AdaMSCoL: Adaptive Multi-Scale Structural Consistency for Unsupervised Underwater Image Enhancement

Aparna Tiwari
NTOU, Taiwan
21353003@mail.ntou.edu.tw

Hitika Tiwari
IIT Madras, Zanzibar Campus
hitika@iitmz.ac.in

Dong-Lin Li
NTOU, Taiwan
ericli@email.ntou.edu.tw

Abstract

*Underwater image enhancement is an ill-posed inverse problem due to wavelength-dependent absorption, scattering, and spatially varying illumination. The problem further becomes complicated by the lack of paired supervision and the ambiguity of visually plausible restorations. Existing unsupervised methods often rely on heuristic priors or pixel-wise constraints that fail to preserve multi-scale structural consistency. In this work, we propose **Ada**ptive **M**ulti-Scale **S**tructural **Co**nsistency **L**oss (**AdaMSCoL**), a novel unsupervised training objective that enforces structural fidelity across spatial scales. AdaMSCoL integrates cross-scale patch correlation, self-reconstruction consistency, and adaptive prior regularization to jointly preserve local structures, global intensity distributions, and perceptual naturalness. Our experiments on multiple underwater benchmarks demonstrate that AdaMSCoL significantly improves structural preservation, color realism, and perceptual quality compared to existing unsupervised objectives, emphasizing it as a strong generic loss for underwater image enhancement tasks.*

1. Introduction

Underwater image enhancement (UIE) plays a crucial role in enabling reliable underwater vision for applications such as marine exploration, autonomous underwater robotics, and biological monitoring [20, 22, 31]. Unlike in-air imaging, underwater imaging is inherently challenging due to wavelength-dependent absorption and scattering, which cause severe color distortion, low contrast, haze, and the loss of fine structural details [11, 39, 57]. These degradations not only reduce perceptual quality but also significantly impair the performance of downstream vision tasks, including object detection, semantic segmentation, and feature tracking [21, 48]. Consequently, robust UIE methods are essential to ensure both visually appealing reconstructions and reliable task performance in diverse underwater conditions.

Classical UIE approaches typically rely on physical models of light propagation or hand-crafted priors to invert the degradation process [11, 14, 19]. These methods estimate parameters such as attenuation coefficients, backscatter, and transmission maps, or apply heuristic enhancement rules derived from domain knowledge. While such methods can yield satisfactory results in controlled or laboratory environments, they struggle to generalize to real-world underwater scenes where water types, depth, turbidity, and illumination vary widely [16, 33]. The limited adaptability of prior-based methods often leads to residual color casts, over- or under-enhanced regions, and loss of structural details in complex scenes.

Recently, deep learning–based approaches have demonstrated improved robustness and flexibility by learning enhancement mappings directly from data [8, 9, 17, 30, 49]. Despite their success, most existing methods adopt deterministic training with pixel-wise losses or single-scale structural metrics such as SSIM [53]. While these objectives encourage local or global reconstruction fidelity, they fail to adequately capture the multi-scale and spatially varying nature of underwater degradations. As a result, enhanced images often suffer from over-smoothing, texture distortion, or unnatural color artifacts, particularly in regions with severe scattering or complex illumination patterns [22, 58]. The abovementioned limitation emphasizes the need for objectives that explicitly model structural consistency across multiple spatial scales while preserving perceptual realism.

To address these challenges, we propose **Ada**ptive **M**ulti-**S**cale Structural **Co**nsistency **L**oss (**AdaMSCoL**), a novel unsupervised training objective which enforces structural fidelity at both local and global levels. AdaMSCoL integrates three complementary mechanisms. First, cross-scale patch correlation aligns local structures across multiple resolutions, ensuring that fine-grained edges and textures are preserved consistently. Second, self-reconstruction consistency maintains global intensity distributions and spatial layouts, preventing unnatural distortions and ensur-

ing photometric stability. Third, adaptive prior regularization encourages perceptually natural color distributions and smooth transitions, while preserving important edges. Upon jointly modeling structural information across scales and combining it with photometric and prior constraints, AdaMSCoL mitigates the intrinsic ambiguity of underwater image enhancement and produces visually plausible, structurally coherent outputs. Our AdaMSCoL is extensively evaluated on multiple challenging underwater benchmarks, for instance, EUVP [22], LSUI [37] and UIEBD [31]. Experimental results demonstrate that our method consistently outperforms prior unsupervised objectives, producing both higher perceptual quality and better structural fidelity. For instance, on the EUVP dataset, AdaMSCoL improves SSIM from 0.771 to 0.838 (+0.067) and reduces mean absolute deviation (MAD) from 94.62 to 70.22 (-24.40), showing substantial improvements over state-of-the-art approaches.

Our contributions are as follows:

- We introduce **AdaMSCoL**, a novel unsupervised loss function that enforces adaptive multi-scale structural consistency to address the ill-posed and ambiguous nature of underwater image enhancement.
- AdaMSCoL integrates cross-scale patch correlation, self-reconstruction consistency, and adaptive prior regularization, enabling the network to preserve both local textures and global intensity distributions.
- Our AdaMSCoL shows substantial improvements on various datasets, e.g., SSIM +0.067, MAD -24.40, and UIQM +1.279, emphasizing the efficacy of our method.

2. Related Work

Underwater image enhancement is significantly important in marine exploration, robotics, and vision-based underwater applications. Existing methods [2, 3, 26, 44, 50] are broadly categorized into prior-based, model-free, deep learning-based, and probabilistic-based approaches. Each category leverages different assumptions and strategies, showing distinct trade-offs between interpretability, performance, and robustness.

2.1. Prior-Based Methods

Prior-based methods exploit physical models of underwater image formation and handcrafted priors to estimate degradation parameters such as light attenuation, scattering, and backscatter, and subsequently reconstruct enhanced images. Common priors include the red channel prior [19], underwater dark channel prior [11], and light attenuation prior [48], which provide visual cues for estimating scene transmission and illumination.

Several works estimate depth or transmission maps using absorption and blurriness cues to restore clean images through underwater imaging models [39]. Extensions of the dark channel prior have been proposed for extreme imaging conditions [38]. More sophisticated models differentiate between direct transmission and backscatter coefficients, enabling accurate restoration, especially when RGBD information is available [1].

Although physically grounded, these approaches often struggle in complex underwater environments where spatially varying water properties, illumination changes, and unknown scene geometry violate model assumptions.

2.2. Model-Free Methods

Model-free methods enhance underwater images without explicitly modeling physical degradation, focusing instead on pixel-level manipulation and feature enhancement. Classical techniques include Contrast-Limited Adaptive Histogram Equalization (CLAHE) [18], White Balance (WB) [13], and Retinex-based methods [27, 41], which redistribute pixel intensities and correct color distortions to improve contrast and visibility [11].

Fusion-based, multi-scale, and optimization-driven methods have further advanced model-free underwater image enhancement, including adaptive color correction, contrast fusion frameworks, wavelet-based fusion, fuzzy set-based approaches, and frequency–spatial domain fusion [4, 6, 23, 29, 40, 52, 59]. These techniques selectively enhance high-frequency details, sharpen edges, and improve both color fidelity and perceptual quality. By combining spatial and frequency cues, they enhance texture, contrast, and saturation without relying on physical priors.

Despite their simplicity and computational efficiency, model-free methods remain challenged in complex underwater environments, where non-uniform illumination, scattering, and spatially varying turbidity limit robustness and generalization.

2.3. Deep Learning-Based Methods

Deep learning-based methods have demonstrated superior performance by automatically learning complex representations for underwater image enhancement. Key approaches include Convolutional Neural Networks (CNNs), transformer-based networks, and Generative Adversarial Networks (GANs), which model nonlinear underwater degradations effectively [60]. Encoder–decoder CNNs are widely adopted for denoising and color restoration, while lightweight CNNs leverage scene-specific priors for improved performance [45].

Attention mechanisms, multi-scale fusion, and transformer-based diffusion models have further improved structural preservation and perceptual fidelity [12, 60]. Adaptive attention networks refine image details while maintaining global consistency [35], and contrastive adversarial learning approaches optimize perceptual quality alongside downstream task performance [34]. GAN-based methods generate synthetic underwater data and support unsupervised

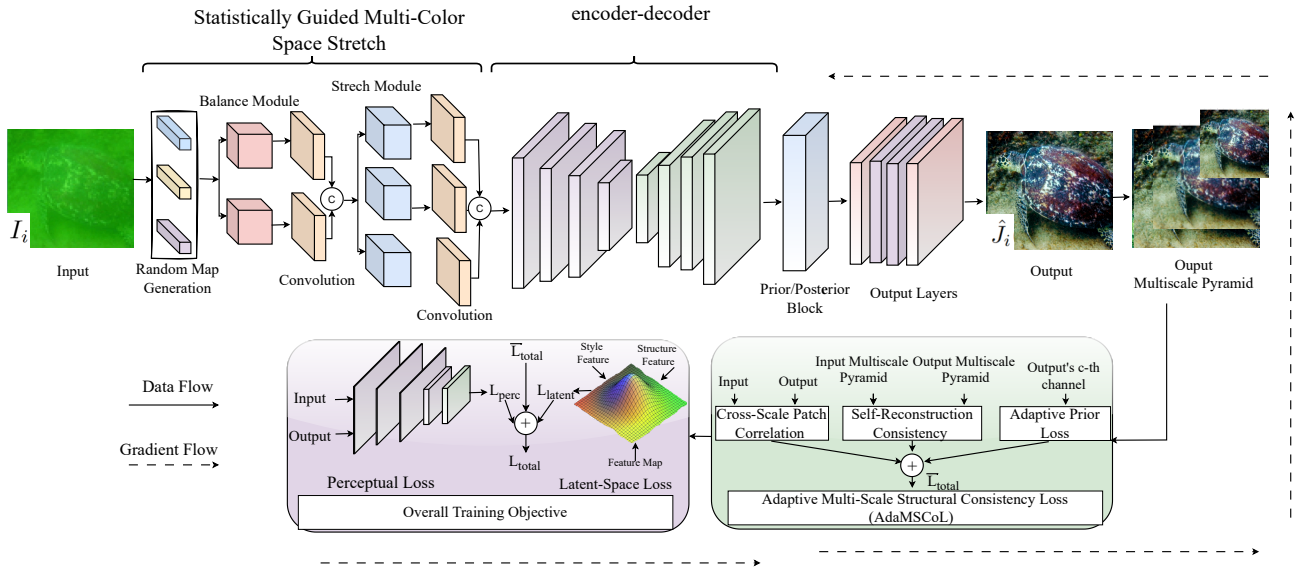


Figure 1. Illustration of our proposed end-to-end training pipeline incorporating novel **Ada**ptive **M**ulti-Scale **S**tructural **Co**nsistency **L**oss (**AdaMSCoL**) training objective for underwater image enhancement task.

or semi-supervised enhancement, with cascaded attention modules and dual-discriminator architectures improving noise suppression and long-range spatial modeling [10, 51]. Hybrid approaches integrate physical priors with deep networks. For instance, degradation information explicitly extracted and leveraged to guide restoration [49]. These strategies demonstrate the potential of combining data-driven learning with physics-informed guidance to improve robustness, generalization, and fidelity.

2.4. Probabilistic-Based Methods

Probabilistic-based approaches explicitly model uncertainty to handle noise, disturbances, and modeling errors in underwater imaging. Unlike deterministic approaches, they represent latent variables as probability distributions, capturing diverse and ambiguous degradations. Variational autoencoders (VAEs) and conditional VAEs (cVAEs) map input images into latent probabilistic spaces and reconstruct images using reconstruction and regularization objectives [28, 47].

Such frameworks have been effective in background modeling, motion generation, and image denoising [5, 32, 55]. In underwater enhancement, uncertainty-aware frameworks predict pixel-level uncertainty maps to improve robustness under challenging conditions [15]. Advanced strategies such as distributional uncertainty modeling and probabilistic adaptive normalization improve generalization across environments by dynamically aligning latent feature distributions, mitigating domain shifts, and capturing the inherent ambiguity of underwater image formation.

3. Methodology

We formulate underwater image enhancement as a conditional probabilistic generation problem to address the lack of paired supervision and the ambiguity of visually plausible restorations. Inspired by [42], given a degraded image $I \in \mathbb{R}^{H\times W\times 3}$, we model a distribution over plausible enhanced outputs J:

$$p(J \mid I) = \int p(J \mid I, z)\, p(z \mid I)\, dz, \tag{1}$$

where the latent variable z captures enhancement uncertainty and the inherent ambiguity of underwater restoration. In the absence of paired supervision, we construct pseudo-reference images R on-the-fly using a stochastic augmentation pipeline [43]. Specifically, input images are transformed via randomized contrast stretching, gamma correction, and saturation adjustments to generate diverse candidates, which are further refined using statistically guided color correction (Gray World and White Patch) and normalization in RGB, HSI, and Lab color spaces. This provides weak supervisory signals that encourage plausible color and intensity distributions. Empirically, this strategy stabilizes optimization and improves convergence. Our backbone then adopts a conditional variational framework with a dual-branch U-Net encoder–decoder that learns a posterior $q_\phi(z \mid I, R)$ and a prior $p_\theta(z \mid I)$. The decoder f_θ synthesizes enhanced images conditioned on latent samples:

$$\hat{J} = f_\theta(I, z). \tag{2}$$

To model diverse enhancement styles, uncertainty is injected via probabilistic adaptive instance normalization,

where channel-wise statistics $(\mu, \sigma) \sim p_\theta(\mu, \sigma \mid I)$ modulate intermediate features. At inference time, multiple latent samples $\{z_k\}_{k=1}^K$ yield candidate outputs $\{\hat{J}_k\}_{k=1}^K$, from which the final result is selected via likelihood estimation or learned ranking. The proposed training objective, which enforces structural and perceptual consistency while regularizing the latent space, is described next.

3.1. Adaptive Multi-Scale Structural Consistency Loss

Underwater images suffer from complex, spatially varying degradations, where simple pixel-wise losses fail to capture local structures. To address this, we propose an ***Ada**ptive **M**ulti-Scale **S**tructural **Co**nsistency **L**oss* (***AdaMSCoL***) for our unsupervised training objective.

3.1.1. Cross-Scale Patch Correlation Loss (CSPC)

Our CSPC loss preserves local structures and global edges by explicitly measuring patch-wise correlations across multiple resolutions. Let $\mathcal{I} = \{I_i\}_{i=1}^N$ denote the set of observed underwater images and $\hat{\mathcal{J}} = \{\hat{J}_i\}_{i=1}^N$ their restored counterparts. For each image I_i at scale k, we extract overlapping patches $\mathcal{P}_{i,x}^k \subset \Omega$ centered at pixel x and compute local patch statistics. For a patch $\mathcal{P}_{i,x}^k$, let $\mathbb{E}_{\mathcal{P}}[\cdot]$ denote patch-wise averaging. We define the patch-wise means

$$\mu_{i,x}^k = \mathbb{E}_{\mathcal{P}}[I_i], \quad \hat{\mu}_{i,x}^k = \mathbb{E}_{\mathcal{P}}\left[\hat{J}_i\right], \tag{3}$$

the corresponding variances

$$\sigma_{i,x}^{k\,2} = \mathbb{E}_{\mathcal{P}}\left[(I_i - \mu_{i,x}^k)^2\right], \quad \hat{\sigma}_{i,x}^{k\,2} = \mathbb{E}_{\mathcal{P}}\left[(\hat{J}_i - \hat{\mu}_{i,x}^k)^2\right], \tag{4}$$

and the cross-covariance

$$\sigma_{i\hat{i},x}^k = \mathbb{E}_{\mathcal{P}}\left[(I_i - \mu_{i,x}^k)(\hat{J}_i - \hat{\mu}_{i,x}^k)\right]. \tag{5}$$

Each patch is then normalized (using Eqs. (3), (4) and (5)) to zero-mean, unit-variance vectors:

$$\tilde{\mathbf{v}}_{i,x}^k = \frac{I_i(\mathcal{P}_{i,x}^k) - \mu_{i,x}^k}{\sigma_{i,x}^k + \epsilon_2}, \quad \hat{\tilde{\mathbf{v}}}_{i,x}^k = \frac{\hat{J}_i(\mathcal{P}_{i,x}^k) - \hat{\mu}_{i,x}^k}{\hat{\sigma}_{i,x}^k + \epsilon_2}, \tag{6}$$

where ϵ_2 is a small constant for numerical stability. The patch correlation score combines (using Eq. (6)) contrast-normalized similarity with luminance alignment:

$$S_i^k(x) = (\tilde{\mathbf{v}}_{i,x}^k \cdot \hat{\tilde{\mathbf{v}}}_{i,x}^k) \cdot \frac{2\mu_{i,x}^k \hat{\mu}_{i,x}^k + \epsilon_1}{(\mu_{i,x}^k)^2 + (\hat{\mu}_{i,x}^k)^2 + \epsilon_1}, \tag{7}$$

where ϵ_1 prevents division by zero. Our score (Eq. (7)) ensures that local patterns are aligned in both contrast and luminance. To enforce cross-scale consistency, we penalize misalignment between patches at adjacent resolutions:

$$L_{\text{cross}} = \sum_{i=1}^{N} \sum_{k=1}^{K-1} \frac{1}{|\Omega|} \sum_{x \in \Omega} \left\| \hat{J}_i(\mathcal{P}_{i,x}^k) - \text{Ups}(\hat{J}_i(\mathcal{P}_{i,x}^{k+1})) \right\|_2^2. \tag{8}$$

Here, Ups $(\cdot)$ denotes upsampling. Finally, the CSPC loss aggregates the patch correlations (from Eq. (7)) and cross-scale regularization (from Eq. (8)):

$$L_{\text{CSPC}} = \sum_{i=1}^{N} \sum_{k=1}^{K} \alpha_k \frac{1}{|\Omega|} \sum_{x \in \Omega} \left(1 - S_i^k(x)\right) + \lambda L_{\text{cross}}, \tag{9}$$

where α_k balances contributions from each scale and λ controls cross-scale regularization.

3.1.2. Self-Reconstruction Consistency Loss (SRC)

To prevent degenerate enhancements and preserve intensity fidelity, we enforce a multi-scale self-reconstruction constraint. Let $\mathcal{D}_k$ denote a downsampling operator on image height H and width W to scale k:

$$\mathcal{D}_k : \mathbb{R}^{H \times W \times 3} \to \mathbb{R}^{H_k \times W_k \times 3},$$
$$H_k = H/2^{k-1},\ W_k = W/2^{k-1}.$$

The SRC loss is defined as:

$$L_{\text{SRC}} = \sum_{k=1}^{K} \gamma_k \left\| \mathcal{D}_k(\hat{J}_i) - \mathcal{D}_k(I_i) \right\|_1, \tag{10}$$

where $\gamma_k > 0$ is a scale-specific weight and $\|\cdot\|_1$ denotes the element-wise ℓ_1 norm. This enforces that enhanced outputs preserve the overall intensity structure of the input.

3.1.3. Adaptive Prior Loss (APR)

To regularize the enhanced images, we combine color constancy (CE) and spatial variation (SV) losses. Let $\hat{J}_i^c$ denote the c-th channel of $\hat{J}_i$, and let

$$\bar{J}_i^c = \frac{1}{HW} \sum_{x \in \Omega} \hat{J}_i^c(x) \tag{11}$$

be its spatial mean. The color constancy loss (using Eq. (11)):

$$L_{\text{CE}} = \sum_{c < c'} (\bar{J}_i^c - \bar{J}_i^{c'})^2, \tag{12}$$

which encourages globally balanced channel intensities. The spatial variation loss (using Eq. (11)) is:

$$L_{\text{SV}} = \frac{1}{|\Omega|} \sum_{x \in \Omega} \|\nabla \hat{J}_i(x)\|_1, \tag{13}$$

where $\nabla \hat{J}_i(x) \in \mathbb{R}^{3 \times 2}$ contains horizontal and vertical gradients at pixel x, encouraging local smoothness while preserving edges. The combined adaptive prior loss (using Eq. (12) and Eq. (13)) is then:

$$L_{\text{APR}} = L_{\text{CE}} + \eta\, L_{\text{SV}}, \tag{14}$$

where $\eta > 0$ balances spatial smoothness.

3.1.4. Proposed AdaMSCoL

Our proposed AdaMSCoL is an unsupervised training objective that integrates all three components: cross scale patch correlation (Eq. (9)), reconstruction loss (Eq. (10)), and adaptive prior loss (Eq. (14)) as follows:

$$\overline{L}_{\text{total}} = \Psi_{\text{struct}} \cdot L_{\text{CSPC}} + \Psi_{\text{recon}} \cdot L_{\text{SRC}} + \Psi_{\text{prior}} \cdot L_{\text{APR}}, \tag{15}$$

where $\Psi_{\text{struct}}, \Psi_{\text{recon}}, \Psi_{\text{prior}}$ (set to 0.5, 0.3, and 0.2 in our experiments) are fixed weighting coefficients that balance structural, reconstruction, and prior constraints. Although the weights are fixed, the formulation is adaptive in that it integrates complementary constraints operating across multiple spatial scales and statistical levels. Eq. (15) ensures that the network preserves local and global structures, maintains intensity fidelity, and produces visually natural enhancements. We note that the term "adaptive" refers to the multi-component and multi-scale nature of the loss rather than dynamic weight learning.

3.2. Deep Perceptual Feature Consistency Loss

While low-level pixel and structural losses enforce local fidelity, they fail to capture high-level semantic and perceptual attributes of natural images. To encourage perceptually meaningful reconstructions, we impose a deep feature consistency constraint using a pretrained VGG16 network. Although the VGG16 network is pretrained on ImageNet, prior work [43] has shown that its intermediate features generalize well to low-level vision tasks, providing effective perceptual guidance even in domain-shifted scenarios such as underwater imaging. Let $\phi_l(\cdot)$ denote the activation map extracted from the l-th layer of a fixed model pretrained on ImageNet. Given an observed underwater image I_i and its enhanced counterpart $\hat{J}_i$, we define the perceptual discrepancy as:

$$L_{\text{perc}}(I_i, \hat{J}_i) = \frac{1}{CHW} \left\| \phi_l(I_i) - \phi_l(\hat{J}_i) \right\|_2^2, \tag{16}$$

where C, H, and W denote the channel and spatial dimensions of the feature map. Notably, similar perceptual losses have been widely adopted in prior work [43].

3.3. Latent-Space Normalization Loss

To stabilize unsupervised training and enforce structured latent representations, we regularize the latent distributions of features corresponding to structure (u) and style (s) against standard Gaussian priors:

$$L_{\text{latent}} = D_{\text{KL}}\big(q(u|F_i) \,\|\, p(u)\big) + D_{\text{KL}}\big(q(s|F_i) \,\|\, p(s)\big), \tag{17}$$

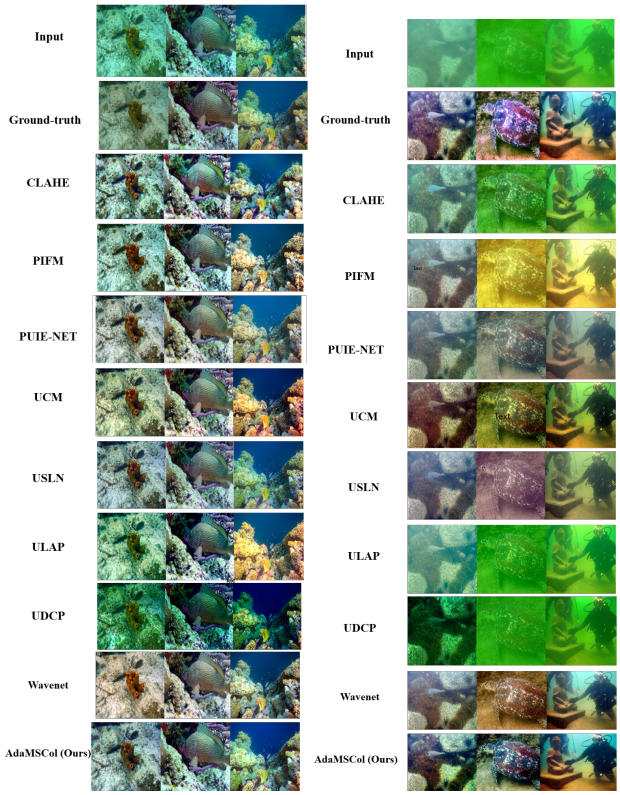
where $q(\cdot|F_i)$ is the posterior estimated from feature map F_i, and $p(\cdot)$ is the standard Gaussian prior. D_{KL} denotes KL divergence. This loss prevents unstable training and mode collapse.

3.4. Total Loss

Our complete training loss for unsupervised underwater image enhancement consists of: (i) the proposed AdaMSCoL (Eq. (15)), (ii) deep feature consistency (Eq. (16)), and (iii) latent space normalization (Eq. (17)) components as follows:

$$L_{\text{total}} = \lambda_1 \overline{L}_{\text{total}} + \lambda_2 L_{\text{perc}} + \lambda_3 L_{\text{latent}}, \tag{18}$$

with weights $\lambda_1 = 1.0$, $\lambda_2 = 0.1$, and $\lambda_3 = 0.01$. The aforementioned formulation achieves a trade-off between structural fidelity, pixel-level consistency, perceptual quality, color realism, smoothness, and latent stability for challenging underwater image enhancement tasks.



(a) EUVP Dataset **(b)** UIEBD Dataset

Figure 2. Qualitative comparison on underwater image enhancement benchmarks across various approaches. Apart from AdaMSCoL (Ours), input, and ground truth, other images are obtained from [43].

4. Experimental Details

In this section, we describe the datasets used for training and evaluation in Sec. 4.1, followed by the evaluation metrics in Sec. 4.2. We then present quantitative and qualitative comparison results with state-of-the-art methods in Sec. 4.4. Finally, detailed ablation studies analyzing the contribution of individual components and hyperparameters

Table 1. Quantitative comparison on various datasets (EUVP [22], LSUI [37], UIEBD [31]). Reference-based metrics: PSNR↑, SSIM↑, MAD↓, GMSD↓. Non-reference metrics: UCIQE↑, UIQM↑. – indicates that results are unavailable. Metrics differ across methods, as they are extracted and merged from different works [43, 46].

Method	EUVP [22]				LSUI [37]				UIEBD [31]			
	PSNR ↑	SSIM ↑	MAD ↓	GMSD ↓	PSNR ↑	SSIM ↑	UCIQE ↑	UIQM ↑	PSNR ↑	SSIM ↑	UCIQE ↑	UIQM ↑
CLAHE [18]	18.97	0.726	138.6	0.090	–	–	–	–	20.64	0.821	–	–
IBLA [39]	22.62	0.719	97.78	0.068	–	–	–	–	17.56	0.614	–	–
PIFM [14]	20.17	0.747	113.5	0.071	–	–	–	–	23.62	0.852	–	–
PUIE-Net [24]	21.01	0.770	94.55	0.052	20.38	0.831	0.542	3.134	23.74	0.844	0.546	3.359
RGHS [19]	21.13	0.753	98.40	0.056	–	–	–	–	23.57	0.803	–	–
UCM [48]	20.91	0.767	99.37	0.062	–	–	–	–	22.03	0.815	–	–
UDCP [11]	15.80	0.572	136.8	0.098	13.83	0.553	0.571	2.104	13.47	0.548	0.588	1.942
ULAP [57]	21.91	0.730	108.4	0.071	–	–	–	–	18.95	0.718	–	–
USLN [40]	20.87	0.771	94.62	0.050	–	–	–	–	24.04	0.849	–	–
Raw [22]	–	–	–	–	21.32	0.765	0.539	2.660	19.17	0.749	0.534	2.698
ACDC [56]	–	–	–	–	18.58	0.724	0.543	3.359	20.27	0.817	0.552	3.300
P2CNet [16]	–	–	–	–	12.24	0.446	0.534	1.774	12.79	0.457	0.525	1.927
TCTL-Net [25]	–	–	–	–	15.14	0.658	0.547	2.071	15.56	0.721	0.574	2.119
AdaMSCoL (Ours)	**21.98**	**0.838**	**70.22**	0.092	**21.64**	**0.846**	0.358	**3.865**	19.91	**0.859**	0.333	2.992

are provided in Sec. 4.5. We note that results for approaches are taken directly from their existing works [46**?**], where only a subset of metrics (e.g., PSNR, SSIM, or UIQM) are provided. To ensure fair and reproducible comparison, we restrict our evaluation to the metrics reported in prior work and avoid recomputing results under potentially different experimental settings. Consequently, certain metrics such as GMSD and UCIQE are not available for all methods. Despite this limitation, our method demonstrates consistent improvements across the commonly reported metrics, indicating robust performance gains.

4.1. Datasets

We evaluate AdaMSCoL on three publicly available underwater image datasets: EUVP [22], UIEBD [31], and LSUI [37]. These datasets include diverse water types, illumination conditions, and degradation levels. Only UIEBD [31] is used for unsupervised training, while all datasets are used for evaluation to assess generalization. Notably, we follow the same testing dataset partitioning as existing methods [42].

4.2. Evaluation Details

Quantitative performance is measured using full-reference metrics such as PSNR [53], SSIM [53], MAD [7], and GMSD [54], and no-reference perceptual metrics such as UIQM [36] and UCIQE [36]. Qualitative analysis evaluates structural fidelity, color realism, and perceptual naturalness.

4.3. Implementation Details

All models are implemented in PyTorch and trained on an NVIDIA RTX 4060 Ti. Inputs are processed as patches of size 256×256, with a batch size of 10 for 500 epochs using the Adam optimizer at a learning rate of 1×10^{-4} with standard momentum. Data augmentation includes random flips and rotations, and learning rates decay progressively for stable convergence. During inference, multiple latent samples are drawn to produce diverse enhancement candidates, with the final output selected via likelihood-based ranking.

4.4. Result Comparison with State-of-the-Arts

We compare the proposed AdaMSCoL against a comprehensive set of state-of-the-art underwater image enhancement methods, including traditional image processing approaches, physical model–based methods, and recent learning-based techniques. Specifically, we consider representative baselines such as CLAHE [18], UDCP [11], UCM [48], ULAP [57], PIFM [14], and recent deep learning methods including PUIE-Net [24] and USLN [40]. We first present qualitative comparisons on the EUVP [22] and UIEBD [31] datasets, as shown in Fig. 2. Traditional enhancement methods such as CLAHE and UDCP tend to over-amplify contrast and introduce color distortion, particularly in scenes dominated by green or blue wavelengths. These methods lack explicit modeling of wavelength-dependent attenuation, resulting in unnatural color balance and loss of visual realism. Model-based approaches such as PIFM and ULAP improve overall visibility by compensating for light scattering and attenuation; however, they often leave residual haze and amplify noise in low-visibility regions due to inaccurate parameter estimation and limited adaptability to complex environments. Learning-based methods, such as PUIE-Net, UCM, and USLN, produce more visually pleasing results compared to traditional approaches, benefiting from their ability to learn enhancement mappings directly from data. Nevertheless, these methods still exhibit limitations in challenging regions, including incomplete color correction, over-smoothing of fine textures, and inconsistent structural restoration. In contrast, our AdaMSCoL produces visu-

ally superior results, achieving more accurate color recovery, improved contrast, and clearer structural details. Importantly, our method preserves fine textures while avoiding over-enhancement artifacts due to multi-scale structural consistency objective which is effective in handling diverse underwater degradations.

We further evaluate our method quantitatively on three challenging underwater benchmarks: EUVP [22], LSUI [37], and UIEBD [31] (Table 1). Our comparison is conducted against CLAHE [18], IBLA [39], PIFM [14], PUIE-Net [24], RGHS [19], UCM [48], UDCP [11], ULAP [57], USLN [40], Raw [22], ACDC [56], P2CNet [16] and TCTL-Net [25]. On the EUVP dataset, our method achieves the SSIM of 0.838, surpassing the strongest competing method by a significant margin of +0.067. In addition, our approach obtains the lowest MAD value of 70.22, improving by -24.33 compared to prior methods. These results show that AdaMSCoL achieves superior structural reconstruction and reduces pixel-level deviation from reference images, emphasizing its ability to preserve spatial consistency. On the LSUI dataset, our method achieves the best performance across multiple metrics, including the PSNR of 21.64 (+0.32 improvement), the SSIM of 0.846 (+0.038 improvement), and the perceptual quality score with a UIQM of 3.865 (+0.506 improvement). Similarly, on the UIEBD dataset, our method achieves the highest SSIM of 0.859, improving over existing approaches by +0.07. These gains show that AdaMSCoL significantly enhances structural similarity and perceptual quality while maintaining competitive PSNR. Although it strongly improves SSIM, MAD, and UIQM, performance on metrics like GMSD and UCIQE is occasionally weaker, reflecting a trade-off between structural fidelity and contrast- or distribution-based measures, where optimizing for perceptual realism may not align with gradient- or global statistical metrics.

4.5. Ablation Study

We conduct an ablation study to evaluate the effectiveness of each component in the proposed training objective and to analyze the influence of key hyperparameters on enhancement performance. Specifically, in Sec. 4.5.1, we detail the contribution of the individual AdaMSCoL terms by varying the structural, reconstruction, and prior weights. In Sec. 4.5.2, we analyze the impact of the perceptual loss weight on balancing semantic feature alignment and pixel-level fidelity. We primarily adopt partial ablation by varying the relative contribution of each component to maintain stable optimization dynamics. Additionally, we include a full ablation setting where AdaMSCoL is entirely removed ($\lambda_1 = 0$) to quantify its overall contribution.

4.5.1. Effect of AdaMSCoL Components

We analyze the contribution of each component in AdaMSCoL. For this purpose, we vary the weights of the structural (Ψ_{struct}), reconstruction (Ψ_{recon}), and prior (Ψ_{prior}) terms. As shown in Table 2, the performance metrics PSNR and SSIM are influenced differently by each component. Increasing Ψ_{prior} does not consistently improve PSNR, as the highest prior weight ($\Psi_{\text{prior}} = 0.5$) does not yield the best performance. Instead, a moderate prior weight contributes to stable performance, indicating that prior regularization helps maintain global color and illumination consistency under challenging underwater conditions. However, excessive emphasis on either the prior or reconstruction term slightly reduces SSIM, showing a trade-off between global photometric regularization and local structural fidelity. In contrast, higher Ψ_{struct} values lead to a more balanced improvement, enhancing edge preservation and structural consistency without negatively affecting pixel-level accuracy. The reconstruction term, Ψ_{recon}, primarily impacts low-level photometric fidelity, and excessive weighting slightly suppresses the benefits of structural guidance. Among the tested configurations, $\Psi_{\text{struct}} = 0.5$, $\Psi_{\text{recon}} = 0.3$, and $\Psi_{\text{prior}} = 0.2$ achieves the best overall performance, yielding a PSNR of 19.91 dB and the SSIM of 0.8598. This confirms that effective underwater image enhancement requires a careful balance between structural preservation and photometric regularization.

Table 2. Effect of various AdaMSCoL components on reconstruction quality.

Ψ_{struct}	Ψ_{recon}	Ψ_{prior}	**PSNR (dB)** ↑	**SSIM** ↑
0.2	0.3	0.5	19.74	0.8589
0.2	0.5	0.3	19.61	0.8498
0.5	**0.3**	**0.2**	**19.91**	**0.8598**
0.3	0.5	0.2	19.71	0.8495

4.5.2. Effect of Perceptual Loss

We analyze the effect of the perceptual loss weight λ_2 on enhancement performance, as detailed in Table 3. The perceptual loss introduces high-level feature supervision, and its weighting (λ_2) directly controls the balance between semantic consistency and pixel-level reconstruction accuracy. As λ_2 varies, PSNR and SSIM exhibit complementary behavior, showing the trade-off between photometric fidelity and structural alignment.

When λ_2 is small (e.g., 0.01), the perceptual constraint has limited influence, and optimization is dominated by reconstruction objectives, resulting in reasonable pixel-level accuracy (PSNR 19.72 dB) but lower structural similarity (SSIM 0.854), indicating insufficient recovery of fine structural and textural details. Increasing λ_2 to 0.03 improves both PSNR and SSIM to 19.89 dB and 0.857, showing that moderate perceptual supervision helps preserve structural information while maintaining photometric consis-

Table 3. Effect of perceptual loss weight λ_2.

λ_2	PSNR (dB) ↑	SSIM ↑
0.01	19.72	0.854
0.03	19.89	0.857
0.05	**19.91**	**0.859**
0.07	19.76	0.851
0.09	19.70	0.851

tency. The best performance occurs at $\lambda_2 = 0.05$, where PSNR and SSIM reach their maximum values of 19.91 dB and 0.859, respectively. This demonstrates that perceptual guidance enhances semantic and structural fidelity without compromising low-level reconstruction, with SSIM highlighting the role of feature-level supervision in preserving structure in underwater images prone to texture degradation and contrast loss. Further increasing λ_2 degrades performance: at 0.07, PSNR and SSIM drop to 19.76 dB and 0.851, and at 0.09, PSNR decreases to 19.70 dB while SSIM remains 0.851, indicating that excessive perceptual weighting overemphasizes feature alignment, reducing pixel-level accuracy and structural fidelity.

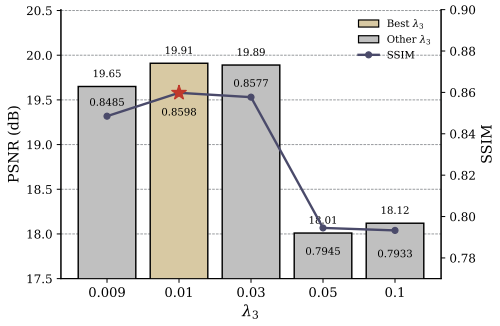


Figure 3. Effect of latent regularization.

4.5.3. Effect of Latent Regularization

We analyze the effect of the latent regularization weight λ_3 on enhancement performance. This hyperparameter controls the degree of constraint on the learned latent representation, promoting structured and smooth feature embeddings that improve generalization and stabilize training. However, excessive weighting can degrade reconstruction quality. As shown in Fig. 3, varying λ_3 significantly impacts PSNR and SSIM, reflecting the trade-off between reconstruction fidelity and latent space consistency. When λ_3 is very small (e.g., 0.009), weak regularization allows greater flexibility in encoding image-specific details, yielding strong pixel-level reconstruction (PSNR 19.65 dB, SSIM 0.8485) but less structured latent representations, slightly reducing spatial consistency. Conversely, larger values (e.g., 0.05–0.1) overly constrain the latent space, degrading PSNR (18.01–18.12 dB) and SSIM (0.793–0.795) and producing overly smooth reconstructions with diminished detail. A moderate regularization of $\lambda_3 = 0.01$ balances these effects, maintaining stable latent representations while preserving fine structural details and photometric variations, achieving the highest PSNR of 19.91 dB and SSIM of 0.8598.

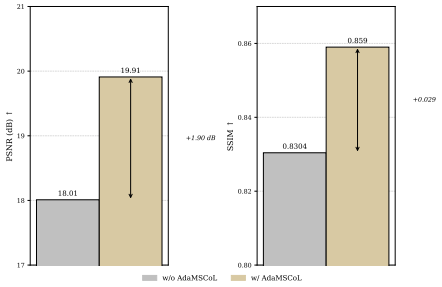


Figure 4. Effect of AdaMSCoL on the network performance.

4.5.4. Impact of AdaMSCoL Loss

We conduct an ablation study to analyze the significance of each loss component in our training objective (Fig. 4). When setting $\lambda_1 = 0$ (removing AdaMSCoL), the performance decreases to PSNR 18.01 and SSIM 0.8304, demonstrating that AdaMSCoL plays a key role in improving both structural similarity and reconstruction accuracy. These results emphasize that AdaMSCoL contributes significantly to the overall performance of the proposed method.

5. Conclusion

We proposed **Ada**ptive **M**ulti-Scale **S**tructural **Co**nsistency **L**oss (**AdaMSCoL**), a novel unsupervised loss for underwater image enhancement that enforces adaptive multi-scale structural consistency, combining cross-scale patch correlation, self-reconstruction consistency, and prior-based regularization. Our experiments across multiple benchmarks demonstrate that AdaMSCoL consistently improves structural fidelity, color realism, and perceptual quality, e.g., SSIM +0.067, MAD -24.40, and UIQM +1.279, outperforming existing unsupervised objectives. In future work, we will explore extending AdaMSCoL to multi-modal underwater image enhancement tasks, such as jointly leveraging RGB, depth, or sonar data, and integrating it with diffusion-based frameworks to further improve generalization, enhance structural and color fidelity, and ensure consistent perceptual quality.

Acknowledgment

We acknowledge the financial support of the National Science and Technology Council (NSTC), R.O.C., under Grants 114-2637-8-027-005 and 114-2637-8-992-009, which enabled this research.

References

[1] D. Akkaynak and T. Treibitz. Sea-thru: A method for removing water from underwater images. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019. 2

[2] Omar Almutiry, Khalid Iqbal, Shariq Hussain, Awais Mahmood, and Habib Dhahri. Underwater images contrast enhancement and its challenges: a survey. *Multimedia Tools and Applications*, 83(5):15125–15150, 2024. 2

[3] Yasmin M Alsakar, Nehal A Sakr, Shaker El-Sappagh, Tamer Abuhmed, and Mohammed Elmogy. Underwater image restoration and enhancement: a comprehensive review of recent trends, challenges, and applications. *The Visual Computer*, 41(6):3735–3783, 2025. 2

[4] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and P. Bekaert. Color balance and fusion for underwater image enhancement. *IEEE Transactions on Image Processing*, 2017. 2

[5] G. Balakrishnan, A. V. Dalca, A. Zhao, J. V. Guttag, F. Durand, and W. T. Freeman. Visual deprojection: Probabilistic recovery of collapsed dimensions. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 171–180, 2019. 3

[6] M. Braik. Hybrid enhanced whale optimization algorithm for contrast and detail enhancement of color images. *Cluster Computing*, 27(1):231–267, 2024. 2

[7] D. M. Chandler. Most apparent distortion: Full-reference image quality assessment and the role of strategy. *Journal of Electronic Imaging*, 19(1):011006, 2010. 6

[8] X. Chen, P. Zhang, L. Quan, C. Yi, and C. Lu. Underwater image enhancement based on deep learning and image formation model. *Computers & Electrical Engineering*, 2021. 1

[9] Z. Cheng, G. Fan, J. Zhou, M. Gan, and C. P. Chen. Fdce-net: Underwater image enhancement with embedding frequency and dual color encoder. *IEEE Transactions on Circuits and Systems for Video Technology*, 2024. 1

[10] R. et al. Cong. Pugan: Physical model-guided underwater image enhancement using gan with dual discriminators. *IEEE Transactions on Image Processing*, 2023. 3

[11] P. Drews, E. Nascimento, F. Moraes, S. Botelho, and M. Campos. Transmission estimation in underwater single images. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 825–830, 2013. 1, 2, 6, 7

[12] 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. 2

[13] Yi-Ning Fan, Geng-Kun Wu, Jia-Zheng Han, Bei-Ping Zhang, and Jie Xu. Innovative underwater image enhancement algorithm: Combined application of adaptive white balance color compensation and pyramid image fusion to submarine algal microscopy. *Image and Vision Computing*, 156:105466, 2025. 2

[14] X. Fu, P. Zhuang, Y. Huang, Y. Liao, X.-P. Zhang, and X. Ding. A retinex-based enhancing approach for single underwater image. In *ICIP*, pages 4572–4576, 2014. 1, 6, 7

[15] X. et al. Fu. Uncertainty-inspired underwater image enhancement. *IEEE Transactions on Image Processing*, 2022. 3

[16] Z. Fu, H. Lin, Y. Yang, S. Chai, L. Sun, Y. Huang, and X. Ding. Unsupervised underwater image restoration: From a homology perspective. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 643–651, 2022. 1, 6, 7

[17] Z. Fu, W. Wang, Y. Huang, X. Ding, and K.-K. Ma. Uncertainty inspired underwater image enhancement. In *Computer Vision – ECCV 2022*, pages 465–482. Springer, 2022. 1

[18] M. S. Hitam, E. A. Awalludin, W. N. J. Wan Yussof, and Z. Bachok. Mixture contrast limited adaptive histogram equalization for underwater image enhancement. In *ICCAT*, pages 1–5, 2020. 2, 6, 7

[19] D. Huang, Y. Wang, W. Song, J. Sequeira, and S. Mavromatis. Shallow-water image enhancement using relative global histogram stretching based on adaptive parameter acquisition. In *International Conference on Multimedia Modeling*, pages 453–465, 2018. 1, 2, 6, 7

[20] M. J. Islam, C. Edge, Y. Xiao, P. Luo, M. Mehtaz, C. Morse, S. S. Enan, and J. Sattar. Semantic segmentation of underwater imagery: Dataset and benchmark. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1769–1776, 2020. 1

[21] M. J. Islam, P. Luo, and J. Sattar. Simultaneous enhancement and super-resolution of underwater imagery for improved visual perception. In *Robotics: Science and Systems*, 2020. 1

[22] M. J. Islam, Y. Xia, and J. Sattar. Fast underwater image enhancement for improved visual perception. *IEEE Robotics and Automation Letters*, 5(2):3227–3234, 2020. 1, 2, 6, 7

[23] J. R. Jebadass and P. Balasubramaniam. Color image enhancement technique based on interval-valued intuitionistic fuzzy set. *Information Sciences*, 653:119811, 2024. 2

[24] X. Ji, X. Wang, N. Leng, L.-Y. Hao, and H. Guo. Dual-branch underwater image enhancement network via multi-scale neighborhood interaction attention learning. *Image and Vision Computing*, 151:105256, 2024. 6, 7

[25] Q. Jiang, Y. Zhang, F. Bao, X. Zhao, C. Zhang, and P. Liu. Two-step domain adaptation for underwater image enhancement. *Pattern Recognition*, 122:108324, 2022. 6, 7

[26] Naveena Tresa Joseph, SN Kumar, and Kannadhasan Suriyan. State-of-the-art techniques for optical underwater image enhancement. *International Journal of Image and Data Fusion*, 16(1):1–32, 2025. 2

[27] Amandeep Kaur, Shalli Rani, and Mohammad Shabaz. Underwater image dehazing using a hybrid gan with bottleneck attention and improved retinex-based optimization. *Scientific Reports*, 15(1):26132, 2025. 2

[28] D. P. Kingma and M. Welling. An introduction to variational autoencoders. *Foundations and Trends in Machine Learning*, 12(4):307–392, 2019. 3

[29] X. Lei, H. Wang, J. Shen, Z. Chen, and W. Zhang. A novel intelligent underwater image enhancement method via color correction and contrast stretching. *Microprocessors and Microsystems*, 107:104040, 2024. 2

[30] Dawa Chyophel Lepcha, Bhawna Goyal, Ayush Dogra, Vivek Vullikanti, J Albert Mayan, Prabhat Kumar Sahu, Sachin Kumar, and U Siddaraj. Underwater image enhancement using colour balancing and morphological residual processing through gamma correction. *Scientific Reports*, 2026. 1

[31] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao. An underwater image enhancement benchmark dataset and beyond. *IEEE Transactions on Image Processing*, 29: 4376–4389, 2020. 1, 2, 6, 7

[32] X. et al. Li. Probabilistic salient object detection via variational inference. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. 3

[33] Risheng Liu, Xin Fan, Ming Zhu, Min Hou, and Zhihua Luo. Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(12):4861–4875, 2020. 1

[34] R. et al. Liu. Twin adversarial contrastive learning for underwater image enhancement and beyond. *IEEE Transactions on Image Processing*, 2022. 2

[35] S. et al. Liu. Adaptive learning attention network for underwater image enhancement. *IEEE Robotics and Automation Letters*, 2022. 2

[36] K. Panetta, C. Gao, and S. Agaian. Human-visual-system-inspired underwater image quality measures. *IEEE Journal of Oceanic Engineering*, 41(3):541–551, 2016. 6

[37] Lintao Peng, Chunli Zhu, and Liheng Bian. U-shape transformer for underwater image enhancement. *IEEE transactions on image processing*, 32:3066–3079, 2023. 2, 6, 7

[38] Y. Peng, K. Cao, and P. Cosman. Generalization of the dark channel prior for single image restoration. *IEEE Transactions on Image Processing*, 2018. 2

[39] Y.-T. Peng and P. C. Cosman. Underwater image restoration based on image blurriness and light absorption. *IEEE Transactions on Image Processing*, 26(4):1579–1594, 2017. 1, 2, 6, 7

[40] S. et al. Raveendran. Fusion-based underwater image enhancement with adaptive color correction. *Expert Systems with Applications*, 2024. 2, 6, 7

[41] Tiago FR Ribeiro, José Areia, Bianca Reis, João N Franco, Fernando Silva, and Rogério Luís de C. Costa. Comparative study of retinex algorithms and deep models for underwater image enhancement. *Signal, Image and Video Processing*, 19(15):1314, 2025. 2

[42] Alzayat Saleh, Marcus Sheaves, Dean Jerry, and Mostafa Rahimi Azghadi. Adaptive uncertainty distribution in deep learning for unsupervised underwater image enhancement. *Available at SSRN 4742854*, 2022. 3, 6

[43] Alzayat Saleh, Marcus Sheaves, Dean Jerry, and Mostafa Rahimi Azghadi. Adaptive deep learning framework for robust unsupervised underwater image enhancement. *Expert Systems with Applications*, 268:126314, 2025. 3, 5, 6

[44] Lyes Saad Saoud, Mahmoud Elmezain, Atif Sultan, Mohamed Heshmat, Lakmal Seneviratne, and Irfan Hussain. Seeing through the haze: A comprehensive review of underwater image enhancement techniques. *IEEE Access*, 12: 145206–145233, 2024. 2

[45] A. et al. Sharma. Lightweight cnn for underwater image enhancement. *IEEE Transactions on Image Processing*, 2023. 2

[46] Xiaowen Shi and Yuan-Gen Wang. Cpdm: Content-preserving diffusion model for underwater image enhancement. *Scientific Reports*, 14(1):31309, 2024. 6

[47] K. Sohn, X. Yan, and H. Lee. Learning structured output representation using deep conditional generative models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2015. 3

[48] W. Song, Y. Wang, D. Huang, and D. Tjondronegoro. A model-based restoration of underwater images. *IEEE Transactions on Image Processing*, 2018. 1, 2, 6, 7

[49] Fukuan Wang, Fei Li, Chaojun Cen, Zhenbo Li, and Qingling Duan. Underwater image enhancement via degradation information extraction and guidance. *Pattern Recognition*, page 113121, 2026. 1, 3

[50] Hao Wang, Weibo Zhang, Lu Bai, and Peng Ren. Metalantis: A comprehensive underwater image enhancement framework. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–19, 2024. 2

[51] N. et al. Wang. Underwater attentional generative adversarial networks for image enhancement. *IEEE Transactions on Human-Machine Systems*, 2023. 3

[52] W. Wang and Y. Yang. A histogram equalization model for color image contrast enhancement. *Signal, Image and Video Processing*, 2024. 2

[53] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612, 2004. 1, 6

[54] Wufeng Xue, Lei Zhang, Xuanqin Mou, and Alan C. Bovik. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE Transactions on Image Processing*, 23(2):684–695, 2014. 6

[55] X. et al. Yan. Unsupervised learning of motion sequences using variational autoencoders. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018. 3

[56] Yuchen Yu and Cheng Qin. An end-to-end underwater-image-enhancement framework based on fractional integral retinex and unsupervised autoencoder. *Fractal and Fractional*, 7(1):70, 2023. 6, 7

[57] S. Zhang, T. Wang, J. Dong, and H. Yu. Underwater image enhancement via extended multi-scale retinex. *Neurocomputing*, 245:1–9, 2017. 1, 6, 7

[58] Wei Zhang, Peng Zhuang, H. H. Sun, Guorui Li, Sam Kwong, and Chun Li. Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement. *IEEE Transactions on Image Processing*, 31:3997–4010, 2022. 1

[59] Weihong Zhang, Xiaobo Li, Yizhao Huang, Shuping Xu, Junwu Tang, and Haofeng Hu. Underwater image enhancement via frequency and spatial domains fusion. *Optics and Lasers in Engineering*, 186:108826, 2025. 2

[60] 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. 2