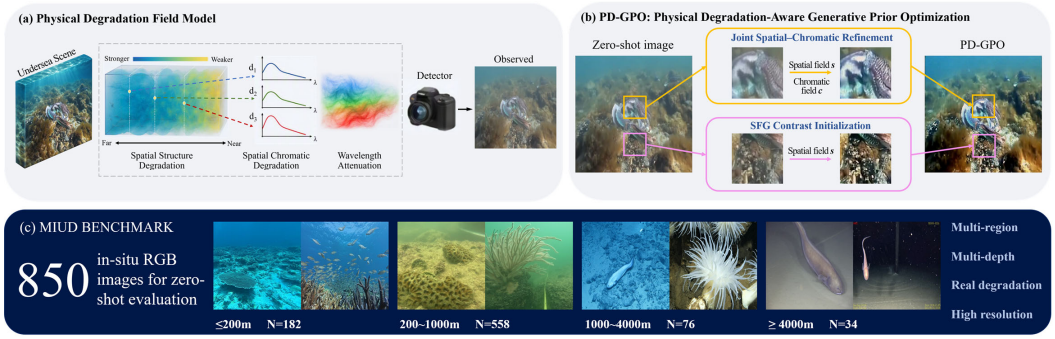

Figure 1: (a) Spatially varying structural degradation and wavelength-dependent chromatic attenuation in underwater imaging. (b) PD-GPO represents these effects with a scalar spatial field s and a chromatic field $\mathbf{c}$. (c) Examples from MIUD, our independently collected multi-region in-situ dataset, grouped by acquisition depth for zero-shot evaluation.

ABSTRACT

Single-image underwater restoration is ill-posed: wavelength-dependent attenuation and scattering produce spatially varying structural and chromatic degradation. Learned underwater restoration methods typically depend on training across underwater image collections, whereas generic generative priors provide no explicit guidance for such nonuniform degradation. We propose Physical Degradation-Aware Generative Prior Optimization (PD-GPO), a zero-shot framework that uses a spatial and a chromatic degradation field to guide image-specific latent optimization under a frozen generative prior. A physical degradation field estimator (PDFE), trained only on synthetic degradations of natural RGB-D images, remains frozen and estimates both fields from each generated candidate. The spatial field measures the local degree of structural degradation; it guides contrast initialization and selects the support on which the chromatic field regularizes channel ratios. A frozen DINOv2 feature-consistency loss limits generative drift. We also introduce MIUD, an 850-image multi-region in-situ underwater benchmark spanning depths from several meters to over 4,000 m in the Beibu Gulf, the South China Sea, and the Eastern Indian Ocean. Extensive experiments demonstrate that PD-GPO outperforms existing zero-shot methods on most evaluated metrics.

1 INTRODUCTION

Underwater image restoration has traditionally followed either physical inversion or data-driven learning. Physical methods estimate optical quantities and invert an assumed image formation model, but estimation errors can be amplified, and the assumptions of homogeneous water and uniform illumination often fail in real scenes. Supervised CNN-based (Li et al., 2020; 2021), Transformer-based (Peng et al., 2023), and diffusion-based methods (Shi & Wang, 2024; Bi et al., 2025; Zhao et al., 2026) learn underwater-to-reference mappings from paired, pseudo-paired, or

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

synthetic data. Unpaired methods relax the need for aligned references but still learn a task-specific mapping from an underwater training distribution (Fu et al., 2022; Jiang et al., 2023; Zeng et al., 2026). The scarcity and ambiguity of reference images, together with domain shifts across water bodies, imaging ranges, and acquisition conditions, can limit generalization to new conditions. Broader restoration frameworks also build on trained components: Reti-Diff (He et al., 2025) learns Retinex-based latent diffusion priors for illumination-degradation restoration, including underwater enhancement, and AgenticIR (Zhu et al., 2025) perceives the input and schedules specialized restoration tools.

Here, zero-shot refers specifically to the absence of cross-image underwater restoration training: no paired or unpaired underwater corpus is used to train or fine-tune a shared underwater-to-clean mapping, and each test observation is restored by independent image optimization (Ulyanov et al., 2018; Chihaoui & Favaro, 2025). This does not mean that every component is untrained, we do not train or fine-tune any component on the evaluated real underwater datasets the generic pretrained models are used as released.

Generic generative priors capture richer natural-image statistics than an untrained network and support blind restoration without an explicit degradation operator. However, such a prior does not indicate which structures and colors are still supported by an underwater observation, so unconstrained image optimization may introduce unsupported color compensation or hallucinated details. Recent physics-guided underwater methods (Du et al., 2025; Qiao et al., 2025) use transmission, background light, depth, or reconstructed radiance as model conditions, sampling weights, or analytic inversion variables. They do not, however, assign complementary spatial and chromatic roles to degradation fields that are re-estimated throughout image generative optimization.

We describe the observable degradation state with two input-dependent fields that play distinct roles. The spatial field s_I measures local structural degradation, while the chromatic field $\mathbf{c}_I$ captures the direction and strength of wavelength-dependent attenuation. Together, they provide complementary guidance on where and how to correct an image without prescribing a target color at every pixel.

Building on this decomposition, we propose Physical Degradation-Aware Generative Prior Optimization (PD-GPO). Following the zero-shot latent optimization paradigm of the diffusion image prior (DIIP) (Chihaoui & Favaro, 2025), PD-GPO optimizes an image latent under a frozen generative prior and adds physical degradation guidance to regulate the restoration. PDFE is trained and validated on 6,400 synthetic samples derived from natural RGB-D images and remains frozen during restoration. Given an observed image I, its spatial field s_I guides automatic adaptive contrast enhancement (AutoACE) (Martens, 1995) to construct a degradation-aware initialization $\hat{I}$. The image latent is then optimized under a frozen generative prior. At iteration k, the frozen PDFE re-estimates $s_J^{(k)}$ and $\mathbf{c}_J^{(k)}$ from the current candidate $J^{(k)}$: the former determines the support of the chromatic objective, and the latter provides the channel values aggregated over that support. We evaluate PD-GPO on three public benchmarks and on our multi-region in-situ underwater dataset (MIUD) of 850 images from the Beibu Gulf, the South China Sea, and the eastern Indian Ocean, complemented by component ablations and a computational-cost analysis.

The main contributions of this work are summarized as follows:

- We construct MIUD, an independently acquired multi-region in-situ dataset benchmark cover shallow to deep underwater observations, diverse scene content, and multiple color-appearance conditions. It extends evaluation beyond established public datasets and helps assess both the applicability and the limitations of image zero-shot restoration.
- We introduce a joint spatial–chromatic degradation representation consisting of a scalar spatial field and a normalized chromatic field. The input spatial field guides initialization; during optimization, spatial field selects the support over which chromatic field is aggregated for chromatic-ratio regularization.
- We develop PD-GPO, a zero-shot underwater restoration framework in which physical degradation cues guide both the initialization and the image optimization under a frozen generative prior, without training an underwater-to-clean mapping on any paired or unpaired underwater data.

2 RELATED WORK

Physics-guided underwater restoration. Underwater image formation couples wavelength- and range-dependent attenuation of direct radiance with backscatter, which motivates model-based recovery when the required optical quantities and depth can be estimated (Akkaynak & Treibitz, 2018; 2019). In unconstrained single-image settings, however, these estimates are ambiguous, and inversion may amplify their errors. A complementary line embeds physical cues in learned restoration through transmission-guided features, degradation-aware adversarial learning, and formation-model-constrained unpaired training (Li et al., 2021; Cong et al., 2023; Yan et al., 2023). These methods use physics to organize restoration, but their trained components still depend on underwater training data, unlike a frozen degradation estimator that guides optimization independently for each test image.

Supervised and unpaired underwater image enhancement. Supervised UIE learns underwater-to-reference mappings from paired or pseudo-paired images, as illustrated by fusion-based and hierarchical Transformer architectures (Li et al., 2019; Peng et al., 2023). Their outputs inherit the quality and coverage of the target images, which are difficult to acquire across water types and depths. Unpaired UIE relaxes the alignment requirement through reconstruction and homology constraints, adversarial or perceptual objectives, task-oriented contrastive feedback, and frequency-statistics transfer (Fu et al., 2022; Jiang et al., 2023; Liu et al., 2022; Zeng et al., 2026). These strategies broaden the usable training data but do not remove the dependence on an underwater image distribution; performance outside the observed acquisition conditions therefore remains an empirical question rather than a guarantee conferred by the absence of paired targets.

Generative priors and zero-shot restoration. Generative models have brought richer priors to UIE through physical or chromatic conditioning, frequency-domain diffusion, posterior guidance, latent scene decomposition, and color–structure disentanglement (Bi et al., 2025; Zhao et al., 2026; 2024; Du et al., 2025; Wu et al., 2025; Fan et al., 2026). A generative prior alone, however, does not make a method zero-shot: task-trained mappings or auxiliary enhancers may still introduce a dependence on underwater training. image approaches instead rely on test-image constraints; UIEAnything (Shao et al., 2025) and Zero-UMSIE (Liu et al., 2024) follow observation-guided physical and decomposition routes, respectively. Two physics-guided diffusion methods are closest to ours. Osmosis (Bar Nathan et al., 2024) samples from a joint RGB-D prior and enforces an underwater formation model through posterior guidance. PhyDiff (Qiao et al., 2025) estimates radiance and transmission for each observation and uses the spatially varying transmission to fuse the radiance with intermediate reverse-diffusion states. In contrast, we study how observable degradation fields should take complementary roles within image generative optimization: s and $\mathbf{c}$ are not identified with transmission or background-light estimates, but jointly form a spatial–chromatic description of the degradation.

3 METHOD

PD-GPO uses physical degradation estimates to guide image optimization under a frozen generative prior (Fig. 2). PDFE is pretrained on synthetic degradations of natural images, without any underwater data, and remains frozen during restoration. Given an observed underwater image I, PDFE predicts a spatial field s_I and a chromatic field $\mathbf{c}_I$, two complementary descriptions of the same spatially varying degradation state. The spatial field s_I guides (SFG) AutoACE to construct a degradation-aware initialization $\hat{I}$, from which an image latent is optimized under the frozen generative prior. At iteration k, the frozen PDFE evaluates the current candidate $J^{(k)}$ to obtain $s_J^{(k)}$ and $\mathbf{c}_J^{(k)}$; the former selects the support for chromatic refinement, and the latter supplies the channel values aggregated over it. A DINOv2-based feature-consistency objective (Oquab et al., 2024) preserves the structural and semantic content supported by the observation. The final candidate is returned as the restored image.

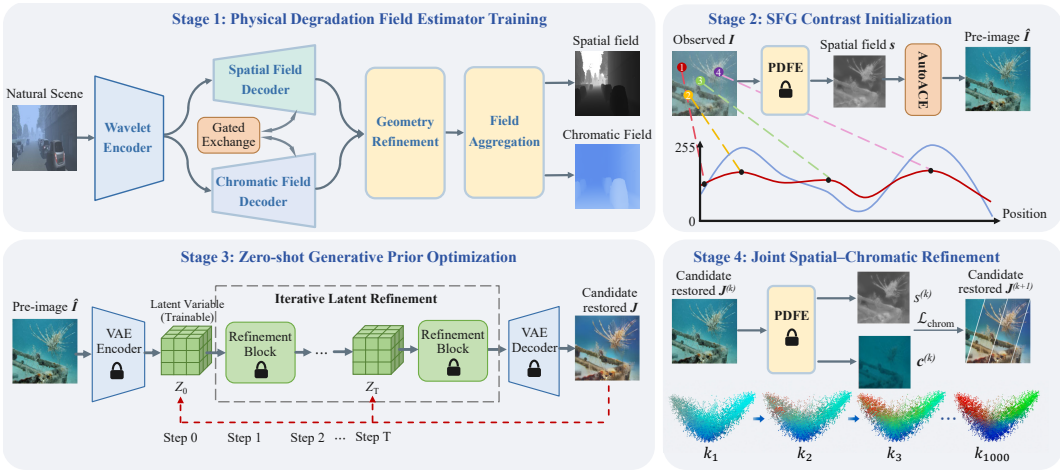

Figure 2: Overview of PD-GPO. PDFE is pretrained on natural images without underwater data and remains frozen during restoration. The spatial field guides the contrast initialization $\hat{I}$. During optimization under a frozen generative prior, $s_J^{(k)}$ selects the support and $\mathbf{c}_J^{(k)}$ supplies support-aggregated channel ratios for refining the image latent, while a DINOv2 feature-consistency loss preserves scene structure and semantics.

3.1 Physical Degradation Field Estimation

PD-GPO organizes physical degradation information into two quantities that directly support the restoration operations: a spatial field s and a chromatic field $\mathbf{c}$. We define them as

$$s(\mathbf{x}) = 1 - e^{-\beta d(\mathbf{x})}, \qquad c_{ch}(\mathbf{x}) = A_{ch} + (1 - A_{ch})e^{-\alpha_{ch} d(\mathbf{x})}, \tag{1}$$

where $d(\mathbf{x}) \geq 0$ denotes an effective propagation range, $\beta \geq 0$ controls how fast scene contrast decays with range, $\alpha_{ch} \geq 0$ is a channel-dependent decay rate, and $A_{ch} \in [0, 1]$ is a channel-dependent asymptotic response. PDFE does not estimate these variables at inference. The two fields enter the factorized degradation model

$$I(\mathbf{x}) = \mathbf{c}(\mathbf{x}) \odot \left[(1 - s(\mathbf{x}))J(\mathbf{x}) + s(\mathbf{x})\right], \tag{2}$$

in which $1 - s$ acts as a spatial contrast-retention factor and $\mathbf{c}(1 - s)$ gives the total channel-wise contrast retention. At zero range, $s = 0$ and $\mathbf{c} = \mathbf{1}$, so Eq. (2) reduces to $I = J$; as the range increases, s approaches one and c_{ch} approaches A_{ch}. The model thus links spatially varying structural degradation with chromatic distortion in a single representation. We use Eq. (2) only to synthesize supervision for PDFE; it is not imposed as a reconstruction constraint during test-time optimization.

PDFE estimates the spatial–chromatic degradation state from a single RGB image. As shown in Fig. 3, it adapts the wavelet encoder (Wen et al., 2023) into spatial and chromatic decoding branches with bidirectional gated feature exchange. A relative-depth head predicts geometric guidance from RGB features to refine both branches and their outputs, so no external depth is required at inference. The network outputs $s_I \in [0, 1]^{H \times W}$ and $\mathbf{c}_I \in [0, 1]^{H \times W \times 3}$. The input field s_I is used to construct the initialization, and during latent optimization the same frozen PDFE evaluates each candidate $J^{(k)}$ to obtain $s_J^{(k)}$ and $\mathbf{c}_J^{(k)}$. Appendices A provide further details.

3.2 Spatial-Field-Guided Contrast Initialization

The initialization modifies only the value channel I_V of the observed image in HSV space and keeps its hue and saturation. At each pixel, AutoACE selects a local window from multi-scale statistics and adapts the contrast gain to local image characteristics:

$$I_{\mathrm{ACE},V}(\mathbf{x}) = \mu_{r(\mathbf{x})}(I_V) + g_{r(\mathbf{x})}\left[I_V(\mathbf{x}) - \mu_{r(\mathbf{x})}(I_V)\right], \tag{3}$$

where $r(\mathbf{x})$ is the adaptively selected window scale, μ is the corresponding local mean, and g is the adaptive gain determined from multi-scale local statistics. We apply automatic gamma correction to $I_{\mathrm{ACE},V}$, recombine it with the original hue and saturation channels, and convert the result to RGB to

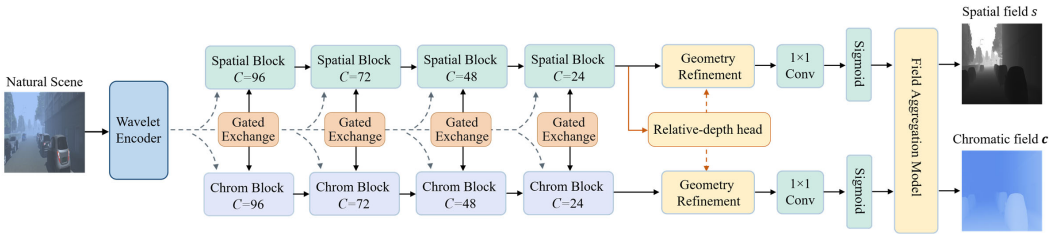

Figure 3: Architecture of PDFE. A shared wavelet encoder feeds separate spatial and chromatic decoders coupled by gated exchange. A relative-depth head guides feature refinement and the depth-affinity aggregation of the output fields.The dashed line represents scale matched with feature.

obtain I_{ACE}. Keeping the observation as the baseline, we then modulate the contrast residual with the spatial field:

$$\hat{I} = I + \eta\, s_I \odot (I_{\mathrm{ACE}} - I)\,, \tag{4}$$

where η controls the overall enhancement strength. Regions with larger estimated degradation thus receive stronger contrast compensation, whereas more reliable regions change less. We normalize $\hat{I}$ to $[-1, 1]$ and encode it with the pretrained VAE; the posterior mean, rather than a random sample, is multiplied by the VAE scaling factor to give the initial latent for generative optimization.

3.3 Zero-shot Generative Prior Optimization

PD-GPO follows the image optimization of DIIP: a latent is optimized for each observation while the generative model stays frozen. We instantiate the prior with FLUX.2-klein-4B (Black Forest Labs, 2026). The VAE encoding of $\hat{I}$ serves both as a fixed visual-conditioning latent $\mathbf{y}_{\hat{I}}$ and, through its posterior mean, as the initial trainable latent $\mathbf{z}^{(0)}$. Given a latent $\mathbf{z}$, the frozen pipeline $\mathcal{G}_\theta$ follows its rectified-flow trajectory (Liu et al., 2023) conditioned on $\mathbf{y}_{\hat{I}}$ and decodes the result into $J(\mathbf{z}) = \mathcal{G}_\theta(\mathbf{z}; \mathbf{y}_{\hat{I}})$. We optimize

$$\mathbf{z}^\star = \arg\min_{\mathbf{z}} \mathcal{L}_{\text{PD-GPO}}\big(J(\mathbf{z}); \hat{I}\big), \tag{5}$$

where the fields used in the objective are re-estimated from each candidate $J(\mathbf{z})$; Section 3.5 defines the objective.

3.4 Joint Spatial–Chromatic Refinement

Spatial-field-guided initialization improves visibility but does not, by itself, correct wavelength-dependent chromatic attenuation. Whole-image gray-world balancing is also unreliable when non-gray objects dominate a scene or when distant regions retain little color information. We therefore restrict the chromatic objective to a support selected from the current candidate.

At iteration k, the frozen PDFE evaluates $J^{(k)}$ to obtain $s_J^{(k)}$ and $\mathbf{c}_J^{(k)}$. Unlike s_I, which guides initialization, $s_J^{(k)}$ defines the support for chromatic regularization through the mask $m_J^{(k)}(\mathbf{x}) = \mathbf{1}[s_J^{(k)}(\mathbf{x}) \le \tau]$, which keeps the pixels that PDFE regards as only mildly degraded; τ is a threshold. With $\Omega_J^{(k)} = \{\mathbf{x} : m_J^{(k)}(\mathbf{x}) = 1\}$, the chromatic loss is

$$\mathcal{L}_{\text{chrom}}^{(k)} = \sum_{ch\in\{G,B\}} \left| \frac{\sum_{\mathbf{x}\in\Omega_J^{(k)}} c_{J,ch}^{(k)}(\mathbf{x})}{\sum_{\mathbf{x}\in\Omega_J^{(k)}} c_{J,R}^{(k)}(\mathbf{x}) + \epsilon\big(|\Omega_J^{(k)}| + \epsilon\big)} - 1 \right|. \tag{6}$$

This expression is the green (or blue) mean divided by the red mean plus ϵ, where each channel mean is computed over $\Omega_J^{(k)}$ with denominator $|\Omega_J^{(k)}| + \epsilon$; both ϵ terms ensure numerical stability. The loss drives the support-aggregated green-to-red and blue-to-red ratios toward one without imposing this balance on the whole image. Because the fields are re-estimated from $J^{(k)}$ at every iteration, both $\Omega_J^{(k)}$ and the channel statistics are updated as optimization proceeds. The hard threshold passes no gradient through $s_J^{(k)}$. If $\Omega_J^{(k)}$ is empty, the feature term alone drives that iteration.

3.5 Physical Degradation-Aware Generative Prior Optimization

For each observation, PD-GPO combines the preceding stages. The field s_I guides the construction of $\hat{I}$, whose VAE encoding supplies $\mathbf{y}_{\hat{I}}$ and initializes $\mathbf{z}^{(0)}$. At each iteration, $\mathcal{G}_\theta$ generates $J^{(k)}$, the frozen PDFE re-estimates $(s_J^{(k)}, \mathbf{c}_J^{(k)})$, and these estimates enter the chromatic loss in Eq. (6). No parameter of the generative pipeline, PDFE, or feature extractor is updated.

To discourage content drift, we compare the candidate with the observation-derived initialization using selected intermediate features of a frozen DINOv2 extractor Φ. These layers capture high-level semantic and structural information, so the constraint preserves content without over-constraining low-level appearance:

$$\mathcal{L}_{\text{feat}}^{(k)} = \left\| \Phi\left(J^{(k)}\right) - \Phi\left(\hat{I}\right) \right\|_1 . \tag{7}$$

The per-iteration objective is

$$\mathcal{L}_{\text{PD-GPO}}^{(k)} = \lambda_{\text{feat}} \mathcal{L}_{\text{feat}}^{(k)} + \lambda_{\text{chrom}} \mathcal{L}_{\text{chrom}}^{(k)}, \tag{8}$$

where the weights balance feature consistency and chromatic refinement. Gradients of the feature loss and, when the support is nonempty, of the chromatic field values propagate through the generated candidate to the latent; the objective contains no reconstruction residual obtained by applying Eq. (2) to the candidate. The overall feedback loop is

$$\mathbf{z}^{(k)} \xrightarrow{\mathcal{G}_\theta} J^{(k)} \xrightarrow{f_{\text{PDFE}}} \left(s_J^{(k)}, \mathbf{c}_J^{(k)}\right) \xrightarrow{\mathcal{L}_{\text{PD-GPO}}^{(k)}} \mathbf{z}^{(k+1)}, \tag{9}$$

and after K updates the final candidate $J^{(K)}$ is returned as the restored image.

4 Experiments

4.1 Experimental Settings

PD-GPO implementation. We implement PD-GPO in PyTorch on a single NVIDIA RTX 3090 GPU (24 GB) with frozen FLUX.2-klein-4B and PDFE models. Each input is resized to 512×512, and the output has the same resolution; Table 5 additionally reports the cost at 256×256. The VAE posterior mean of $\hat{I}$ initializes the latent, which is optimized with $K = 1{,}000$ Adam (Kingma & Ba, 2015) updates at a learning rate of 5×10^{-3}, using $step = 1$ FLUX.2-klein-4B sampling steps per update. We set the support threshold to $\tau = 0.6$. The feature loss uses frozen DINOv2 ViT-S/14 features from layers 8–11, with $\lambda_{\text{feat}} = \lambda_{\text{chrom}} = 1.0$.

PDFE training. PDFE is trained at 512×512 derived from natural RGB-D images: 1,822 from NYU Depth V2 (Silberman et al., 2012) and 4,578 from Virtual KITTI 2 (Cabon et al., 2020). They comprise 6,000 synthetically degraded and 400 undegraded samples, split into 5,760 training and 640 validation samples with no source image shared between the splits; Appendices A and B describe the synthesis. No real underwater image from UIEB, LSUI, EUVP, or MIUD is used for PDFE training or validation.

4.2 Comparison on Public Underwater Datasets

Datasets and protocol. We evaluate on UIEB (Li et al., 2019), LSUI (Peng et al., 2023), and EUVP (Islam et al., 2020). For each dataset, we fixed a subset of 100 images before examining the results of PD-GPO and evaluated all methods on the same subset. The primary comparison covers the zero-shot methods GDP (Fei et al., 2023), PhyDiff (Qiao et al., 2025), UIEAnything (Shao et al., 2025), Zero-UMSIE (Liu et al., 2024), and Osmosis (Bar Nathan et al., 2024), run with their official implementations and recommended configurations when available. Supervised and unpaired methods are compared separately in Appendix C.

Quantitative comparison. We report four complementary no-reference metrics, UCIQE (Yang & Sowmya, 2015), UIQM (Panetta et al., 2016), PAUQA (Lin et al., 2025), and MUSIQ (Ke et al., 2021), and, where references are available, PSNR (Huynh-Thu & Ghanbari, 2008), SSIM (Wang et al., 2004), and LPIPS (Zhang et al., 2018).

Table 1: No-reference scores on the EUVP, LSUI, and UIEB subsets. In Tables 1–3, the best and second-best results are shown in **red** and blue, respectively.

Method	EUVP				LSUI				UIEB			
	UCIQE↑	UIQM↑	PAUQA↑	MUSIQ↑	UCIQE↑	UIQM↑	PAUQA↑	MUSIQ↑	UCIQE↑	UIQM↑	PAUQA↑	MUSIQ↑
GDP	0.354	6.153	0.254	40.043	0.300	6.239	0.252	42.361	0.334	5.279	0.495	52.637
PhyDiff	0.420	7.158	0.374	39.446	0.379	**6.920**	0.351	48.104	0.373	6.103	0.396	49.123
UIEAnything	0.237	6.724	0.422	38.142	0.209	6.444	0.338	20.772	0.235	6.453	0.580	60.774
Zero-UMSIE	0.167	5.900	0.440	35.365	0.179	6.523	0.480	46.267	0.223	5.479	0.608	55.972
Osmosis	0.381	6.115	0.412	35.278	0.352	5.582	0.380	34.048	0.358	5.840	0.404	39.914
Ours	**0.502**	**7.284**	**0.483**	**41.005**	**0.414**	6.737	**0.591**	**48.206**	**0.412**	**7.436**	**0.742**	**62.533**

Table 2: Full-reference scores on the EUVP, LSUI, and UIEB subsets.

Method	PSNR↑			SSIM↑			LPIPS↓		
	EUVP	LSUI	UIEB	EUVP	LSUI	UIEB	EUVP	LSUI	UIEB
GDP	19.623	16.627	15.000	**0.854**	0.829	0.742	0.215	0.175	0.326
PhyDiff	20.268	18.754	18.883	0.810	0.811	0.826	0.242	0.203	0.206
UIEAnything	19.752	19.218	20.392	0.808	0.692	0.814	0.310	0.483	0.175
Zero-UMSIE	17.977	18.525	19.315	0.731	0.792	**0.840**	0.328	0.245	0.168
Osmosis	17.419	17.907	19.738	0.800	0.849	0.824	0.314	0.262	0.283
Ours	**21.468**	**23.630**	**21.544**	0.811	**0.856**	0.820	**0.207**	**0.154**	**0.164**

Table 3: No-reference scores on the MIUD subset.

Method	MIUD			
	UCIQE↑	UIQM↑	PAUQA↑	MUSIQ↑
GDP	0.250	5.523	0.401	59.872
PhyDiff	0.344	6.224	0.392	46.373
UIEAnything	0.241	6.000	0.609	**64.757**
Zero-UMSIE	0.189	5.997	0.565	63.500
Osmosis	0.335	5.130	0.364	40.627
Ours	**0.373**	**6.565**	**0.705**	64.044

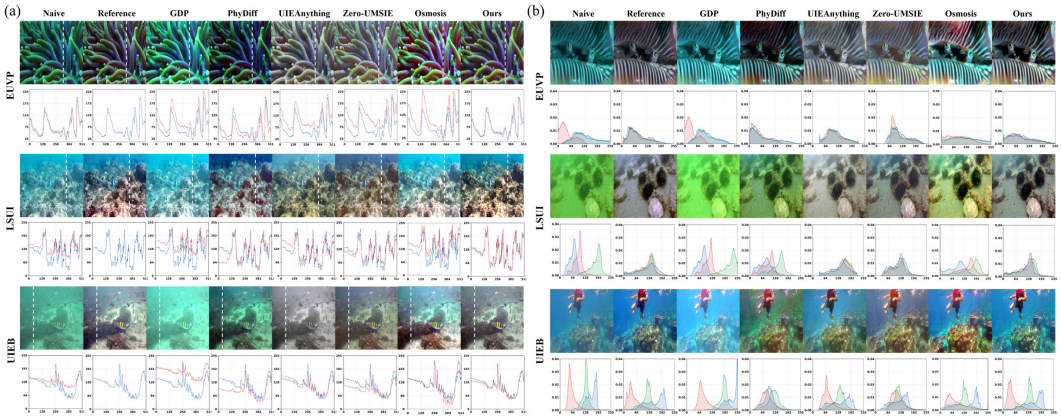

Figure 4: Qualitative comparison on EUVP, LSUI, and UIEB. (a) Representative images and vertical intensity profiles along the white dashed lines; blue curves denote the references and red curves the inputs or restored images (axes: sampling position and 8-bit intensity). (b) RGB intensity histograms of representative images, with red, green, and blue curves for the corresponding channels.

Across EUVP, LSUI, and UIEB, PD-GPO obtains the best UCIQE and PAUQA together with the highest PSNR and lowest LPIPS (Tables 1 and 2). The agreement between no-reference gains and reference fidelity suggests that spatially selective chromatic correction improves underwater appearance while remaining faithful to the reference, rather than enforcing a global color balance. The main exceptions are UIQM on LSUI, where PhyDiff scores higher (6.920 vs. 6.737), and SSIM on EUVP and UIEB, where GDP (0.854) and Zero-UMSIE (0.840) lead, respectively. UIQM combines colorfulness, sharpness, and contrast, and on LSUI PD-GPO still leads all three full-reference metrics.

Qualitative comparison. Figure 4(a) compares intensity profiles along the marked columns. In the illustrated scenes, PD-GPO follows the principal peaks and troughs of the reference profiles more closely, including the low-intensity region in the fish example, which distinguishes recovery of local intensity variation from a uniform increase in brightness. Figure 4(b) compares RGB intensity histograms. Several inputs show pronounced channel separation, and the methods differ in how they

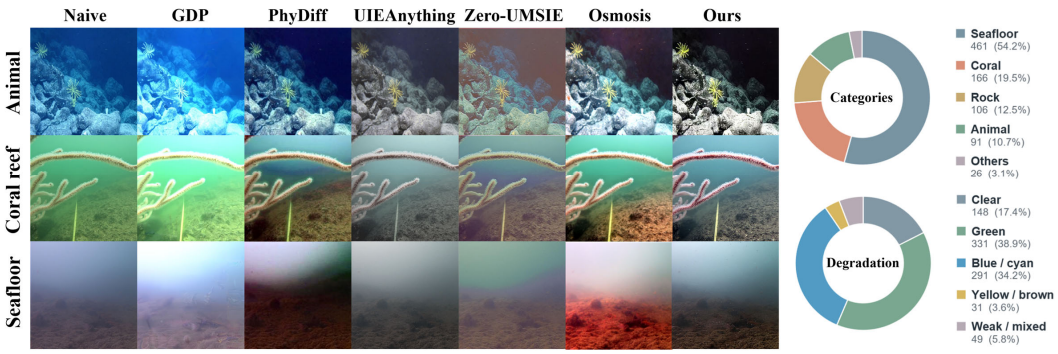

Figure 5: MIUD qualitative comparison and dataset composition. Left: Results for animal (sea anemone), coral-reef, and seafloor scenes (top to bottom). Right: Distributions of scene categories (top) and dominant degradation appearances (bottom).

redistribute the three channels. In the illustrated cases, the distributions of PD-GPO are closer to those of the references, consistent with the visible reduction of color casts.

4.3 Evaluation on the MIUD Benchmark

Multi-region in-situ underwater dataset. MIUD contains 850 independently collected RGB observations from the Beibu Gulf, the South China Sea, and the eastern Indian Ocean, stored at 1024×1024 resolution. The images span four acquisition-depth strata, from several meters to more than 4,000 m, and are annotated with five scene categories and five dominant degradation appearances (Fig. 5). These labels describe the observations; they are not clean-image references.

Quantitative and qualitative comparison. Because MIUD has no clean references, we evaluate all methods with no-reference metrics on 100 images randomly sampled across depth strata and scene categories. As shown in Table 3, PD-GPO ranks first on UCIQE, UIQM, and PAUQA and second on MUSIQ, behind UIEAnything (64.757 vs. 64.044). In the examples of Fig. 5, PD-GPO recovers foreground boundaries and local texture without the over-brightening, halos, or regional color shifts visible in several compared outputs. These subset scores and examples indicate that field-guided optimization is competitive under the MIUD protocol and extend the evaluation beyond established public datasets; they do not establish performance over all 850 images.

4.4 Ablation Study

We ablate spatial-field-guided initialization (SFG) and joint spatial–chromatic refinement on the same UIEB subset (Table 4). *Base* is our internal latent-optimization baseline: the generative prior is frozen, only the image latent is optimized, and $\mathcal{L}_{\text{feat}}$ is the sole objective. Table 4 presents two controlled comparisons: one varies the chromatic loss and its spatial support while retaining the full SFG initialization; the other varies the initialization while retaining the full chromatic loss.

Spatial-field-guided initialization. With the full chromatic loss retained, AutoACE initialization without s improves over removing SFG on UIQM, PAUQA, and MUSIQ (UCIQE is unchanged at 0.400), and guiding it with s improves all four metrics further. This progression separates the value of local contrast initialization from the added value of guiding it with the predicted spatial field. The clearer fine structures and local contrast in the top block of Fig. 6 are consistent with a more informative starting point for latent optimization.

Joint spatial–chromatic refinement. With SFG retained, adding $\mathcal{L}_{\text{chrom}}$ without s improves all four metrics over removing the chromatic loss, and restricting the loss to the support selected by s improves them further. Chromatic guidance is therefore useful on its own, and spatial support makes it more effective. Specifically, Eq. 6 constrains channel ratios aggregated over the selected support; its spatial selectivity lies in that support selection, rather than in prescribing a separate color-correction direction at every pixel. In the bottom block of Fig. 6, the full model corrects foreground

Table 4: Ablation of SFG and chromatic refinement.

Module	Variant	UCIQE↑	UIQM↑	PAUQA↑	MUSIQ↑
–	Base	0.362	6.180	0.467	56.767
SFG	w/o SFG	0.400	6.800	0.631	56.171
	SFG (w/o s)	0.400	7.052	0.720	60.838
L_{chrom}	w/o L_{chrom}	0.370	6.916	0.592	59.280
	L_{chrom} (w/o s)	0.377	7.021	0.710	61.464
–	Ours (full, w/ s)	**0.412**	**7.436**	**0.742**	**62.533**

Table 5: Computational cost.

Method	Params (M)	FLOPs (T)	Time (s/img)
GDP	552.814	9104.607	349.825
PhyDiff	553.672	10348.257	397.698
UIEAnything	37.883	0.337	2.443
Zero-UMSIE	0.127	0.066	1.277
Osmosis	552.821	8913.404	399.313
Ours_{256}	3992.931	862.165	76.174
Ours_{512}	3992.931	2860.404	181.102

color while better preserving the surrounding regions and local texture. These point estimates on the evaluated subset are consistent with complementary contributions from the two components.

Computational cost. Image latent optimization is considerably slower than the fastest baselines, yet PD-GPO requires less time and fewer total FLOPs than GDP, PhyDiff, and Osmosis under the reported settings (Table 5). All methods were evaluated on an NVIDIA RTX 3090. PD-GPO used 1,000 image-latent optimization updates and one rectified-flow step per update. Baselines used their released checkpoints and method-specific inference settings. We measured 100 images per method. The large parameter count comes from the frozen generative prior, VAE, PDFE, and DINOv2 extractor; only the image latent is optimized. Because the cost grows with resolution, PD-GPO is better suited to quality-oriented offline restoration than to latency-sensitive deployment.

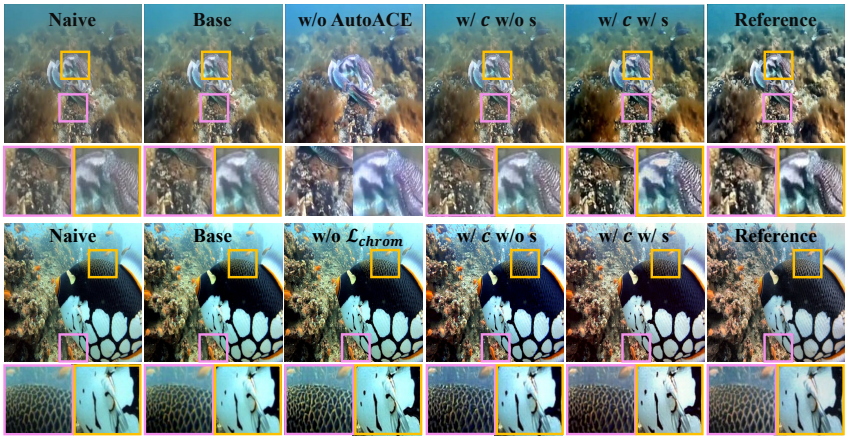

Figure 6: Qualitative ablations on two underwater scenes. Top: spatial-field-guided initialization; Bottom: joint spatial–chromatic refinement. Insets enlarge the marked regions to show differences in local contrast, color, and texture.

5 Conclusion

We presented PD-GPO, a zero-shot underwater restoration framework that combines spatial-field-guided initialization and joint spatial–chromatic refinement with image latent optimization under a frozen generative prior. A frozen PDFE re-estimates both fields from each candidate, and a frozen DINOv2 feature-consistency constraint limits content drift; no paired or unpaired underwater images are used to train an underwater-to-clean mapping. Results on three public datasets and the MIUD benchmark support this design under the evaluated protocol. The method still requires one to three minutes per image, and it can overcorrect under extreme color casts or severe scattering.

AI use statement

We used generative AI tools to help draft and edit the manuscript text and to check the LaTeX formatting. Generative AI also helped create the underwater degradation principle in Fig. 1(a) and the changes of color distribution in Stage 4 of Fig. 2. The authors reviewed and revised all AI-assisted text and visual elements and take full responsibility for the manuscript.

Reproducibility statement

Section 3 specifies the PD-GPO formulation and the image-specific optimization procedure, Section 4 describes the datasets, evaluation protocol, and metrics, and the appendix provides further implementation and experimental details. Upon acceptance, we will release the PD-GPO implementation, the PDFE network and synthetic-data generation code, the MIUD dataset, and the synthetic degradation dataset used for PDFE training and validation.

References

Derya Akkaynak and Tali Treibitz. A revised underwater image formation model. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6723–6732. IEEE, 2018.

Derya Akkaynak and Tali Treibitz. Sea-thru: A method for removing water from underwater images. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1682–1691. IEEE, 2019.

Opher Bar Nathan, Deborah Levy, Tali Treibitz, and Dan Rosenbaum. Osmosis: Rgbd diffusion prior for underwater image restoration. In *European Conference on Computer Vision*, pp. 302–319, 2024.

Hengyue Bi, Long Chen, Jingchao Cao, Jingyang Wang, Jinghao Sun, Yuan Rao, and Junyu Dong. Seadiff: Underwater image enhancement with degradation-aware diffusion model. *IEEE Transactions on Circuits and Systems for Video Technology*, 2025.

Black Forest Labs. FLUX.2 [klein] 4B: Model Card. `https://huggingface.co/black-forest-labs/FLUX.2-klein-4B`, 2026.

Yohann Cabon, Naila Murray, and Martin Humenberger. Virtual KITTI 2. *arXiv preprint arXiv:2001.10773*, 2020. doi: 10.48550/arXiv.2001.10773. URL `https://arxiv.org/abs/2001.10773`.

Hamadi Chihaoui and Paolo Favaro. Diffusion image prior. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 24636–24644. IEEE, 2025.

Runmin Cong, Wenyu Yang, Wei Zhang, Chongyi Li, Chun-Le Guo, Qingming Huang, and Sam Kwong. Pugan: Physical model-guided underwater image enhancement using gan with dual-discriminators. *IEEE Transactions on Image Processing*, 32:4472–4485, 2023.

Dazhao Du, Enhan Li, Lingyu Si, Wenlong Zhai, Fanjiang Xu, Jianwei Niu, and Fuchun Sun. Uiedp: Boosting underwater image enhancement with diffusion prior. *Expert Systems with Applications*, 259:125271, 2025.

Guodong Fan, Yu Zhou, Jingchun Zhou, Yakun Ju, Guang-Yong Chen, Jinjiang Li, and Alex C Kot. Dcd-uie: Decoupled chromatic diffusion model for underwater image enhancement. *IEEE Transactions on Image Processing*, 2026.

Ben Fei, Zhaoyang Lyu, Liang Pan, Junzhe Zhang, Weidong Yang, Tianyue Luo, Bo Zhang, and Bo Dai. Generative diffusion prior for unified image restoration and enhancement. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9935–9946. IEEE, 2023.

Zhenqi Fu, Huangxing Lin, Yan Yang, Shu Chai, Liyan Sun, Yue Huang, and Xinghao Ding. Unsupervised underwater image restoration: From a homology perspective. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pp. 643–651, 2022.

Chunming He, Chengyu Fang, Yulun Zhang, Longxiang Tang, Jinfa Huang, Kai Li, Zhenhua Guo, Xiu Li, and Sina Farsiu. Reti-Diff: Illumination degradation image restoration with retinex-based latent diffusion model. In *The Thirteenth International Conference on Learning Representations*, 2025.

Shirui Huang, Keyan Wang, Huan Liu, Jun Chen, and Yunsong Li. Contrastive semi-supervised learning for underwater image restoration via reliable bank. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 18145–18155, 2023.

Quan Huynh-Thu and Mohammed Ghanbari. Scope of validity of psnr in image/video quality assessment. *Electronics Letters*, 44(13):800–801, 2008. doi: 10.1049/el:20080522.

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

Qiuping Jiang, Yaozu Kang, Zhihua Wang, Wenqi Ren, and Chongyi Li. Perception-driven deep underwater image enhancement without paired supervision. *IEEE Transactions on Multimedia*, 26:4884–4897, 2023.

Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5148–5157, 2021. doi: 10.1109/ICCV48922.2021.00510.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015.

Chongyi Li, Chunle Guo, Wenqi Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao. An underwater image enhancement benchmark dataset and beyond. *IEEE transactions on image processing*, 29:4376–4389, 2019.

Chongyi Li, Saeed Anwar, and Fatih Porikli. Underwater scene prior inspired deep underwater image and video enhancement. *Pattern recognition*, 98:107038, 2020.

Chongyi Li, Saeed Anwar, Junhui Hou, Runmin Cong, Chunle Guo, and Wenqi Ren. Underwater image enhancement via medium transmission-guided multi-color space embedding. *IEEE Transactions on Image Processing*, 30:4985–5000, 2021.

Yuanyuan Li, Zetian Mi, Peng Lin, and Xianping Fu. Underwater optical image enhancement via an unsupervised contrastive learning-guided diffusion model. *Optics and Lasers in Engineering*, 194:109144, 2025.

Bosen Lin, Junyu Dong, and Xinghui Dong. Perception-aware underwater image quality assessment: Dataset, perceptual quality scores, and assessment network. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(11):11113–11128, 2025. doi: 10.1109/TCSVT.2025.3574860.

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

Tong Liu, Kaiyan Zhu, Weiye Cao, Bolin Shan, and Fangyi Guo. Zero-umsie: a zero-shot underwater multi-scale image enhancement method based on isomorphic features. *Optics Express*, 32 (23):40398–40415, 2024.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *International Conference on Learning Representations*, 2023.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019.

Jean-Bernard Martens. Adaptive contrast enhancement through residue-image processing. *Signal Processing*, 44(1):1–18, 1995.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, et al. DINOv2: Learning robust visual features without supervision. *Transactions on Machine Learning Research*, 2024.

Karen Panetta, Chen Gao, and Sos S. Agaian. Human-visual-system-inspired underwater image quality measures. *IEEE Journal of Oceanic Engineering*, 41(3):541–551, 2016. doi: 10.1109/JOE.2015.2469915.

Lintao Peng, Chunli Zhu, and Liheng Bian. U-shape transformer for underwater image enhancement. *IEEE Transactions on Image Processing*, 32:3066–3079, 2023.

Yuanjian Qiao, Mingwen Shao, Lingzhuang Meng, and Wangmeng Zuo. Learning physical-aware diffusion priors for zero-shot restoration of scattering-affected images. *Pattern Recognition*, 163: 111473, 2025.

Jinxin Shao, Haosu Zhang, and Jianming Miao. Uieanything: zero-shot underwater image enhancement via advanced depth estimation, white balance models, and improved sea-thru. *Pattern Analysis and Applications*, 28(2):68, 2025.

Gaurav Sharma, Wencheng Wu, and Edul N. Dalal. The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research & Application*, 30(1):21–30, 2005. doi: 10.1002/col.20070.

Xiaowen Shi and Yuan-Gen Wang. Cpdm: Content-preserving diffusion model for underwater image enhancement. *Scientific Reports*, 14(1):31309, 2024.

Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. Indoor segmentation and support inference from rgbd images. In *European Conference on Computer Vision*, pp. 746–760. Springer, 2012.

Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9446–9454. IEEE, 2018.

Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612, 2004. doi: 10.1109/TIP.2003.819861.

Junjie Wen, Jinqiang Cui, Zhenjun Zhao, Ruixin Yan, Zhi Gao, Lihua Dou, and Ben M Chen. Syreanet: A physically guided underwater image enhancement framework integrating synthetic and real images. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5177–5183. IEEE, 2023.

Jiayi Wu, Tianfu Wang, Md Abu Bakr Siddique, Md Jahidul Islam, Cornelia Fermuller, Yiannis Aloimonos, and Christopher A. Metzler. Single-step latent diffusion for underwater image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025. doi: 10.1109/TPAMI.2025.3599775.

Shuaizheng Yan, Xingyu Chen, Zhengxing Wu, Min Tan, and Junzhi Yu. Hybrur: A hybrid physical-neural solution for unsupervised underwater image restoration. *IEEE Transactions on image processing*, 32:5004–5016, 2023.

Miao Yang and Arcot Sowmya. An underwater color image quality evaluation metric. *IEEE Transactions on Image Processing*, 24(12):6062–6071, 2015. doi: 10.1109/TIP.2015.2491020.

Runhao Zeng, Xionglin Zhu, Wenfu Peng, Jiezhang Cao, Yong Guo, Zhihua Wang, and Qiuping Jiang. Amplitude exchanging network for unsupervised underwater image enhancement. *Pattern Recognition*, pp. 113119, 2026.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586–595, 2018. doi: 10.1109/CVPR.2018.00068.

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

Chen Zhao, Chenyu Dong, Weiling Cai, and Yueyue Wang. Learning a physical-aware diffusion model based on transformer for underwater image enhancement. *IEEE transactions on geoscience and remote sensing*, 2026.

Kaiwen Zhu, Jinjin Gu, Zhiyuan You, Yu Qiao, and Chao Dong. An intelligent agentic system for complex image restoration problems. In *The Thirteenth International Conference on Learning Representations*, 2025.

A SYNTHETIC SPATIAL–CHROMATIC FIELD DATASET

As shown in Table 6, to cover diverse spatial–chromatic conditions, the 6,000 degraded samples are evenly distributed across blue, green, yellow, and approximately neutral chromatic profiles. Five chromatic-strength levels are considered, together with multiple spatial-degradation and simulated-distance intervals. These factors are crossed during sampling rather than varied independently in isolated subsets. In addition, 900 samples contain a weak subpixel blur augmentation with a standard deviation of 0.25, 0.35, or 0.45 pixels. This augmentation is applied to the source appearance before degradation synthesis and does not alter the definitions of the s and $\mathbf{c}$ labels. Depth is used only for generating the synthetic fields and providing auxiliary geometric supervision during training. At inference, PDFE takes a single RGB image as input and predicts s_I and $\mathbf{c}_I$ without external depth.

Table 6: Distribution of 6,400 synthetic samples by chromatic condition and spatial-degradation level. The dataset contains 6,000 degraded samples across four chromatic profiles and 400 clean samples.

Chromatic Condition	**Spatial Degradation Level**				**Total**
	None	Light	Medium	Heavy	
Blue	375	450	450	225	1,500
Green	375	450	450	225	1,500
Yellow	375	450	450	225	1,500
Neutral	375	450	450	225	1,500
Clean	400	0	0	0	400
Total	**1,900**	**1,800**	**1,800**	**900**	**6,400**

B PDFE IMPLEMENTATION DETAILS AND ESTIMATION RESULTS

PDFE Implementation Details PDFE adopts the contour-response architecture described in the main paper. The shared wavelet encoder and the two field decoders are jointly optimized; the encoder is not frozen during PDFE training (Wen et al., 2023). The decoder channel widths from the coarsest to the finest scale are 96, 72, 48, and 24, respectively. Both decoders progressively recover spatial resolution through bilinear upsampling and convolutional fusion with encoder skip features. The spatial decoder incorporates normalized high-frequency wavelet energies, local contrast and gradient cues, and spatial context. The chromatic decoder combines low-frequency features, linear-RGB descriptors of intensity and channel relationships, and global intensity context. Bidirectional gated feature exchange is performed at every scale.

The final fused spatial features are passed to a relative-depth prediction head. The predicted relative depth guides boundary-sensitive feature refinement in both branches and the subsequent depth-affinity aggregation of the four output channels, limiting field mixing across inferred depth discontinuities. Separate 1×1 convolutional heads followed by sigmoid activations produce the single-channel spatial field $\hat{s}$ and the three-channel chromatic field $\hat{\mathbf{c}}$. The bounded chromatic components retain variation in both RGB orientation and magnitude. Relative depth is predicted from RGB at inference; no external depth input is used.

We train PDFE for 100 epochs using AdamW (Loshchilov & Hutter, 2019) with a batch size of 4 and weight decay of 10^{-4}. The shared encoder-decoder body uses a learning rate of 1×10^{-4}, while the newly introduced field-prediction heads use 2×10^{-4}. We retain the checkpoint with the best validation performance. All learnable components, including the wavelet encoder, the spatial and chromatic decoders, the gated feature-exchange modules, the relative-depth head, and the field prediction heads, are jointly optimized. Source depth is used for synthesis and auxiliary supervision, but inference requires only RGB input.

PDFE is optimized using four complementary objectives:

$$\mathcal{L}_{\mathrm{PDFE}} = 1.0\,\mathcal{L}_{\mathrm{field}} + 0.1\,\mathcal{L}_{\mathrm{chrom}} + 0.2\,\mathcal{L}_{\mathrm{struct}} + 0.1\,\mathcal{L}_{\mathrm{cons}}. \tag{10}$$

Each grouped loss retains the internal weights of its constituent terms. The field loss $\mathcal{L}_{\mathrm{field}}$ directly supervises the predicted spatial field $\hat{s}$ and chromatic field $\hat{\mathbf{c}}$ using their synthetic labels, constraining

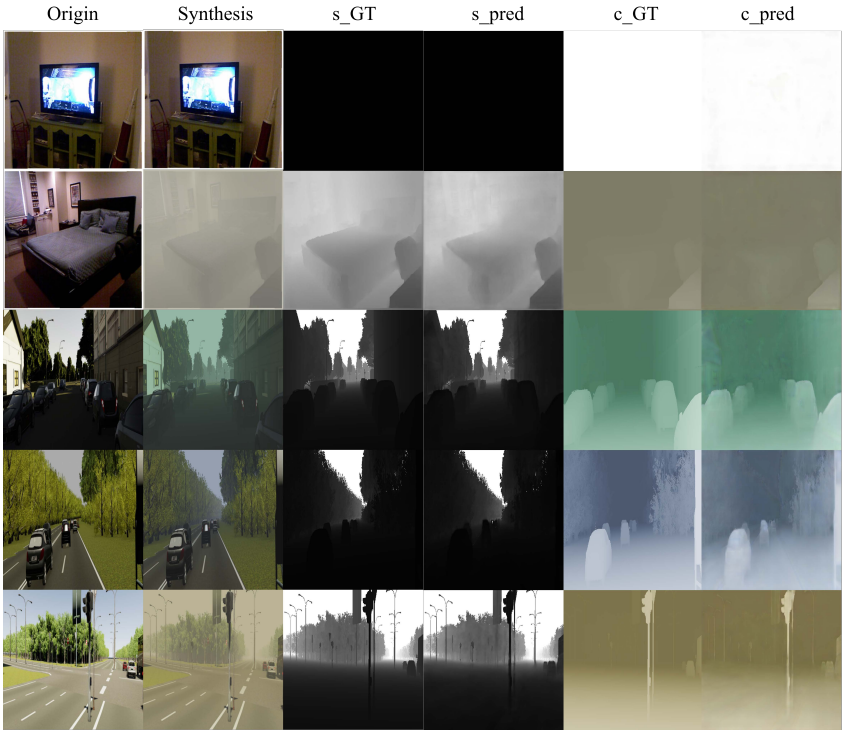

Figure 7: PDFE predictions compared with known field labels on held-out synthetic validation scenes. Columns, from left to right: natural scene image $\mathbf{J}$, synthesized observation $\mathbf{I}$, target spatial field s, predicted spatial field s_I, target chromatic field $\mathbf{c}$, and predicted chromatic field $\mathbf{c}_I$. The indoor and driving examples originate from NYU Depth V2 (Silberman et al., 2012) and Virtual KITTI 2 (Cabon et al., 2020), respectively.

both field values and their multi-scale spatial distributions. The chromatic loss $\mathcal{L}_{\text{chrom}}$ constrains the overall magnitude and relative channel pattern of $\hat{\mathbf{c}}$, thereby preserving its attenuation strength and orientation in color space. The structure loss $\mathcal{L}_{\text{struct}}$ incorporates local field variations, field boundaries, and relative-depth guidance to improve spatial coherence while retaining discontinuities associated with scene geometry. Finally, the consistency loss $\mathcal{L}_{\text{cons}}$ substitutes the predicted fields into the proposed spatial–chromatic degradation formulation and requires the re-synthesized image to reproduce the corresponding synthetic observation. It also anchors undegraded samples to $\hat{s} = 0$ and $\hat{\mathbf{c}} = \mathbf{1}$. All objectives are evaluated only over valid supervision regions, and the reference image and depth are used during training only rather than being supplied as network inputs.

Quantitative and qualitative analysis. We evaluate PDFE by comparing its predictions with the synthetic field labels on a deterministic subset of the validation split. Specifically, 200 samples are selected without replacement from the 640 validation samples using a fixed random seed. The evaluated subset contains 57 NYU Depth V2 samples and 143 Virtual KITTI 2 samples, including 185 degraded samples and 15 undegraded samples. PSNR and SSIM (Huynh-Thu & Ghanbari, 2008; Wang et al., 2004), together with RMSE, are computed separately for the scalar spatial field s and the three-channel chromatic field $\mathbf{c}$. All metrics are evaluated only over valid field-support pixels. For $\mathbf{c}$, we additionally report ΔE_{00} (Sharma et al., 2005). For this diagnostic only, the three channels of $\mathbf{c}$ are interpreted as a linear-RGB triplet and converted to CIELAB using the sRGB primaries and the D65 reference white.

Table 7 reports the agreement between the predicted fields and their synthetic labels. The spatial field achieves a PSNR of 35.918 dB, an SSIM of 0.977, and an RMSE of 0.016. The chromatic field obtains a PSNR of 30.458 dB, an SSIM of 0.977, an RMSE of 0.030, and a mean ΔE_{00} of 1.568. Both fields have an SSIM of 0.977 when rounded to three decimal places. In this synthetic-label evaluation, the chromatic field has a larger RMSE and a lower PSNR than the spatial field. Qualitative examples are presented in Fig. 7. Across both indoor and outdoor scenes, the predicted

Table 7: PDFE field-estimation accuracy against synthetic label. ΔE_{00} is reported for the chromatic field only.

Physical component	PSNR (dB)	SSIM	RMSE	ΔE_{00}
Spatial field s	35.918	0.977	0.016	-
Chromatic field $\mathbf{c}$	30.458	0.977	0.030	1.568

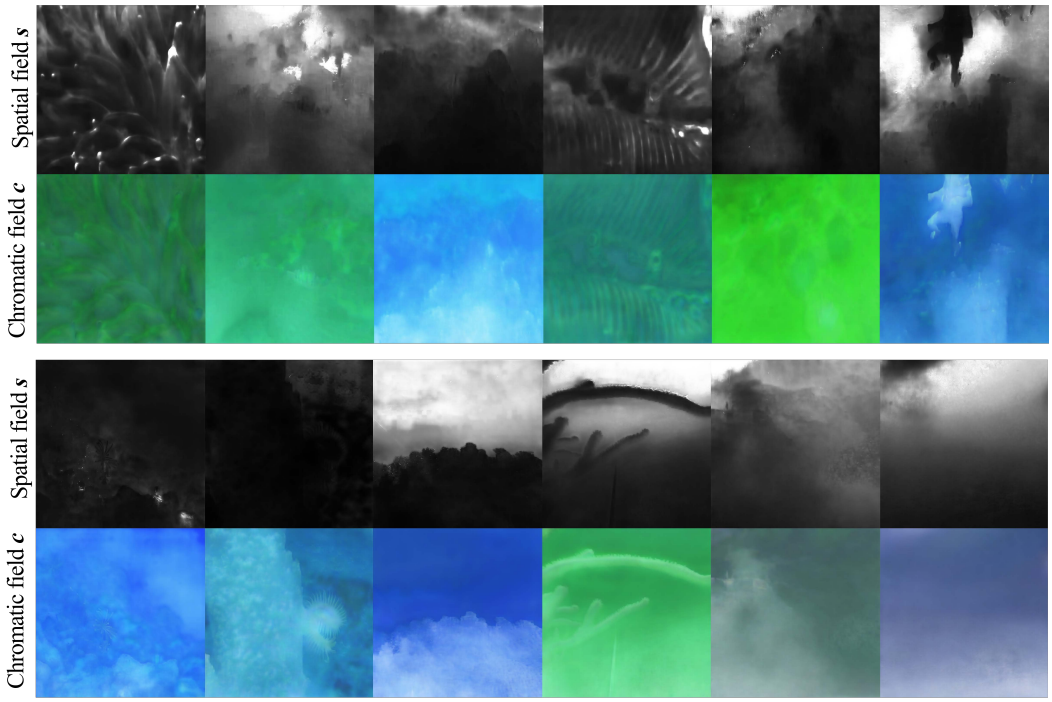

Figure 8: PDFE field predictions on real underwater images, shown for qualitative inspection because field ground truth is unavailable. The upper group contains examples from EUVP, LSUI, and UIEB; the lower group contains MIUD examples. In each group, the top row shows the predicted spatial field s_I in grayscale (brighter indicates greater predicted degradation), and the bottom row shows the normalized three-channel chromatic field $\mathbf{c}_I$.

fields follow the principal depth-dependent variation and the major object boundaries present in the synthetic labels. In particular, the estimates preserve the broad foreground–background ordering and the dominant local chromatic pattern. Residual differences remain around depth discontinuities, distant regions, and strongly illuminated areas.

Field Predictions on Underwater Images Fig. 8 visualizes the spatial and chromatic fields predicted for the real underwater images presented in the main paper. The predicted s_I fields exhibit clear spatial variations rather than assigning a uniform degradation level to the entire scene. Brighter regions indicate a larger degree of predicted structural degradation. Across different examples, s_I distinguishes foreground objects, distant backgrounds, and scattering-dominated regions while retaining the principal scene boundaries. These spatial variations allow s_I to guide the contrast initialization according to the local degradation state.

The predicted $\mathbf{c}_I$ fields also vary across scenes and spatial locations. Their blue-, cyan-, and green-dominant patterns are consistent with different channel-dependent attenuation conditions, while local magnitude variations suggest differences in relative attenuation strength. The foreground and surrounding water regions frequently receive different chromatic estimates, demonstrating that PDFE represents chromatic degradation as a spatially varying field rather than a global color vector. Moreover, s_I and $\mathbf{c}_I$ are related but not spatially identical, which is consistent with their complementary roles: s_I guides the spatial initialization, whereas $\mathbf{c}_I$ records spatially varying channel responses.

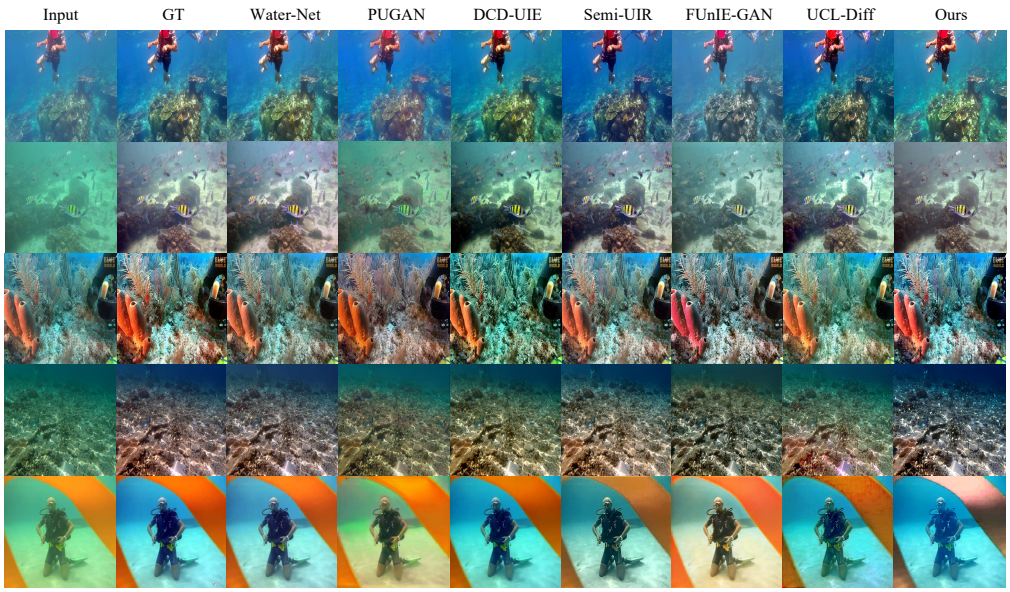

Figure 9: Qualitative UIEB comparison across training regimes. Each row shows one of five scenes. Columns, from left to right: input, provided reference (GT), Water-Net, PUGAN, DCD-UIE, Semi-UIR, FUnIE-GAN, UCL-Diff, and PD-GPO (Ours). The examples include varying color casts, visibility, and foreground–background colors.

During latent optimization, the corresponding candidate fields select support and supply support-aggregated channel ratios, respectively.

C Additional Comparisons on UIEB and MIUD

UIEB comparisons across supervision regimes. We compare PD-GPO with paired supervised, unpaired, and semi-supervised methods on the same fixed subset of 100 UIEBD images (Li et al., 2019). PD-GPO optimizes each test image using a PDFE trained on synthetic degradations and a frozen generative prior, without training a shared enhancement mapping on real underwater images. The compared methods are Water-Net (Li et al., 2019), PUGAN (Cong et al., 2023), DCD-UIE (Fan et al., 2026), FUnIE-GAN (Islam et al., 2020), UCL-Diff (Li et al., 2025), and Semi-UIR (Huang et al., 2023). their training regimes are indicated in Tables 8 and 9.

Figure 9 shows five scenes with different color casts and foreground–background combinations. PD-GPO generally improves foreground visibility while retaining the distinction between warm objects and cooler water, although its colors do not always match the provided references. This behavior is consistent with spatial-field-guided contrast initialization and region-based chromatic regularization.

Tables 8 and 9 show complementary outcomes: PD-GPO leads all four no-reference metrics (Yang & Sowmya, 2015; Panetta et al., 2016; Lin et al., 2025; Ke et al., 2021), whereas methods trained with real underwater data generally achieve better full-reference scores. The full-reference measures are PSNR (Huynh-Thu & Ghanbari, 2008), SSIM (Wang et al., 2004), and LPIPS (Zhang et al., 2018); they measure agreement with the provided references, which PD-GPO is not trained to reproduce. The remaining gap in reference color and fine-structure fidelity is a limitation.

MIUD comparison with zero-shot methods. Figure 10 compares five MIUD scenes with GDP (Fei et al., 2023), PhyDiff (Qiao et al., 2025), UIEAnything (Shao et al., 2025), Zero-UMSIE (Liu et al., 2024), and Osmosis (Bar Nathan et al., 2024), illustrating differences in brightness, local contrast, and color correction. Because MIUD has no paired clean references, the examples contextualize the no-reference scores in Table 3 rather than establish ground-truth fidelity.

Table 8: No-reference image-quality scores. Higher is better for all metrics. Methods are grouped by paired supervised, unpaired, semi-supervised, and zero-shot training regimes.

Type	Method	UCIQE ↑	UIQM ↑	PAUQA ↑	MUSIQ ↑
Supervised	Water-Net	0.303	5.232	0.508	57.179
	PUGAN	0.351	5.864	0.552	62.442
	DCD-UIE	0.356	5.070	0.602	61.943
Unsupervised	FUnIE-GAN	0.327	5.820	0.549	52.279
	UCL-Diff	0.328	5.420	0.629	58.372
Semi-Supervised	Semi-UIR	0.331	5.670	0.646	61.068
Zero-shot	Ours	**0.412**	**7.436**	**0.742**	**62.533**

Table 9: Full-reference image-quality scores. Higher PSNR and SSIM and lower LPIPS are better. Methods are grouped by training regime.

Type	Method	PSNR ↑	SSIM ↑	LPIPS ↓
Supervised	Water-Net	25.725	0.913	0.090
	PUGAN	21.016	0.877	0.134
	DCD-UIE	25.172	0.926	**0.070**
Unsupervised	FUnIE-GAN	21.184	0.875	0.150
	UCL-Diff	24.450	0.927	0.092
Semi-Supervised	Semi-UIR	**26.019**	**0.940**	0.091
Zero-shot	Ours	21.544	0.820	0.164

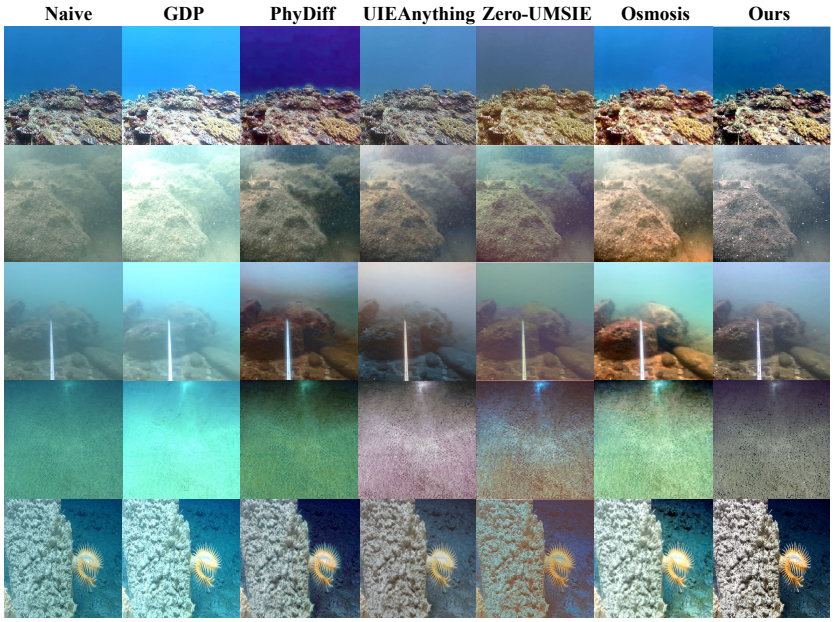

Figure 10: Qualitative comparison on five MIUD scenes without paired clean references. Columns, from left to right: unprocessed input (Naive), GDP, PhyDiff, UIEAnything, Zero-UMSIE, Osmosis (Bar Nathan et al., 2024), and PD-GPO (Ours). Each row shows the same scene across methods.

D Downstream Tasks

Downstream instance segmentation. To assess the practical value of PD-GPO beyond visual enhancement, we examine its effect on underwater instance segmentation. An effective restoration method should provide informative inputs for subsequent perception while improving visual appearance. We apply the same segmentation model to the original image and the outputs of the compared restoration methods, and qualitatively compare the resulting predictions.

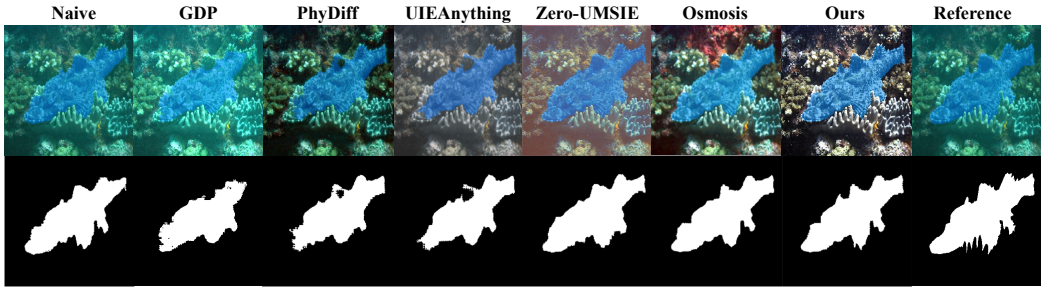

Figure 11: Qualitative underwater instance-segmentation comparison using one fixed segmentation model. Columns show the unprocessed input (Naive), six restored images (GDP, PhyDiff, UIEAnything, Zero-UMSIE, Osmosis, and PD-GPO), and the provided reference. The top row overlays predicted masks in blue on each input or restored image; the bottom row shows binary predicted masks, with white denoting the target. The reference column shows the provided reference image above its manual target mask.

As shown in Fig. 11, several competing predictions exhibit irregular boundaries or overly smoothed contour segments, indicating that changes in image appearance do not necessarily translate into more accurate object delineation. The mask obtained from the PD-GPO result captures a coherent target region while retaining distinct local contour variations. Its overall shape broadly follows the reference mask, although the finer pointed structures along the lower boundary remain incompletely resolved. This qualitative example suggests that the PD-GPO output provides useful structural cues for target delineation, supporting its potential as a pre-processing step for downstream underwater perception.

E Hyperparameter Comparison

We examine the FLUX sampling-step setting s and the objective weights λ_{feat} and λ_{chrom} on UIEB. The two comparisons share the configuration $(step, \lambda_{\text{feat}}, \lambda_{\text{chrom}}) = (1, 1, 1)$: Table 10 varies the sampling steps with both weights fixed, whereas Table 11 varies the weights with $step = 1$.

Sampling-step selection. With both weights fixed at 1, $step = 1$ achieves the highest MUSIQ (62.533), PSNR (21.544), and SSIM (0.820), and the lowest LPIPS (0.164). Increasing to $step = 2$ gives the best UIQM (7.766) and PAUQA (0.765), while $step = 4$ gives the best UCIQE (0.435). Thus, additional sampling steps do not improve the seven metrics uniformly. The lower full-reference scores at larger $step$ are consistent with the possibility of reference-inconsistent changes introduced during generative sampling. Since $step = 1$ leads on four metrics and requires the fewest sampling evaluations, we select it as the final quality–efficiency setting.

Loss-weight comparison. With $step = 1$ fixed, all tested configurations retain both loss terms, this comparison changes their weights rather than removing either objective. The chromatic-emphasized setting $(\lambda_{\text{feat}}, \lambda_{\text{chrom}}) = (0.5, 2)$ achieves the highest MUSIQ (64.789), but has the lowest PSNR and SSIM and the highest LPIPS. The feature-emphasized setting $(2, 0.5)$ achieves the highest SSIM (0.834). Equal weights $(1, 1)$ lead on UCIQE, UIQM, PAUQA, PSNR, and LPIPS—five of the seven metrics. We therefore use $(1, 1)$ in the final configuration.

F Failure Cases

Although our method generally produces stable color correction and contrast enhancement, it remains challenging under extreme or ambiguous underwater degradation. Figure 12 presents five representative failure cases by comparing the input, our result, and the provided reference.

In the leftmost example, PD-GPO darkens the foreground and gives it a warmer tone than the provided reference. In the second, a strong green cast leads to excessive red compensation on the foreground object. The third shows local reddish artifacts around a dark foreground boundary, while

Table 10: Quantitative comparison of different step configurations on UIEB with $\lambda_{\text{feat}} = \lambda_{\text{chrom}} = 1$. The best result in each column is highlighted in bold.

step	No-reference				Reference-based		
	UCIQE↑	UIQM↑	PAUQA↑	MUSIQ↑	PSNR↑	SSIM↑	LPIPS↓
1	0.412	7.436	0.742	**62.533**	**21.544**	**0.820**	**0.164**
2	0.384	**7.766**	**0.765**	62.132	21.387	0.810	0.171
4	**0.435**	7.341	0.733	61.911	21.294	0.812	0.173

Table 11: Quantitative comparison of different loss-weight configurations on UIEB with $step = 1$.

λ_{feat}	λ_{chrom}	No-reference				Reference-based		
		UCIQE↑	UIQM↑	PAUQA↑	MUSIQ↑	PSNR↑	SSIM↑	LPIPS↓
0.5	2	0.373	7.061	0.738	**64.789**	18.469	0.723	0.303
2	0.5	0.362	6.365	0.602	63.038	21.404	**0.834**	0.167
1	1	**0.412**	**7.436**	**0.742**	62.533	**21.544**	0.820	**0.164**

the restored water color differs from the cooler reference. The fourth begins with an already warm scene; unnecessary compensation shifts its colors toward saturated cyan and purple. In the rightmost example, severe scattering leaves few reliable structural cues. PD-GPO increases foreground contrast but produces dark reddish tones, while distant structures remain poorly resolved.

These failures may arise from uncertainty in estimating the physical degradation fields from a single real underwater observation. The physical estimator is trained using synthetic degradations; consequently, unusual water types, extreme color distributions, or degradation patterns outside the synthetic training distribution may lead to inaccurate chromatic and spatial fields. Because these fields guide latent optimization, estimation bias may be amplified when chromatic regularization repeatedly acts on regions with unreliable field predictions. Moreover, under severe scattering, much of the scene information is physically missing rather than merely attenuated. In such cases, neither the content-preservation constraint nor the frozen generative prior can reliably recover the lost structures.

Future work could improve robustness by incorporating confidence-aware physical-field estimation, adaptively weighting the physical constraints according to prediction reliability, and expanding the synthetic degradation distribution to cover more diverse real-world water conditions. For severely degraded scenes, additional observations or complementary depth and temporal cues may also be required to recover reliable scene content.

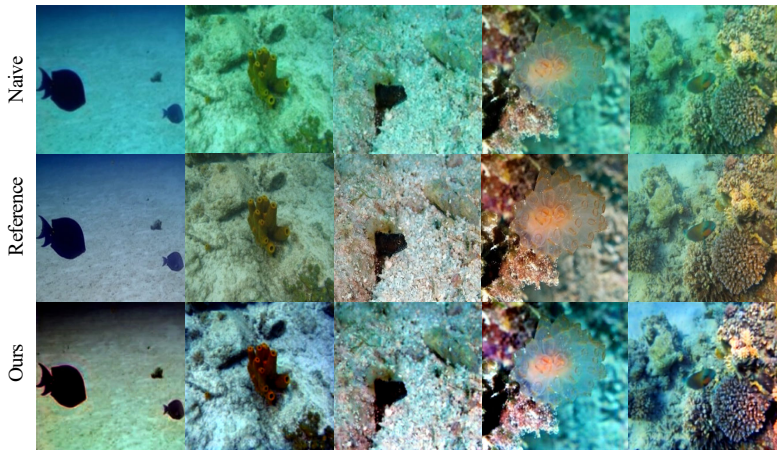

Figure 12: Five representative failure cases (columns). Rows show the unprocessed input (Naive), provided reference, and PD-GPO output (Ours), from top to bottom. The first three examples show local color shifts or overcompensation; the fourth shows unnecessary correction of an already warm scene; the fifth shows limited recovery under severe scattering.

G Discussions and Limitations

The spatial and chromatic fields are restoration-oriented summaries of degradation rather than uniquely identifiable optical parameters. A single underwater RGB observation does not uniquely determine A_{ch}, α_{ch}, β, or metric depth. PDFE instead estimates fields that guide the operations in PD-GPO. Table 7 and Fig. 7 measure agreement with labels generated by the synthetic degradation model, whereas Fig. 8 shows qualitative predictions on real underwater images without field ground truth. The overcorrection and incomplete recovery in Fig. 12 are consistent with this risk. Accuracy against synthetic labels therefore does not establish the physical accuracy of the predicted fields under all real water conditions.

The two fields have distinct roles in restoration: s_I modulates the contrast initialization, while $s_J^{(k)}$ selects the support over which the channel predictions in $\mathbf{c}_J^{(k)}$ are aggregated. The chromatic loss in Eq. (6) regularizes channel ratios on that support; it does not enforce a unique pixelwise inversion of attenuation or exact agreement with a reference color. Hard support selection does not transmit gradients through $s_J^{(k)}$, and an empty support makes the displayed chromatic loss constant. The component differences in Table 4 are point estimates for the evaluated subset; they do not isolate the effect of re-estimating the fields at every iteration.

The reported metrics also have different interpretations. Full-reference scores measure agreement with the provided reference appearance, whereas the non-reference scores and visual examples on the 100-image MIUD evaluation subset cannot establish fidelity to an unavailable clean image or performance over the full 850-image collection. Finally, image-specific latent optimization requires substantial computation: Table 5 reports 181.102 s per 512×512 image on an RTX 3090 under the evaluated setting. PD-GPO is therefore more suitable for offline, quality-oriented restoration than real-time deployment, and more efficient optimization are useful directions for future work.