NOT ALL PIXELS SINK: PHASE-GUIDED REPRESENTATION LEARNING FOR UNDERWATER IMAGE RESTORATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Underwater images suffer from color absorption, light scattering, and non-uniform haze, making reliable restoration crucial for marine science and autonomous navigation. We propose NemoNet, a novel encoder–decoder architecture that leverages phase-guided representation learning to overcome these challenges. The architecture incorporates Spectral–Spatial Attention (SSA) block that couples Fourier phase-based pixel refinement with spatial attention to recover fine textures. These details are most severely degraded in underwater conditions and are critical for perceptually convincing restoration more broadly. Phase-based attention in skip connections ensures that they enhance useful representations instead of propagating artifacts. We introduce a hybrid Un/Supervised loss framework, where comprehensive supervised objectives are complemented by an unsupervised color consistency loss that mitigates wavelength-dependent color shifts in underwater scenes. We further introduce a no-reference Color-Plausibility Quality Index (CPQI) that augments Perceptual Index with a color consistency prior, which conventional metrics fail to capture. Comprehensive experiments demonstrate that the proposed approach outperforms existing state-of-the-art methods on supervised (UIEB, LSUI, EUVP) and unsupervised (U45) underwater image datasets across conventional and proposed metrics. The source code is available at `https://github.com/FindingNemo26/NOT-ALL-PIXELS-SINK.git`.

1 INTRODUCTION

Underwater imaging is essential for tasks such as resource exploration (Zhou et al., 2024b; Chen et al., 2024), marine biology research (Shi et al., 2022; Cheng et al., 2023; Ludvigsen et al., 2007) and autonomous underwater vehicle (AUV) navigation (Sun et al., 2019; Ahn et al., 2018). However, due to absorption and scattering of light, such images suffer from reduced contrast and dominant blue-green color tones, thereby posing significant challenges for both human interpretation and downstream computer vision tasks such as depth estimation and object detection (Zhou et al., 2023; Jaffe, 2014). To address these challenges, underwater image restoration (UIR) helps fix visual distortions by improving image quality and restoring scene visibility (Chi et al., 2019; Henderson et al., 2013).

Traditional approaches primarily rely on hand-crafted priors and transmission map estimation for quality restoration, but they struggle to generalize across dynamic underwater environments due to their reliance on pre-defined models or specific assumptions (Drews et al., 2013; Li et al., 2016). Deep learning has emerged as a powerful paradigm for underwater image restoration, offering flexibility and the ability to directly model complex degradation processes from data. This has led to the development of various successful convolutional neural networks (CNNs) (Zhao et al., 2024; Lin et al., 2024) and adversarial networks (Liu et al., 2022; Cong et al., 2023). Most existing methods emphasize spatial-domain processing, but these approaches often overlook long-range dependencies, amplify noise, and fail to restore global color balance. To address these limitations, Khan et al. (2025) proposed leveraging the phase component of an image, which preserves structural information and aids in enhancing degraded underwater images. Figure 1 presents a t-SNE visualization of the amplitude and phase components of clean and degraded underwater images. Distinct clusters formed by the amplitude of clean and degraded images indicate that degradation significantly alters

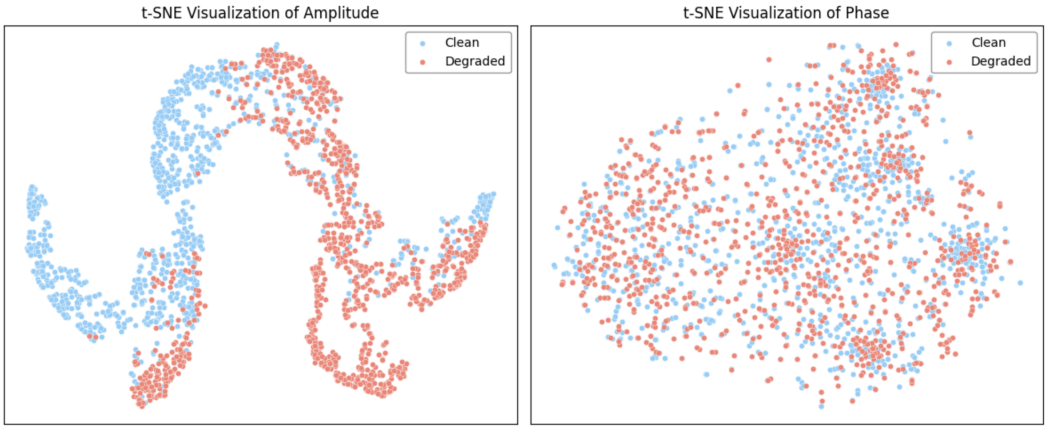


Figure 1: t-SNE projection of amplitude and phase components for clean and degraded underwater images shows that amplitude features form distinct clusters, indicating strong degradation effects. In contrast, phase features largely overlap, suggesting that structural information is preserved.

amplitude information. In contrast, the overlapping phase clusters suggest that phase remains relatively unaffected. Since phase primarily encodes structural details, leveraging it is advantageous for enhancing feature representations of degraded underwater images.

Existing no-reference metrics often show poor alignment with subjective perception. For instance, BRISQUE (Mittal et al., 2012a) is sensitive to rotation, UCIQE (Yang & Sowmya, 2015) and UIQM (Panetta et al., 2015) tend to favor oversaturated images, and in some cases, NIQE (Mittal et al., 2012b) assigns nearly identical scores to images with different color shifts, reflecting its insensitivity to color distortions. Thus, existing no-reference measures lack a reliable way to truly evaluate color fidelity. Overall, the main contributions of our work are:

- We propose NemoNet, an encoder–decoder network with phase-guided representation learning and a Spectral–Spatial Attention (SSA) block to recover fine textures. We include Large Kernel Attention (LKA) for medium-range context and Location-Aware Attention (LAA) to capture directional, long-range dependencies. At the bottleneck, the Omni-Kernel Module handles orientation-specific structures. Optimized Phase-Based Attention in skip connections ensures only useful features are passed, reducing artifacts and improving restoration quality.
- We introduce a Hybrid Un/Supervised Loss for underwater image restoration. It combines conventional supervised objectives with an unsupervised color consistency loss. The unsupervised component specifically addresses color distortions caused by water absorption. This approach ensures more accurate and visually plausible color correction in underwater scenes.
- We present a no-reference Color-Plausibility Quality Index (CPQI), which complements Perceptual Index (PI) (Blau et al., 2018) metric by incorporating a LAB-space color consistency prior, providing a more reliable assessment of quality by capturing both structural detail and color consistency.

2 Related Work

2.1 Underwater Image Enhancement

Early underwater image restoration (UIR) approaches relied on physical priors, such as the dark channel prior and Rayleigh scattering assumptions, to estimate transmission and ambient light (Berman et al., 2020; Ghani & Isa, 2015; Li et al., 2017a). While effective at correcting global color casts, these methods often fail to recover fine details and struggle in complex underwater conditions.

To overcome these limitations, data-driven methods have emerged in two main directions: parameter-guided and end-to-end learning. Parameter-guided CNNs estimate transmission or scattering parameters using physical priors (Kar et al., 2021; Gogireddy & Gogireddy, 2024), but their performance is limited by parameter accuracy. End-to-end frameworks directly map degraded inputs to restored images, improving structural and perceptual quality. Early examples include WaterGAN (Li et al., 2017b), which synthesizes underwater images for unsupervised training, and CycleGAN-based variants (Park et al., 2019; Yan et al., 2023) for unpaired domain adaptation. UWCNN (Li et al., 2020) and WaterNet (Li et al., 2019) further enhanced robustness through diverse water-type modeling and gated fusion strategies, respectively. More recent adversarial and hybrid models, such as perceptual-oriented GANs (Gonzalez-Sabbagh et al., 2024; Islam et al., 2020b; Cong et al., 2023) and Ucolor (Li et al., 2021), combine physical priors with network learning to restore severely degraded regions. Frequency-domain techniques, e.g., wavelet corrections (Jamadandi & Mudenagudi, 2019) and multi-color feature fusion (Li et al., 2021), improve high-frequency detail recovery. Recently, Transformers have been explored to capture long-range dependencies beyond CNN receptive fields. Models such as U-Transformer (Peng et al., 2023a), URSCT (Ren et al., 2022), and related variants (Mu et al., 2023; Liu et al., 2022) leverage global attention for superior restoration. Despite strong performance, their high computational cost poses challenges for real-time applications, motivating lightweight architectures that retain global contextual reasoning.

2.2 Attention Mechanisms in Image Enhancement

Attention-based methods have become increasingly important in underwater image enhancement, enabling models to focus on degraded regions while suppressing noise. Works such as SGUIE-Net (Qi et al., 2022) used semantic region-aware attention across multiple scales to distinguish degradation by object type, improving robustness under varied conditions. Building on this, Walia et al. (2025) incorporated CBAM (Woo et al., 2018), which extends attention to the spatial domain, allowing finer localization of degraded regions. Similarily, RAUNE-Net (Peng et al., 2023b) integrates attention modules in the down-sampling path along with residual learning to capture high-level features, yielding better visual fidelity and generalization. Meanwhile, PhaseFormer (Khan et al., 2025) further introduces phase-aware attention to better preserve structural cues in underwater conditions. These approaches highlight the growing role of specialized attention in underwater image enhancement.

2.3 Underwater Image Quality Assessment

No-reference image quality assessment (NR-IQA) is essential for evaluating underwater image enhancement, yet conventional natural-scene metrics fail to capture underwater distortions. To address this gap, a series of underwater-specific image quality assessment (UIQA) metrics have been proposed. Representative examples include UCIQE (Yang & Sowmya, 2015), which linearly combines chroma, saturation, and contrast statistics, and UIQM (Panetta et al., 2015), which models quality from a human-visual-system perspective by integrating color, contrast, and sharpness cues. Zheng et al. (2022) proposed the Underwater Image Fidelity (UIF) metric, which assesses image naturalness, sharpness, and structural quality using the CIELab color space. Similarily, Yang et al. (2021) developed the frequency-domain UIQA metric (FDUM), which analyzes color, contrast, and sharpness in the frequency domain and leverages the dark channel prior (DCP) (Lee et al., 2016) for improved quality evaluation. However, these methods often favor oversaturated images or fail to capture subtle color shifts, limiting their alignment with human perception and highlighting the need for more reliable underwater image quality assessment techniques.

3 METHOD

Figure 2 illustrates NemoNet, a U-shaped encoder–decoder designed for underwater image restoration. The network integrates spatial and phase information across scales using Spectral-Spatial Attention (SSA) to balance global color correction with local detail recovery. It employs large-kernel attention for medium-range patterns, location-aware attention for positional context, and at the bottleneck, the Omni-Kernel Module to recover structures and long-range context. The Optimized Phase-Based Attention Module further refines features by using phase priors to suppress degraded

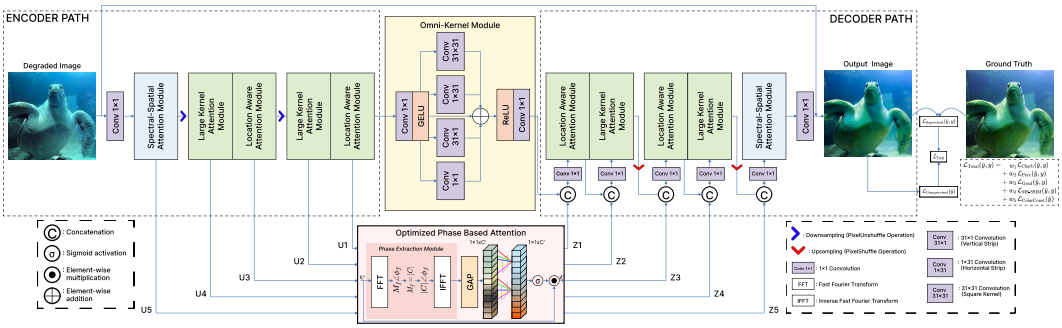


Figure 2: Overview of our proposed NemoNet architecture for UIE. The model takes a degraded underwater image as input and restores it with enhanced visual quality. NemoNet comprises of Spectral-Spatial Block, Large Kernel Attention Module, Location Aware Attention module, Omni-Kernel Module, Optimized Phase Attention Block, and Hybrid Un/Supervised Loss.

signals while emphasizing meaningful structures. Training adopts a Hybrid Un/Supervised loss enabling NemoNet to generate natural, color-corrected images.

3.1 Omni-Kernel Module

At the bottleneck, feature maps from the encoder are reduced to one-fourth of the input resolution. Capturing long-range dependencies is crucial to model relationships between distant regions with similar structures or color degradations. Guo et al. (2025) demonstrated that combining multiple depth-wise convolutions with different kernel shapes in parallel can efficiently capture global contextual features. Building on the strategies proposed in (Cui et al., 2024; Lau et al., 2024), we adopt and extend these approaches in our design. The overall structure of the Omni-Kernel Module (OKM) is illustrated in Figure 2. The Omni-Kernel Module (OKM) applies multiple depth-wise convolutions in parallel with different kernel shapes:

$$F_{OKM} = Conv^{dw}_{31\times1}(F) + Conv^{dw}_{1\times31}(F) + Conv^{dw}_{31\times31}(F) + Conv^{dw}_{1\times1}(F), \tag{1}$$

capturing both anisotropic and isotropic features. The outputs are summed and passed through a 1×1 convolution to model inter-channel relationships. This design efficiently combines local and long-range context, expanding the receptive field without heavy computation.

3.2 Large Kernel and Location Aware Attention

The Spectral-Spatial Attention (SSA) module preserves fine textures and structural details but struggles to connect local features with global context. To address this, we adopt Large Kernel Attention (LKA) Module as proposed in (Li et al., 2024; Lau et al., 2024; Guo et al., 2023), which captures medium-range spatial dependencies using a sequence of a 5 × 5 depth-wise convolution, a 7 × 7 dilated convolution, and a 1 × 1 pointwise convolution. This approximates a 19 × 19 receptive field efficiently, producing attention maps that highlight important medium-scale features. However, LKA lacks explicit positional encoding, limiting its ability to model direction-specific long-range dependencies. To overcome this, we employ Location-Aware Attention (LAA). LAA encodes axis-specific relationships by applying vertical average pooling and horizontal max pooling, followed by a 1 × 1 convolution and GELU activation. A learnable Mix module balances vertical and horizontal contributions, and 3 × 3 convolutions restore channel dimensions and refine local interactions. The fused directional features generate location-sensitive attention maps via matrix multiplication and sigmoid activation, highlighting degraded or important regions while preserving fine details. The overall attention pipeline, integrating LKA and LAA, is illustrated in Figure 3. By combining LKA for medium-range context with LAA for directional positional encoding, our attention mechanism captures both medium-scale dependencies and axis-specific long-range cues, enabling robust modeling of underwater degradations and improving restoration quality.

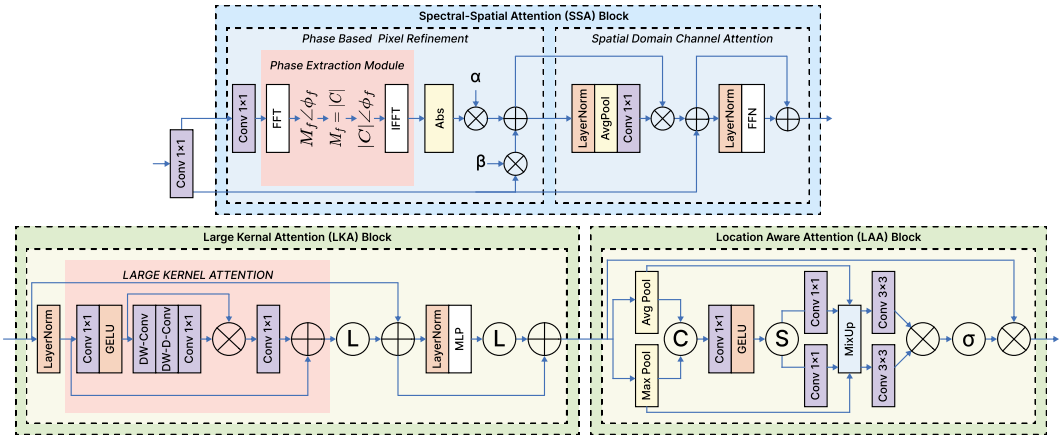


Figure 3: The proposed Spectral–Spatial Attention (SSA) block fuses phase-guided pixel refinement with channel-wise modulation to preserve structural and spectral details. The Large Kernel Attention (LKA) captures medium-range spatial dependencies, while the Location-Aware Attention (LAA) models axis-specific long-range cues, together enhancing context and structure for underwater image restoration.

3.3 Spectral-Spatial Attention Module

Prior studies show that the phase of an image contains the most relevant information (Hansen & Hess, 2007; Oppenheim et al., 1979; Xu et al., 2021), often sufficient for complete image reconstruction. The Spectral-Spatial Attention (SSA) block leverages both spatial and phase information to refine features. As shown in Figure 3, it consists of two sequential components: Phase-Based Pixel Refinement (PBPR) and Spatial-Domain Channel Attention (SDCA).

In PBPR, input features $X \in \mathbb{R}^{H \times W \times C}$ are split into two streams. The first stream is processed with a 1×1 convolution and a Fourier-based phase extraction to obtain a phase-only feature map F_{phase}. The second stream retains spatial-domain features, which are then combined with the phase features via learnable coefficients α and β:

$$Y = \alpha \odot F_{\text{phase}} + \beta \odot X. \tag{2}$$

The SDCA block applies channel-wise attention to Y. Features are normalized, globally pooled, and passed through a 1×1 convolution and sigmoid to produce attention weights S, which modulate the features:

$$F_{\text{attn}} = S \odot Y. \tag{3}$$

Residual connections and a feed-forward network further refine the output:

$$F_{\text{Out}} = F_{\text{attn}} + X + \text{FFN}(\text{LayerNorm}(F_{\text{attn}} + X)). \tag{4}$$

Overall, SSA combines phase-based pixel attention with spatial-domain channel attention, capturing fine local details and complementing coarser processing in other modules.

3.4 Optimized Phase Based Attention

Encoder-decoder networks use skip connections between the encoder and decoder to reduce vanishing gradients and information loss during continuous upsampling and downsampling. However, these connections can also pass redundant or degraded features. Prior works address this via residual enhancement blocks (Zhou et al., 2024a) or channel-wise multi-scale transformer attention (Peng et al., 2023a). However, these attention blocks may forward degraded features because they are processed in spatial domain. Khan et al. (2025) proposed Optimized Phase Based Attention that builds upon Wang et al. (2020) which only involves a handful of parameters while bringing clear performance gain. To address the limitations of conventional skip connections, we integrate the Optimized Phase Attention Block (OPAB) into our network. The workflow of OPAB, including

phase extraction and attention weighting, is illustrated in Figure 2. Unlike direct feature forwarding, OPAB focuses on phase information, which is more robust to scattering and color distortion than amplitude. Given encoder features U_i, a Phase Extraction Module (PEM) first extracts phase cues, followed by global average pooling (GAP) to aggregate spatial information. Cross-channel dependencies are then modeled using a learnable 1D convolution with an adaptive kernel size:

$$k = \left| \frac{\log_2(C')}{\gamma} + \frac{b}{\gamma} \right|_{odd} \tag{5}$$

where $C' = 2^{i-1}C$ is the channel dimension, and $|t|_{odd}$ indicates the nearest odd number of t. This adaptive strategy mitigates the over-smoothing or under-smoothing that occurs with fixed kernels, allowing the network to dynamically adjust receptive fields according to feature dimensionality. The resulting attention weights are applied to the phase features to yield phase-aware attentive maps:

$$Z_i = U_i \otimes \sigma\Big(\omega_k\big(\text{GAP}(\text{PEM}(U_i))\big)\Big), \tag{6}$$

where ω_k is the 1D convolution operator with kernel size k, $\sigma(\cdot)$ denotes the sigmoid activation, and $\otimes$ represents channel-wise multiplication. By emphasizing structurally reliable phase cues and suppressing degraded information, OPAB ensures that skip connections deliver robust, informative features to the decoder. This lightweight mechanism enhances restoration quality without the computational cost of spatial-domain transformer blocks.

3.5 Hybrid Un/Supervised Loss (HUSL)

To optimize training, we propose a Hybrid Un/Supervised Loss (HUSL) that integrates complementary objectives through a weighted summation. The supervised loss combines Charbonnier loss (L_C) (Bruhn et al., 2005), Perceptual loss (L_P) (Johnson et al., 2016), Gradient loss (L_G) (Ribeiro & Elsayed, 1995), and Multi-Scale SSIM (MS-SSIM) loss (L_M) (Wang et al., 2003) which addresses structural fidelity, perceptual quality, edge preservation, and multi-scale similarity, respectively. An unsupervised Color Constancy loss (L_{CC}) is further incorporated to mitigate underwater color distortions. The final objective is

$$L_{Total} = L_{Supervised} + L_{Unsupervised}, \tag{7}$$

$$L_{Supervised} = \Omega_1 * L_C + \Omega_2 * L_P + \Omega_3 * L_G + \Omega_4 * L_M \tag{8}$$

$$L_{Unsupervised} = \Omega_5 * L_{CC} \tag{9}$$

The weights $\Omega_1 = 0.2741$, $\Omega_2 = 0.1680$, $\Omega_3 = 0.2222$, $\Omega_4 = 0.3357$, and $\Omega_5 = 0.1500$ are empirically determined. This formulation provides a flexible and effective loss for robust underwater image restoration.

4 Underwater Image Quality Assessment

Existing unsupervised image quality assessment methods struggle to capture the diverse degradations present in underwater imagery. For instance, UCIQE and UICM often overemphasize saturation, leading to inflated scores for visually unrealistic outputs. NIQE, while widely used, is prone to introducing color shift artifacts that are inconsistent with human perception. Similarly, UISM tends to amplify noise when estimating sharpness, while BRISQUE exhibits sensitivity to image orientation, reducing its robustness for underwater scenarios. These shortcomings highlight the need for a metric that jointly considers perceptual fidelity and color plausibility, ensuring a more reliable evaluation of underwater image quality.

4.1 Color-Plausibility Quality Index (CPQI)

To address the unique challenges of underwater image quality assessment, we introduce Color-Plausibility Quality Index (CPQI) that jointly accounts for perceptual realism and color fidelity. Our method builds upon the Perceptual Index (PI), and augments it with a Color Plausibility (CP).

The final quality score is obtained through a weighted combination of the perceptual and color plausibility components, defined as:

$$Q_{\text{ours}} = \text{PI} + \lambda \cdot \text{CP}, \tag{10}$$

where λ controls the contribution of the color plausibility term.

The Perceptual Index (PI) is an unsupervised metric reflecting human perception of image naturalness. It combines the Ma-score, which predicts aesthetic quality, and NIQE, which measures deviations from natural scene statistics:

$$\text{PI} = \frac{1}{2}\big((10 - \text{Ma}) + \text{NIQE}\big). \tag{11}$$

Lower PI indicates better perceptual quality.

Our Color Plausibility (CP) term explicitly evaluates chromatic naturalness in underwater images. It is computed in the *LAB* color space, which separates luminance from chromatic components, making it well-suited to quantify color realism. Let a and b denote the chromatic channels. CP is defined as:

$$\text{CP}(I) = \sqrt{\mu_a^2 + \mu_b^2} + \big|\sigma_a - \sigma_b\big|, \tag{12}$$

The first term, $\sqrt{\mu_a^2 + \mu_b^2}$, measures the deviation of mean chromaticity from neutral, penalizing global color shifts such as excessive blue or green tones, making it particularly robust for underwater images. The second term, $|\sigma_a - \sigma_b|$, quantifies imbalance between chromatic channel variances, penalizing unnatural color distributions that frequently occur in restored underwater images. By combining these two measures, CP provides a robust assessment of color realism.

Table 1: CPQI demonstrates the best overall performance in terms of SRCC and KRCC among the compared IQA metrics, while achieving competitive PLCC and RMSE values.

Method	SRCC ↑	PLCC ↑	KRCC ↑	RMSE ↓
NIQE (Mittal et al., 2012b)	0.049	0.138	0.034	2.026
BRISQUE (Mittal et al., 2012a)	0.147	0.157	0.099	2.012
UCIQE (Yang & Sowmya, 2015)	0.156	0.122	0.107	**1.998**
UIQM (Panetta et al., 2015)	0.064	0.075	0.044	2.031
ILNIQE (Zhang et al., 2015)	0.149	**0.187**	0.100	2.037
PIQE (Venkatanath et al., 2015)	0.024	0.041	0.017	2.053
NRQM (Ma et al., 2017)	0.036	0.038	0.024	2.035
WaDIQaM (Bosse et al., 2017)	0.021	0.042	0.013	2.036
DBCNN (Zhang et al., 2018)	0.037	0.073	0.024	2.034
MUSIQ (Ke et al., 2021)	0.105	0.105	0.070	2.027
LIQE (Zhang et al., 2023)	0.161	0.109	0.105	2.025
CNNIQA (Kang et al., 2014)	0.069	0.076	0.046	2.029
NIMA (Talebi & Milanfar, 2018)	0.033	0.060	0.022	2.032
HyperIQA Su et al. (2020)	0.087	0.098	0.055	2.030
CPQI	**0.172**	0.124	**0.113**	2.024

Figure 4: Comparison of underwater image quality metrics (UCIQE / CPQI) against MOS: Top-left (0.50/10.50/4.26), Top-right (0.57/17.91/0.88), Bottom-left (0.62/15.22/6.65), Bottom-right (0.62/18.95/1.42). Images appear perceptually low-quality or oversaturated. CPQI aligns better with MOS, with lower values indicating higher perceived quality.

4.2 Experiments

Datasets. We evaluate our method on the SAUD2.0 dataset, which contains 200 raw and 2,400 enhanced underwater images captured in real-world scenarios. Subjective quality scores are provided

as mean opinion scores (MOS) obtained via the single-stimulus absolute category rating (SS-ACR) protocol.

Implementation Details. To ensure unbiased evaluation, we adopt a random 75–25 train–test split, using 75% of the images to fit the non-linear mapping function and the remaining 25% exclusively for testing. Since objective predictions do not directly align with MOS, we apply a four-parameter cubic polynomial regression to map predicted scores to subjective assessments.

Evaluation Metrics. We quantify the agreement between predicted and subjective scores using standard correlation metrics: Pearson linear correlation coefficient (PLCC), Spearman rank correlation coefficient (SRCC), Kendall rank correlation coefficient (KRCC), and root mean square error (RMSE). Results are summarized in Table 1. Our proposed metric achieves the highest SRCC and KRCC while maintaining competitive PLCC and RMSE.

5 Experiments

5.1 Datasets

For training, we use both synthetic and real-world datasets. Synthetic datasets include UIEB (Li et al., 2019), though models trained on them often struggle with real scenes. To mitigate this, we also use real-world datasets LSUI (Peng et al., 2023a) and EUVP (Islam et al., 2020b), sampling 800 images from UIEB, and 2000 each from EUVP and LSUI for training. This combination enables the model to generalize better across diverse underwater scenarios.

For evaluation, we assembled paired test sets consisting of 90 UIEB images, 200 EUVP samples, and 200 LSUI samples, with no overlap with the training split. To further assess cross-domain generalization, we also included 200 unpaired samples from the SUIM dataset (Islam et al., 2020a). This comprehensive test collection allows us to rigorously examine both performance and generalization under diverse underwater imaging conditions.

5.2 Implementation Details

The implementation of the proposed network was carried out using PyTorch and all experiments were conducted on NVIDIA Tesla T4x2 GPU. Training was performed with a batch size of 8, using an initial learning rate of 2×10^{-4}, which was gradually decayed to a minimum of 1×10^{-6}. We employed the AdamW optimizer in combination with a Cosine Annealing learning rate scheduler. The models were trained for 500 epochs, with input images resized to 256×256 pixels. To improve robustness, standard data augmentation techniques like including random cropping, flipping, rotation, transposition, and scaling were applied.

Table 2: Quantitative comparison of underwater image restoration methods on the UIEB dataset.

Method	PSNR ↑	SSIM ↑	MS-SSIM ↑	LPIPS ↓	GMSD ↓	BRISQUE ↓	CIEDE2000 ↓	VIF ↑	CPQI ↓
TACL (Liu et al., 2022)	22.33	0.841	0.940	0.137	0.0528	15.58	8.81	0.545	12.545
PUIE-Net (Fu et al., 2022b)	21.97	0.883	0.953	0.108	0.0395	21.86	9.86	0.720	11.144
UsUIR (Fu et al., 2022a)	20.65	0.864	0.936	0.137	0.0510	29.91	11.55	0.638	11.772
Phaseformer (Khan et al., 2025)	22.44	0.867	0.950	0.131	0.0429	26.40	7.90	0.634	12.478
Spectroformer (Khan et al., 2024)	24.02	0.881	0.947	0.147	0.0411	35.59	**5.51**	0.610	12.629
UShape (Peng et al., 2023a)	19.46	0.645	0.841	0.376	0.1293	55.06	10.83	0.303	12.831
SyreaNet (Wen et al., 2023)	16.61	0.814	0.913	0.184	0.0638	19.50	15.64	0.549	13.170
UDNet (Saleh et al., 2025)	19.16	0.819	0.910	0.160	0.0629	22.53	12.77	**0.758**	11.003
CCLNet (Liu et al., 2024)	20.78	0.875	0.940	0.130	0.0443	21.92	10.90	0.588	12.611
CEVAE (Martinel & Pucci, 2025)	19.06	0.637	0.856	0.434	0.1251	60.87	11.81	0.354	11.699
NemoNet	**24.23**	**0.914**	**0.956**	**0.073**	**0.0350**	**15.41**	8.00	0.735	**10.380**

5.3 Analysis on Synthetic Datasets

We perform a quantitative evaluation of the proposed approach against state-of-the-art (SOTA) methods. The results on the UIEB datasets are presented in Table 2, while qualitative comparisons are illustrated in Figure 4. The performance of our method is competitive with existing techniques.

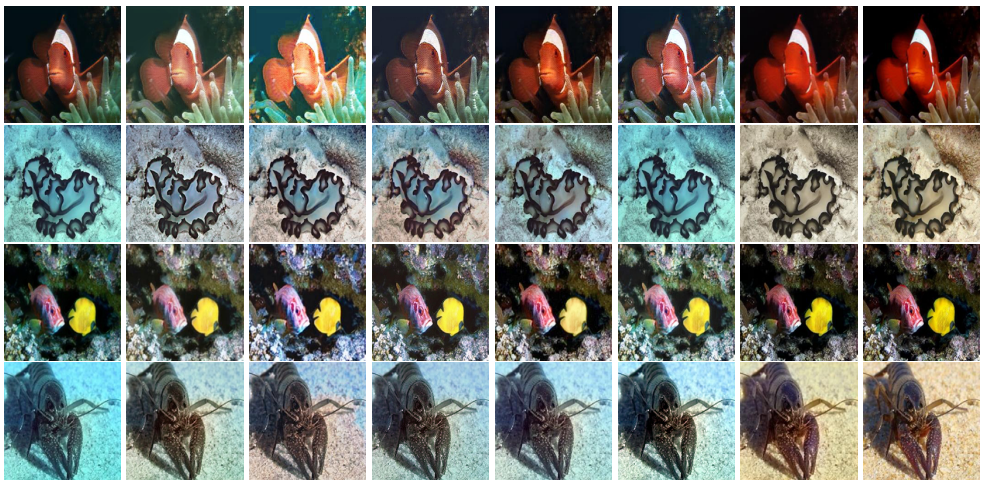
5.4 Analysis on Real-world Datasets

To evaluate the practicality of the proposed approach in real-world conditions, we present results obtained from the LSUI and EUVP datasets. The quantitative assessment employs standard metrics, with detailed scores provided in Table 3 and 4, respectively. These results demonstrate that our method achieves notable improvements in both color balance and scene visibility. Moreover, we further validate the generalization ability of our approach on the unsupervised U45 dataset as presented in Table 5.

In addition, qualitative comparisons on these datasets are presented in Figure 5.

Table 3: Quantitative comparison of underwater image restoration methods on the LSUI dataset.

Method	PSNR ↑	SSIM ↑	MS-SSIM ↑	LPIPS ↓	GMSD ↓	BRISQUE ↓	CIEDE2000 ↓	VIF ↑	CPQI ↓
TACL (Liu et al., 2022)	20.402	0.780	0.924	0.189	0.059	21.401	11.892	0.391	11.189
PUIE-Net (Fu et al., 2022b)	20.744	0.827	0.942	0.193	0.041	28.234	11.902	0.521	9.914
UsUIR (Fu et al., 2022a)	19.237	0.783	0.924	0.235	0.052	30.222	14.132	0.422	8.073
Phaseformer (Khan et al., 2025)	20.393	0.791	0.935	0.142	0.045	21.969	11.335	0.489	12.557
Spectroformer (Khan et al., 2024)	20.421	0.801	0.931	0.222	0.049	29.656	12.142	0.447	13.637
UShape (Peng et al., 2023a)	24.369	0.792	0.940	0.257	0.056	38.824	7.373	0.391	12.023
SyreaNet (Wen et al., 2023)	17.891	0.785	0.913	0.275	0.066	31.927	14.069	0.385	11.347
UDNet (Saleh et al., 2025)	19.572	0.796	0.931	0.227	0.044	29.913	13.330	0.543	10.195
CCLNet (Liu et al., 2024)	18.914	0.783	0.886	0.252	0.060	31.114	13.857	0.394	10.822
CEVAE (Martinel & Pucci, 2025)	26.812	0.809	0.953	0.230	0.046	39.464	4.877	0.414	11.811
NemoNet	**28.519**	**0.923**	**0.980**	**0.057**	**0.017**	**20.670**	**4.830**	**0.652**	**7.009**



(a) Input (b) PUIE-Net (c) CCL-Net (d) Phase (e) UShape (f) Spectro (g) NemoNet (h) GT

Figure 5: Visual comparison showing that images restored by NemoNet closely align with the ground truth, demonstrating effective color correction and detail preservation

5.5 Computational Complexity Analysis

We further evaluate the computational complexity of our method against existing state-of-the-art approaches, considering the number of trainable parameters and floating-point operations (FLOPs). As summarized in Table 6, the results show that our method achieves significantly lower computational complexity compared to current state-of-the-art techniques for underwater image restoration.

6 Ablation Studies

We perform ablation studies to assess the contribution of each component in NemoNet. Table 7 shows results averaged across the UIEB, EUVP, and LSUI datasets.

Table 4: Quantitative comparison of underwater image restoration methods on the EUVP dataset.

Method	PSNR ↑	SSIM ↑	MS-SSIM ↑	LPIPS ↓	GMSD ↓	BRISQUE ↓	CIEDE2000 ↓	VIF ↑	CPQI ↓
TACL (Liu et al., 2022)	19.172	0.753	0.917	0.210	0.0629	27.471	13.623	0.352	9.997
PUIE-Net (Fu et al., 2022b)	20.011	0.786	0.929	0.212	0.0517	26.832	12.332	0.416	9.397
UsUIR (Fu et al., 2022a)	18.310	0.749	0.914	0.250	0.0594	27.865	15.721	0.387	9.183
Phaseformer (Khan et al., 2025)	19.536	0.749	0.928	0.176	0.0494	24.532	12.536	0.377	11.429
Spectroformer (Khan et al., 2024)	19.682	0.780	0.923	0.236	0.0543	25.677	13.091	0.394	12.743
UShape (Peng et al., 2023a)	22.551	0.822	0.947	0.204	0.0462	28.482	10.015	0.416	11.023
SyreaNet (Wen et al., 2023)	18.755	0.774	0.912	0.253	0.0722	26.603	13.872	0.356	11.183
UDNet (Saleh et al., 2025)	19.642	0.768	0.926	0.231	0.0486	26.880	13.961	0.440	9.510
CCLNet (Liu et al., 2024)	18.707	0.762	0.887	0.250	0.0665	28.301	14.850	0.372	9.435
CEVAE (Martinel & Pucci, 2025)	23.781	0.803	0.953	0.190	0.0406	27.562	8.323	0.391	10.829
NemoNet	**27.591**	**0.908**	**0.973**	**0.098**	**0.0213**	**24.351**	**5.722**	**0.554**	**7.235**

Table 5: Quantitative comparison of underwater image restoration methods on the unsupervised U45 dataset.

Method	BRISQUE ↓	CPQI ↓
Phaseformer (Khan et al., 2025)	38.341	13.074
UDNet (Saleh et al., 2025)	8.543	10.554
Spectroformer (Khan et al., 2024)	19.521	14.368
UShape (Peng et al., 2023a)	13.402	13.524
SyreaNet (Wen et al., 2023)	8.729	9.603
CCLNet (Liu et al., 2024)	8.830	10.814
TACL (Liu et al., 2022)	10.826	12.447
PUIE-Net (Fu et al., 2022b)	8.673	13.103
UsUIR (Fu et al., 2022a)	7.466	**8.630**
CEVAE (Martinel & Pucci, 2025)	19.511	11.082
NemoNet	**7.274**	10.981

Table 6: Comparison of model complexity and computational cost for underwater image restoration methods.

Method	**#Param** ($\times 10^6$)	**FLOPs** ($\times 10^6$)
WaterNet (Li et al., 2019)	24.8	193.7
U-shape (Peng et al., 2023a)	65.6	66.2
UIECL (Li et al., 2022)	13.3	31.0
TACL (Liu et al., 2022)	11.3	56.8
UGAN (Fabbri et al., 2018)	57.1	18.3
Ucolor (Li et al., 2021)	157.4	443.9
SGUIE-Net (Qi et al., 2022)	18.5	123.5
Ours	**2.615**	**14.1**

Table 7: Each row shows the effect of adding or removing specific components, highlighting the contribution of the color loss and other modules to overall restoration performance.

Method	PSNR ↑	SSIM ↑	MS-SSIM ↑	LPIPS ↓	GMSD ↓	BRISQUE ↓	CIEDE2000 ↓	VIF ↑	CPQI ↓
SSA	23.568	0.883	0.956	0.113	0.0311	5.623	8.659	0.621	15.297
SSA+LKA	24.964	0.899	0.964	0.090	0.0269	5.626	7.586	0.630	14.085
SSA+LAA	24.910	0.907	0.960	0.102	0.0300	5.611	8.164	0.621	14.421
SSA+OKM	24.257	0.889	0.960	0.101	0.0292	5.529	7.782	0.603	13.218
SSA+LKA+OKM	25.610	0.905	0.967	0.084	0.0261	5.675	6.952	0.632	13.599
SSA+LAA+OKM	25.081	0.898	0.963	0.093	0.0275	5.598	7.265	0.622	15.382
SSA+LKA+LAA+OKM	25.904	0.902	0.966	0.086	0.0262	5.682	7.016	0.639	12.072
NemoNet (without color loss)	26.321	0.912	0.969	0.078	0.0249	5.757	6.526	0.636	12.468
NemoNet	**26.777**	**0.915**	**0.970**	**0.076**	**0.0244**	**5.837**	**6.185**	**0.647**	**11.651**

The results show steady improvements at each step, with a significant gain observed after introducing the color loss, which helps the model handle underwater color distortions more effectively.

7 Conclusion

This paper introduces NemoNet, a phase-guided encoder–decoder for underwater image restoration. By combining Spectral–Spatial Attention with phase-based skip connections, NemoNet recovers fine textures while suppressing artifacts. A hybrid Un/Supervised loss with color consistency and the proposed CPQI metric further enhance restoration quality. Extensive experiments show that NemoNet surpasses state-of-the-art methods and offers strong potential for applications in underwater vision and navigation.

REFERENCES

Jonghyun Ahn, Shinsuke Yasukawa, Takashi Sonoda, Yuya Nishida, Kazuo Ishii, and Tamaki Ura. An optical image transmission system for deep sea creature sampling missions using autonomous underwater vehicle. *IEEE Journal of Oceanic Engineering*, 45(2):350–361, 2018.

Dana Berman, Deborah Levy, Shai Avidan, and Tali Treibitz. Underwater single image color restoration using haze-lines and a new quantitative dataset. *IEEE transactions on pattern analysis and machine intelligence*, 43(8):2822–2837, 2020.

Yochai Blau, Roey Mechrez, Radu Timofte, Tomer Michaeli, and Lihi Zelnik-Manor. The 2018 pirm challenge on perceptual image super-resolution. In *Proceedings of the European conference on computer vision (ECCV) workshops*, pp. 0–0, 2018.

Sebastian Bosse, Dominique Maniry, Klaus-Robert Müller, Thomas Wiegand, and Wojciech Samek. Deep neural networks for no-reference and full-reference image quality assessment. *IEEE Transactions on image processing*, 27(1):206–219, 2017.

Andrés Bruhn, Joachim Weickert, and Christoph Schnörr. Lucas/kanade meets horn/schunck: Combining local and global optic flow methods. *International journal of computer vision*, 61(3): 211–231, 2005.

Weiling Chen, Boqin Cai, Sumei Zheng, Tiesong Zhao, and Ke Gu. Perception-and-cognition-inspired quality assessment for sonar image super-resolution. *IEEE Transactions on Multimedia*, 26:6398–6410, 2024.

Jinguang Cheng, Zongwei Wu, Shuo Wang, Cédric Demonceaux, and Qiuping Jiang. Bidirectional collaborative mentoring network for marine organism detection and beyond. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(11):6595–6608, 2023.

Zhixiang Chi, Xiao Shu, and Xiaolin Wu. Joint demosaicking and blind deblurring using deep convolutional neural network. In *2019 IEEE International conference on image processing (ICIP)*, pp. 2169–2173. IEEE, 2019.

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.

Yuning Cui, Wenqi Ren, and Alois Knoll. Omni-kernel network for image restoration. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 1426–1434, 2024.

Paul Drews, Erickson Nascimento, Filipe Moraes, Silvia Botelho, and Mario Campos. Transmission estimation in underwater single images. In *Proceedings of the IEEE international conference on computer vision workshops*, pp. 825–830, 2013.

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

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, 2022a.

Zhenqi Fu, Wu Wang, Yue Huang, Xinghao Ding, and Kai-Kuang Ma. Uncertainty inspired underwater image enhancement. In *European conference on computer vision*, pp. 465–482. Springer, 2022b.

Ahmad Shahrizan Abdul Ghani and Nor Ashidi Mat Isa. Underwater image quality enhancement through integrated color model with rayleigh distribution. *Applied soft computing*, 27:219–230, 2015.

Yugandhar Reddy Gogireddy and Jithendra Reddy Gogireddy. Advanced underwater image quality enhancement via hybrid super-resolution convolutional neural networks and multi-scale retinex-based defogging techniques. *arXiv preprint arXiv:2410.14285*, 2024.

Salma Gonzalez-Sabbagh, Antonio Robles-Kelly, and Shang Gao. Dgd-cgan: A dual generator for image dewatering and restoration. *Pattern Recognition*, 148:110159, 2024.

Meng-Hao Guo, Cheng-Ze Lu, Zheng-Ning Liu, Ming-Ming Cheng, and Shi-Min Hu. Visual attention network. *Computational visual media*, 9(4):733–752, 2023.

Xiaojiao Guo, Yihang Dong, Xuhang Chen, Weiwen Chen, Zimeng Li, FuChen Zheng, and Chi-Man Pun. Underwater image restoration via polymorphic large kernel cnns. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5. IEEE, 2025.

Bruce C Hansen and Robert F Hess. Structural sparseness and spatial phase alignment in natural scenes. *Journal of the Optical Society of America A*, 24(7):1873–1885, 2007.

Jon Henderson, Oscar Pizarro, Matthew Johnson-Roberson, and Ian Mahon. Mapping submerged archaeological sites using stereo-vision photogrammetry. *International Journal of Nautical Archaeology*, 42(2):243–256, 2013.

Md Jahidul Islam, Chelsey Edge, Yuyang Xiao, Peigen Luo, Muntaqim Mehtaz, Christopher Morse, Sadman Sakib Enan, and Junaed Sattar. Semantic segmentation of underwater imagery: Dataset and benchmark. In *2020 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pp. 1769–1776. IEEE, 2020a.

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, 2020b.

Jules S Jaffe. Underwater optical imaging: the past, the present, and the prospects. *IEEE Journal of Oceanic Engineering*, 40(3):683–700, 2014.

Adarsh Jamadandi and Uma Mudenagudi. Exemplar-based underwater image enhancement augmented by wavelet corrected transforms. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pp. 11–17, 2019.

Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pp. 694–711. Springer, 2016.

Le Kang, Peng Ye, Yi Li, and David Doermann. Convolutional neural networks for no-reference image quality assessment. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1733–1740, 2014.

Aupendu Kar, Sobhan Kanti Dhara, Debashis Sen, and Prabir Kumar Biswas. Zero-shot single image restoration through controlled perturbation of koschmieder's model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16205–16215, 2021.

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.

MD Raqib Khan, Anshul Negi, Ashutosh Kulkarni, Shruti S Phutke, Santosh Kumar Vipparthi, and Subrahmanyam Murala. Phaseformer: Phase-based attention mechanism for underwater image restoration and beyond. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 9618–9629. IEEE, 2025.

Raqib Khan, Priyanka Mishra, Nancy Mehta, Shruti S Phutke, Santosh Kumar Vipparthi, Sukumar Nandi, and Subrahmanyam Murala. Spectroformer: Multi-domain query cascaded transformer network for underwater image enhancement. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pp. 1454–1463, 2024.

Kin Wai Lau, Lai-Man Po, and Yasar Abbas Ur Rehman. Large separable kernel attention: Rethinking the large kernel attention design in cnn. *Expert Systems with Applications*, 236:121352, 2024.

Sungmin Lee, Seokmin Yun, Ju-Hun Nam, Chee Sun Won, and Seung-Won Jung. A review on dark channel prior based image dehazing algorithms. *EURASIP Journal on Image and Video Processing*, 2016(1):4, 2016.

Chenghao Li, Pengbo Shi, Qingzi Chen, Jirui Liu, and Lingyun Zhu. Lkca: large kernel convolutional attention. In *International Conference on Optics and Machine Vision (ICOMV 2024)*, volume 13179, pp. 142–147. SPIE, 2024.

Chong-Yi Li, Ji-Chang Guo, Run-Min Cong, Yan-Wei Pang, and Bo Wang. Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior. *IEEE Transactions on Image Processing*, 25(12):5664–5677, 2016.

Chongyi Li, Jichang Guo, Chunle Guo, Runmin Cong, and Jiachang Gong. A hybrid method for underwater image correction. *Pattern Recognition Letters*, 94:62–67, 2017a.

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.

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

Kunqian Li, Li Wu, Qi Qi, Wenjie Liu, Xiang Gao, Liqin Zhou, and Dalei Song. Beyond single reference for training: Underwater image enhancement via comparative learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(6):2561–2576, 2022.

Peng Lin, Yafei Wang, Yuanyuan Li, Zihao Fan, and Xianping Fu. Underwater color correction network with knowledge transfer. *IEEE Transactions on Multimedia*, 26:8088–8103, 2024.

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.

Yi Liu, Qiuping Jiang, Xinyi Wang, Ting Luo, and Jingchun Zhou. Underwater image enhancement with cascaded contrastive learning. *IEEE Transactions on Multimedia*, 2024.

Martin Ludvigsen, Bjørn Sortland, Geir Johnsen, and Hanumant Singh. Applications of geo-referenced underwater photo mosaics in marine biology and archaeology. *Oceanography*, 20 (4):140–149, 2007.

Chao Ma, Chih-Yuan Yang, Xiaokang Yang, and Ming-Hsuan Yang. Learning a no-reference quality metric for single-image super-resolution. *Computer Vision and Image Understanding*, 158:1–16, 2017.

Niki Martinel and Rita Pucci. Physics informed capsule enhanced variational autoencoder for underwater image enhancement. *arXiv preprint arXiv:2506.04753*, 2025.

Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on image processing*, 21(12):4695–4708, 2012a.

Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a “completely blind” image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012b.

Pan Mu, Hanning Xu, Zheyuan Liu, Zheng Wang, Sixian Chan, and Cong Bai. A generalized physical-knowledge-guided dynamic model for underwater image enhancement. In *Proceedings of the 31st ACM international conference on multimedia*, pp. 7111–7120, 2023.

A Oppenheim, Jae Lim, Gary Kopec, and SC Pohlig. Phase in speech and pictures. In *ICASSP'79. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pp. 632–637. IEEE, 1979.

Karen Panetta, Chen Gao, and Sos Agaian. Human-visual-system-inspired underwater image quality measures. *IEEE Journal of Oceanic Engineering*, 41(3):541–551, 2015.

Jaihyun Park, David K Han, and Hanseok Ko. Adaptive weighted multi-discriminator cyclegan for underwater image enhancement. *Journal of Marine Science and Engineering*, 7(7):200, 2019.

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

Wangzhen Peng, Chenghao Zhou, Runze Hu, Jingchao Cao, and Yutao Liu. Raune-net: a residual and attention-driven underwater image enhancement method. In *International Forum on Digital TV and Wireless Multimedia Communications*, pp. 15–27. Springer, 2023b.

Qi Qi, Kunqian Li, Haiyong Zheng, Xiang Gao, Guojia Hou, and Kun Sun. Sguie-net: Semantic attention guided underwater image enhancement with multi-scale perception. *IEEE Transactions on Image Processing*, 31:6816–6830, 2022.

Tingdi Ren, Haiyong Xu, Gangyi Jiang, Mei Yu, and Ting Luo. Reinforced swin-convs transformer for underwater image enhancement. *arXiv preprint arXiv:2205.00434*, 2022.

JL Ribeiro and EA Elsayed. A case study on process optimization using the gradient loss function. *International Journal of Production Research*, 33(12):3233–3248, 1995.

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.

Zhensheng Shi, Cheng Guan, Qianqian Li, Ju Liang, Liangjie Cao, Haiyong Zheng, Zhaorui Gu, and Bing Zheng. Detecting marine organisms via joint attention-relation learning for marine video surveillance. *IEEE Journal of Oceanic Engineering*, 47(4):959–974, 2022.

Shaolin Su, Qingsen Yan, Yu Zhu, Cheng Zhang, Xin Ge, Jinqiu Sun, and Yanning Zhang. Blindly assess image quality in the wild guided by a self-adaptive hyper network. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3667–3676, 2020.

Sibo Sun, Yingchun Chen, Longhao Qiu, Guangpu Zhang, and Chunhui Zhao. Inverse synthetic aperture sonar imaging of underwater vehicles utilizing 3-d rotations. *IEEE Journal of Oceanic Engineering*, 45(2):563–576, 2019.

Hossein Talebi and Peyman Milanfar. Nima: Neural image assessment. *IEEE transactions on image processing*, 27(8):3998–4011, 2018.

Narasimhan Venkatanath, D Praneeth, S Channappayya Sumohana, S Medasani Swarup, et al. Blind image quality evaluation using perception based features. In *2015 twenty first national conference on communications (NCC)*, pp. 1–6. IEEE, 2015.

Jaskaran Walia, Shravan Venkatraman, et al. Fusion: Frequency-guided underwater spatial image reconstruction. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 843–852, 2025.

Qilong Wang, Banggu Wu, Pengfei Zhu, Peihua Li, Wangmeng Zuo, and Qinghua Hu. Eca-net: Efficient channel attention for deep convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11534–11542, 2020.

Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *The thrity-seventh asilomar conference on signals, systems & computers, 2003*, volume 2, pp. 1398–1402. Ieee, 2003.

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. *arXiv preprint arXiv:2302.08269*, 2023.

Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, pp. 3–19, 2018.

Qinwei Xu, Ruipeng Zhang, Ya Zhang, Yanfeng Wang, and Qi Tian. A fourier-based framework for domain generalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 14383–14392, 2021.

Haorui Yan, Zhenwei Zhang, Jing Xu, Tingting Wang, Ping An, Aobo Wang, and Yuping Duan. Uw-cyclegan: Model-driven cyclegan for underwater image restoration. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–17, 2023.

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

Ning Yang, Qihang Zhong, Kun Li, Runmin Cong, Yao Zhao, and Sam Kwong. A reference-free underwater image quality assessment metric in frequency domain. *Signal Processing: Image Communication*, 94:116218, 2021.

Lin Zhang, Lei Zhang, and Alan C Bovik. A feature-enriched completely blind image quality evaluator. *IEEE Transactions on Image Processing*, 24(8):2579–2591, 2015.

Weixia Zhang, Kede Ma, Jia Yan, Dexiang Deng, and Zhou Wang. Blind image quality assessment using a deep bilinear convolutional neural network. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(1):36–47, 2018.

Weixia Zhang, Guangtao Zhai, Ying Wei, Xiaokang Yang, and Kede Ma. Blind image quality assessment via vision-language correspondence: A multitask learning perspective. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 14071–14081, 2023.

Chen Zhao, Weiling Cai, Chenyu Dong, and Ziqi Zeng. Toward sufficient spatial-frequency interaction for gradient-aware underwater image enhancement. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3220–3224. IEEE, 2024.

Yannan Zheng, Weiling Chen, Rongfu Lin, Tiesong Zhao, and Patrick Le Callet. Uif: An objective quality assessment for underwater image enhancement. *IEEE Transactions on Image Processing*, 31:5456–5468, 2022.

Jingchun Zhou, Qian Liu, Qiuping Jiang, Wenqi Ren, Kin-Man Lam, and Weishi Zhang. Underwater camera: Improving visual perception via adaptive dark pixel prior and color correction. *International Journal of Computer Vision*, pp. 1–19, 2023.

Jingchun Zhou, Jiaming Sun, Chongyi Li, Qiuping Jiang, Man Zhou, Kin-Man Lam, Weishi Zhang, and Xianping Fu. Hclr-net: Hybrid contrastive learning regularization with locally randomized perturbation for underwater image enhancement. *International Journal of Computer Vision*, 132 (10):4132–4156, 2024a.

Jingchun Zhou, Shiyin Wang, Dehuan Zhang, Qiuping Jiang, Kui Jiang, and Yi Lin. Decoupled variational retinex for reconstruction and fusion of underwater shallow depth-of-field image with parallax and moving objects. *Information Fusion*, 111:102494, 2024b.

A Appendix

You may include other additional sections here.