PhISH-Net: Physics Inspired System for High Resolution Underwater Image Enhancement

Aditya Chandrasekar, Manogna Sreenivas, Soma Biswas
Indian Institute of Science, Bangalore
{caditya, manognas, somabiswas}@iisc.ac.in

Abstract

Underwater imaging presents numerous challenges due to refraction, light absorption, and scattering, resulting in color degradation, low contrast, and blurriness. Enhancing underwater images is crucial for high-level computer vision tasks, but existing methods either neglect the physics-based image formation process or require expensive computations. In this paper, we propose an effective framework that combines a physics-based Underwater Image Formation Model (UIFM) with a deep image enhancement approach based on the retinex model. Firstly, we remove backscatter by estimating attenuation coefficients using depth information. Then, we employ a retinex model-based deep image enhancement module to enhance the images. To ensure adherence to the UIFM, we introduce a novel Wideband Attenuation prior. The proposed PhISH-Net framework achieves real-time processing of high-resolution underwater images using a lightweight neural network and a bilateral-grid-based upsampler. Extensive experiments on two underwater image datasets demonstrate the superior performance of our method compared to state-of-the-art techniques. Additionally, qualitative evaluation on a cross-dataset scenario confirms its generalization capability. Our contributions lie in combining the physics-based UIFM with deep image enhancement methods, introducing the wideband attenuation prior, and achieving superior performance and efficiency.

1. Introduction

Underwater imaging is important for various applications such as marine life studies, coral reef monitoring, ocean exploration, and underwater robotics [35]. However, capturing high-quality underwater images is challenging due to factors like refraction, selective light absorption, and scattering. Colors are attenuated differently underwater, with shorter wavelength colors (green and blue) appearing more pronounced due to scattering and longer wavelength colors (red and orange) being absorbed. This effect is amplified at greater depths, resulting in color degradation, low contrast, and blurriness in underwater images. These challenges make it difficult for computer vision tasks like segmentation, object detection, and tracking. To address this, we propose an effective framework that combines deep image enhancement techniques with a physics-based image model that considers light scattering in water.

Early works [1, 28] perform Underwater Image Enhancement (UIE) by correcting contrast, brightness, saturation etc. However, unlike images captured in the air, it is crucial to consider the physical process of underwater image formation when processing these images. Another line of works [2, 3, 52] use Underwater Image Formation Models (UIFM) to improve the visual clarity of underwater images. These models mathematically simulate how light interacts with the underwater environment and how imaging devices capture the images. UIFM-based underwater image enhancement methods often estimate unknown color and light scattering attenuation coefficients to recover clean image signals based on the mathematical model. However, these methods are computationally expensive for real-time applications as they require estimating attenuation coefficients for each image. Recent studies [17, 20, 43] have been inspired by the success of deep neural networks in tasks like image dehazing and low-light image enhancement, and have applied deep networks for underwater image enhancement. These approaches learn the mapping between degraded images and their enhanced counterparts. Additionally, most underwater image datasets consist of high-resolution images [10, 40], but these models often process them at lower resolutions, limiting their performance.

In this work, we propose an effective framework that leverages the formulation of physics based UIFM with retinex model based deep image enhancement framework. We employ the well established UIFM model proposed in Sea-Thru [4], where an underwater image is modeled to have two components (i) Direct signal representing the primary source of image formation and contains the true characteristics of the subject being captured, however with attenuated colors as a function of distance and wavelength;
Figure 1. Sample results for images from the UIEB dataset. **Top:** input; **Bottom:** corresponding enhanced image.

(ii) **Backscatter signal** arising from light scattering by particles suspended in the water. Backscatter tends to be more prominent in underwater imaging due to the higher concentration of suspended particles compared to the air.

Firstly, we propose to estimate and remove backscatter by estimating the attenuation coefficients as proposed in [4]. However, this step requires depth information, which is not available for most underwater image datasets as it is difficult to obtain. To circumvent this, we propose the utilization of an off-the-shelf monocular depth estimation method [48], which we found to be effective. On observing that these backscatter-removed images resemble low-light underexposed images, we propose to use a retinex model-based deep image enhancement module to perform UIE.

Physical model based methods estimate attenuation coefficients for each image individually, while data driven methods ignore the physical image formation process and treat UIE as low light image enhancement problem. In this context, we aim to leverage the effectiveness of both lines of work by proposing a novel **Wideband Attenuation** prior. Through this, while performing image enhancement with PhISH-Net, we ensure that it also adheres to the UIFM model [4], thus considering the characteristics of light scattering in water. We perform most of the computation on low-resolution images and subsequently employ a bilateral grid-based upsampling module to generate the high-resolution enhanced images.

Our study encompasses comprehensive experiments on two underwater image datasets: UIEB [40] and EUVP [35]. In addition, we qualitatively evaluate the cross-dataset performance by testing the model trained on EUVP on another dataset named SQUID [10]. Our method consistently outperforms state-of-the-art (SOTA) UIE methods. Furthermore, the proposed PhISH-Net framework is computationally efficient and lightweight with 0.55M parameters. The key contributions of our work are as follows:

- The proposed PhISH-Net framework combines the physics-based UIFM with a deep image enhancement approach based on the retinex model.
- We propose a novel wideband attenuation prior conditioned on the UIFM while learning PhISH-Net. This takes into account the optical characteristics and color attenuation properties of water.
- Our approach exhibits superior performance compared to SOTA on two underwater image datasets. Further, we demonstrate its effectiveness by qualitatively evaluating its cross-dataset generalization performance.

2. Related Work

In this section, we provide a comprehensive review of the latest advancements in UIE.

**Image Processing based Methods** aim to restore the clarity of underwater images by modifying their appearance based on prior visual characteristics. They employ various image processing techniques, such as contrast adjustment [34], histogram equalization (HE) [49], white balance [47], and fusion-based approaches [7, 8]. Hassan et al. [32] propose to use contrast limited adaptive HE as a pre-processing step for further retinex based enhancement.

**Physical Model based Methods** extend the principles of prior model-based dehazing algorithms to underwater scenes, treating UIE as an inversion problem. The objective is to restore the image by effectively reversing the degradation caused by the underwater imaging process. Various approaches have been developed based on different priors, such as the dark channel prior [13–15, 53, 58], red channel prior [24], attenuation curve prior [66], backscatter pixel prior [75] and minimum information prior [42]. Notable model-based methods include works by Akkaynak et al. [3], Berman et al. [10], Pei et al. [52], Xie et al. [68], HLRP [76], MLLE [74], and ICSP [33].

**Data Driven Methods** with deep learning have recently gained prominence in UIE. CNN based methods include WaterNet [40], UWCNN [38], SGUIE-Net [56], UICoE-
Net [57] and UIE-Net [64]. GAN based approaches include UGAN [19], UW-GAN [60], Spiral-GAN [31], FunIE-GEN [35], WaterGAN [44], Dense GAN [27], FE-GAN [29], TOPAL [36], and CycleGAN [41]. However, such methods perform poorly on real-world images due to their reliance on large synthetic datasets for training as they fail to take into account the domain shift problem between synthetic and real-world data [67]. They also disregard the physical process of image formation under water.

3. Methodology

Here, we present our two main components: (i) UIFM, which is used for backscatter estimation; (ii) PhISH-Net, the proposed deep image enhancement framework to enhance the backscatter-removed image.

3.1. Underwater Image Formation Model

In this work, we employ the UIFM proposed in SeaThru [4]. According to this model, the formation of any image \( I \) captured underwater can be modeled using two components, the direct signal \( D \) and backscatter \( B \) [5]. For each channel \( c \in \{r, g, b\} \), the image can be decomposed as

\[
I_c = D_c + B_c
\]

\( D \) and \( B \) are governed by two coefficients, the wideband (RGB) attenuation \( \beta_d \) and the backscatter coefficient \( \beta_b \). Equation 1 is further expanded as follows:

\[
I = J e^{-\beta_d z} + B^\infty (1 - e^{-\beta_b z})
\]

Here, \( J \) is the unattenuated scene that would have been captured if there was no water medium attenuating colors with depth \( z \). \( B^\infty \) is the backscatter color. In this model, the coefficients \( \beta_d \) and \( \beta_b \) are considered to be distinct. \( \beta_b \) and \( \beta_d \) depend on several factors including object reflectance, spectrum of ambient light, spectral response of the camera, the physical scattering and beam attenuation coefficients of the water body, all of which are functions of wavelength. Now, we describe the algorithm to estimate backscatter \( B \), which is then used to obtain the direct signal \( D \) from the captured underwater image \( I \).

3.2. Backscatter Estimation

When light penetrates through the water surface and interacts with underwater objects, it undergoes scattering in various directions. Additionally, suspended particles in the water column contribute to the scattering phenomenon towards the surface, resulting in a form of noise known as backscatter. This is modelled as an additive signal, which degrades the quality of underwater images.

The amount of backscatter in an image increases exponentially with the depth \( z \) and eventually reaches saturation. In areas where the scene reflectance or illumination tends to zero, the captured image \( I \rightarrow B \). Backscatter can hence be estimated by identifying such areas. Inspired by the dark channel prior, this is done by searching for the darkest RGB triplets based on the depth map. The depth map is first partitioned into 10 evenly spaced clusters. In each cluster, we search \( I \) for RGB triplets in the bottom 1%, the set of which we denote as \( \Omega \). An overestimate of backscatter is \( \hat{B}(\Omega) \approx I(\Omega) \), which can be modelled using Equation 2 as

\[
\hat{B} = J' \cdot e^{-\beta_d' z} + B^\infty (1 - e^{-\beta_b' z})
\]

Residual term

Given an input image \( I \) and its corresponding depth map \( z \), the coefficients \( B^\infty, \beta_b, \beta_d' \) and \( J' \) are estimated using a non-linear least squares fitting subject to bounds as described in [4]. This is summarized in Algorithm 1.

Depth Estimation

To accurately estimate backscatter, an absolute depth map is required. While some datasets utilize structure-from-motion [4] with known objects for scale to obtain these depth maps, this approach is limited in datasets without pre-existing depth information. To overcome this challenge, we employ an off-the-shelf depth estimation network to generate depth maps. Specifically, we adopt a boosted version of monocular depth estimation proposed by [48], which builds upon MiDaS [37] without retraining the base model. This depth boosting framework enhances the depth map quality by iteratively incorporating scene structure information using double estimation and patch selection at different resolutions. We demonstrate the quality of depth images with and without boosting in the supplementary material. We further employ a heuristic to estimate the minimum and maximum depth values in the absence of explicit information.
Algorithm 1 Backscatter Estimation

Input: Underwater image $I$, Depth map $z$.
1: Initialization: $\Omega = \emptyset$, $D = \emptyset$.
2: Partition $z$ in 10 evenly spaced clusters $\Psi$ between $\max(z)$ and $\min(z)$.
3: for cluster in $\Psi$ do
4: Add RGB triplets of $I$ in the bottom 1% to $\Omega$.
5: end for
6: An overestimate of the backscatter is given by $\hat{B}(\Omega) \approx I(\Omega)$, which can be modelled using Eq. (2).
7: for $c \in \{r, g, b\}$ do
8: Estimate $B_c$ using a non-linear least squares fit on Eq. (3) and then obtain the direct signal.
9: $D_c = I_c - \hat{B}_c$.
10: end for
Output: Estimated backscatter $\hat{B}$ and Direct signal $D$.

3.3. PhISH-Net

The direct signal estimated using the backscatter is often dimly lit. The task of enhancing this direct signal mimics the well explored low light image enhancement problem, which we draw inspiration from. At its core, image enhancement involves finding a mapping that enhances an input image $I$ to produce an improved version $\tilde{I}$.

A well established image enhancement model is the retinex model, where an image captured $I$ is composed of two components, Reflectance $\hat{I}$ and Illumination $S$. Mathematically, this is given as $I = S \ast \hat{I}$, where $\ast$ denotes a pixel-wise multiplication. In our case, $I = D$, the direct signal obtained after backscatter removal. As in several prior works [21, 26], we treat $I$ as the enhanced image. Now, given $I$, the problem reduces to estimating the illumination map $S$, using which the enhanced image $\tilde{I}$ can be obtained as $\tilde{S}^{-1} \ast \hat{I}$. Although early works such as LIME [26] and DUAL [73] model $S$ as a single-channel illumination map, recent studies [61] have shown that representing the illumination map using three channels yields better performance, especially in terms of colour enhancement and handling non-linearity across different colour channels.

In this work, we propose PhISH-Net to predict a three-channel illumination map $S$, which is then used to obtain the enhanced output image as $\tilde{I} = I / (S + \epsilon)$.

We propose to use a simple convolutional architecture equipped with a network designed to extract both global and local features, aiding in the estimation of bilateral grid coefficients [25]. These coefficients are then applied to the output of a guide network, to obtain a high-resolution illumination map $S_{hr}$. Subsequently, the enhanced image is computed as $I_{out} = D_{hr} / (S_{hr} + \epsilon)$. This allows most network computations to be performed at a reduced resolution, facilitating real-time processing of high-resolution underwater images. This approach not only addresses computational efficiency but also enables the model to handle images of varying sizes. The overall pipeline for this image enhancement process is illustrated in Figure 2. We impose a novel prior on the wideband (RGB) attenuation coefficient $\beta_d$ based on the experimental findings in [4]. PhISH-Net is optimized to predict the illumination map $S_{hr}$ by minimizing the proposed attenuation loss, image reconstruction loss, color loss and smoothness loss. We first describe the proposed wideband attenuation prior and then describe each of the loss components.

Wideband Attenuation Prior The objective is to recover the unattenuated scene $J$ from the estimated direct signal $D$, following the UIFM model described in Section 3.1. Considering the visual properties of the estimated direct signal mimicking that of a low light image, the objective can alternatively be formulated based on retinex based image model to recover the reflectance component by estimating the illumination map. Sea-Thru draws inspiration from image-dehazing models, and [23] suggest that there is a linear relationship between image-dehazing and the retinex model. This in turn suggests that the objective of estimating the unattenuated scene $J$ based on UIFM can be related to that of estimating reflectance based on Retinex model. Here, we propose a novel wideband attenuation prior through attenuation loss $L_a$, which constrains the predicted output to follow both the image formation models. The attenuation component $e^{-\beta_dz}$ being analogous to the illumination map $S$, we can obtain a coarse estimate of the wideband attenuation coefficient, given the depth map $z$ and the predicted illumination map $S_{hr}$ as follows.

$$\hat{\beta}_d(z) = -\frac{\log S_{hr}}{z}$$

This coefficient is known to decay with depth $z$ following a two-term exponential curve as shown below. This was established in [3] by performing extensive experimentation on a variety of underwater image datasets.

$$\beta_d(z) = a * e^{-b z} + c * e^{-d z}$$

PhISH-Net predicts the illumination map $S_{hr}$ which we use to constrain the coarse estimate of the coefficient $\beta_d$ obtained from Equation 4 to follow the curve in Equation 5 by using the following wideband attenuation prior loss $L_a$.

$$L_a = \left\| \left( -\frac{\log S_{hr}}{z} \right) - (a * e^{-b z} + c * e^{-d z}) \right\|^2$$

Here, $a, b, c, d \in \mathbb{R}^+$ are obtained from a learnable coefficient efficient predictor network using features extracted by the encoder block. We denote these coefficients by the vector $\mathcal{V} = [a, b, c, d]$ in Figure 2.
Reconstruction loss is obtained as the $L_2$ norm between the predicted enhanced image and the ground truth enhanced image $I_{gt}$. Here, $c$ is the channel and $c \in \{r, g, b\}$.

\[
L_r = \| (S_{hr}^{-1} \ast D_{hr}) - I_{gt} \|^2 \text{ s.t. } (D_{hr})_c \leq (S_{hr})_c \leq 1 \tag{7}
\]

Color Loss is employed to match the colors of the predicted output $I_{out}$ with the ground truth image $I_{gt}$. This is achieved by minimizing the angle $\angle(\cdot, \cdot)$ between each color pair for all pixels $p$. Here, each RGB color is considered as a vector in three dimensional space.

\[
L_c = \sum_p \angle((I_{out})_p, (I_{gt})_p) \tag{8}
\]

Smoothness Loss is a weighted $L_2$ norm of the gradients of the illumination map $S$. This is imposed based on the prior observation that illumination in natural images is typically locally smooth.

\[
L_s = \sum_p \sum_c \omega_{p,c} \left( \frac{\partial S}{\partial p} S_c \right)^2 + \omega_{p,c} \left( \frac{\partial p}{\partial S} S_c \right)^2 \tag{9}
\]

The weights $\omega_{p,c}$ and $\omega_{p,c}$ are spatially varying smoothness weights designed to promote locally smooth variations on pixels with small gradients while allowing for discontinuous illumination on pixels with large gradients in the illumination map. In the context of underwater images, which are prone to inconsistent illumination, the smoothness loss helps mitigate the introduction of abrupt changes in the illumination map.

Training PhISH-Net Given an underwater image dataset $D = \{(I^i, I_{gt}^i)\}_{i=1}^N$, where $I_{gt}^i$ is the enhanced version of the image $I^i$ and $N$ is the number of images, we train PhISH-Net using the following loss:

\[
L^i = w_r L_r^i + w_c L_c^i + w_s L_s^i + w_a L_a^i \tag{10}
\]

where $L^i$ corresponds to the loss for the $i$th image pair and $w_r, w_c, w_s$ and $w_a$ are the weights assigned to each loss component. The values are empirically set to $w_r = 10, w_s = 2, w_c = 1$ and $w_a = 0.5$.

3.4. Photofinishing

Color correction is a crucial step in enhancing underwater images. To eliminate undesirable color casts, images are first white balanced, which improves their overall appearance. Previous work [8] has shown that artifacts are typically present in the red and blue channels, and are corrected.

\[
I_{rc} = I_r + \alpha_1 \ast (I_g - \bar{I}_r) \quad I_{bc} = I_b + \alpha_2 \ast (I_g - \bar{I}_b) \tag{11}
\]

where $I_{rc}$ and $I_{bc}$ are the corrected red ($I_r$) and blue ($I_b$) channels respectively. $\bar{I}_c$ denotes the mean value of channel $c \in \{r, g, b\}$. The combining coefficients $\alpha_1$ and $\alpha_2$ are and are chosen between $[0, 1]$.

The final output is obtained through a multi-scale fusion technique based on a laplacian pyramid [11]. First, a gamma-corrected and unsharp masked version of the white-balanced image are generated to enhance global colour contrast and edge sharpness. The two versions are weighted using local image quality metrics such as the laplacian contrast weight, saliency weight, and saturation weight. The weights are chosen to preserve the desired qualities and reject undesired defaults of the inputs. The output of this step is convexly combined with the model’s output to allow for a more natural and visually appealing final image.

4. Experimental Results

Datasets In order to comprehensively assess the performance of our proposed approach, we conducted extensive experiments on two real-world underwater datasets: UIEB [40], EUVP [35]. The UIEB dataset consists of 890 underwater images with diverse scenes and varying levels of degradation. The EUVP dataset contains 5550 paired dark
images captured during oceanic explorations. We use 80:20 train-test split for these two datasets in all experiments. Further, we also use the SQUID [10] dataset which comprises 57 underwater images from different locations, representing various water properties and conditions. Due to the limited number of available images in SQUID dataset, here we evaluate the model trained on EUVP and qualitatively compare our method with prior approaches in Figure 3. This experiment also demonstrates the generalization ability of this method across datasets.

4.1. Implementation Details

**Backscatter Estimation** We use an off-the-shelf depth boosting method [48] based on MiDaS [37], a monocular depth estimation method, without further retraining. The obtained depth map is used for backscatter estimation on the full-resolution underwater images.

**PhISH-Net** The model is built using PyTorch [51] and trained for 100 epochs with a batch size of 64 on an NVIDIA RTX A5000 GPU using the loss defined in Equation 10 and the Adam optimizer with a learning rate of 1e−4. During training, the input images are resized to 512 × 512 pixels, and a fixed size of 256 × 256 pixels is used for the low-resolution input. An intermediate representation is obtained using a series of five convolutional blocks, which is then fed into the local and global feature extractor. The resulting local and global features are concatenated and upsampled using a bilateral grid-based module [25]. The architecture of each module is described in detail in the supplementary material. Given that the guide network is fully convolutional, it can handle images of any resolution at test time, enabling us to operate at the original resolution.

**Evaluation Metrics** To quantitatively evaluate the effectiveness of our proposed model, we employ a comprehensive set of both reference-based and non-reference-based quality metrics. The reference based metrics include Peak Signal-to-Noise Ratio (PSNR), PSNR for the Luminance channel (PSNR-L), Structural Similarity Index Measure (SSIM), and Patch-based Contrast Quality Index (PCQI) [62]. These metrics provide a comparative assessment by comparing the enhanced images with reference images. In addition, we also evaluate the model based on five non-reference-based metrics which evaluate the quality of enhanced images without relying on reference images. These include Underwater Colour Image Quality Evaluation (UCIQE) [71], Colorfulness Contrast Fog density index (CCF) [65], Underwater Image Quality Measure (UIQM), Underwater Image Colorfulness Measure (UICM) and Underwater Image Contrast Measure (UIConM) [50]. It is important to note that higher values indicate better results and improved image quality for all these metrics.

4.2. Additional Analysis

In this section, we thoroughly evaluate and discuss the performance of the proposed framework. To provide a comprehensive assessment, we benchmark our approach against several existing UIE methods, enabling a comparative analysis. Moving beyond quantitative evaluations, we delve into qualitative results and present visual comparisons with previous methods, allowing for a deeper understanding of the improvements achieved. Furthermore, we conduct an ablation study to explore the individual contributions of each loss component in our framework. These discussions shed light on the importance of each component of our approach.

**Comparative Analysis** We present a thorough comparison of the proposed framework with various UIE methods on two datasets, UIEB and EUVP, including both traditional/model-based approaches [4, 14, 33, 54, 76] and deep learning-based methods [19, 30, 35, 36, 39, 60].
<table>
<thead>
<tr>
<th>Method</th>
<th>PSNR</th>
<th>PSNR_{L}</th>
<th>SSIM</th>
<th>PCQI</th>
<th>UIQF</th>
<th>UIQM</th>
<th>UICM</th>
<th>UICM</th>
<th>UICM</th>
<th>CCF</th>
</tr>
</thead>
<tbody>
<tr>
<td>FU-IE-GAN [35]</td>
<td>17.382</td>
<td>20.0596</td>
<td>0.7285</td>
<td>0.6393</td>
<td>0.5459</td>
<td>1.3993</td>
<td>5.7775</td>
<td>1.1605</td>
<td>21.0605</td>
<td></td>
</tr>
<tr>
<td>UG-GAN [60]</td>
<td>16.228</td>
<td>19.1524</td>
<td>0.7644</td>
<td>0.7187</td>
<td>0.5655</td>
<td>1.3467</td>
<td>5.5550</td>
<td>1.1219</td>
<td>22.8449</td>
<td></td>
</tr>
<tr>
<td>UWCNN [39]</td>
<td>12.024</td>
<td>13.7820</td>
<td>0.6469</td>
<td>0.3921</td>
<td>0.5058</td>
<td>1.0647</td>
<td>1.9216</td>
<td>0.9398</td>
<td>11.0019</td>
<td></td>
</tr>
<tr>
<td>HLRP [76]</td>
<td>13.031</td>
<td>13.7891</td>
<td>0.2874</td>
<td>0.2326</td>
<td>0.6356</td>
<td>1.6477</td>
<td>9.5089</td>
<td>1.1915</td>
<td>36.9312</td>
<td></td>
</tr>
<tr>
<td>MLLE [74]</td>
<td>18.107</td>
<td>19.4733</td>
<td>0.7985</td>
<td>0.9105</td>
<td>0.6044</td>
<td>1.6287</td>
<td>4.8271</td>
<td>1.0124</td>
<td>36.2740</td>
<td></td>
</tr>
<tr>
<td>IBIA [54]</td>
<td>15.561</td>
<td>17.7756</td>
<td>0.7390</td>
<td>0.6949</td>
<td>0.6020</td>
<td>1.4397</td>
<td>7.3312</td>
<td>1.0143</td>
<td>28.9752</td>
<td></td>
</tr>
<tr>
<td>TOPAL [36]</td>
<td>20.587</td>
<td>22.6706</td>
<td>0.8674</td>
<td>0.7148</td>
<td>0.5841</td>
<td>1.2209</td>
<td>5.0093</td>
<td>1.0321</td>
<td>21.2175</td>
<td></td>
</tr>
<tr>
<td>UDCP [14]</td>
<td>11.928</td>
<td>12.7288</td>
<td>0.6441</td>
<td>0.6117</td>
<td>0.5956</td>
<td>1.6477</td>
<td>9.5089</td>
<td>1.1915</td>
<td>36.9312</td>
<td></td>
</tr>
<tr>
<td>Water-Net [40]</td>
<td>18.700</td>
<td>19.7733</td>
<td>0.8674</td>
<td>0.6949</td>
<td>0.6020</td>
<td>1.4397</td>
<td>7.3312</td>
<td>1.0143</td>
<td>28.9752</td>
<td></td>
</tr>
<tr>
<td>MLLE [74]</td>
<td>18.107</td>
<td>19.4733</td>
<td>0.7985</td>
<td>0.9105</td>
<td>0.6044</td>
<td>1.6287</td>
<td>4.8271</td>
<td>1.0124</td>
<td>36.2740</td>
<td></td>
</tr>
<tr>
<td>IBIA [54]</td>
<td>15.561</td>
<td>17.7756</td>
<td>0.7390</td>
<td>0.6949</td>
<td>0.6020</td>
<td>1.4397</td>
<td>7.3312</td>
<td>1.0143</td>
<td>28.9752</td>
<td></td>
</tr>
<tr>
<td>TOPAL [36]</td>
<td>20.587</td>
<td>22.6706</td>
<td>0.8674</td>
<td>0.7148</td>
<td>0.5841</td>
<td>1.2209</td>
<td>5.0093</td>
<td>1.0321</td>
<td>21.2175</td>
<td></td>
</tr>
<tr>
<td>UDCP [14]</td>
<td>11.928</td>
<td>12.7288</td>
<td>0.6441</td>
<td>0.6117</td>
<td>0.5956</td>
<td>1.6477</td>
<td>9.5089</td>
<td>1.1915</td>
<td>36.9312</td>
<td></td>
</tr>
<tr>
<td>Water-Net [40]</td>
<td>18.700</td>
<td>19.7733</td>
<td>0.8674</td>
<td>0.6949</td>
<td>0.6020</td>
<td>1.4397</td>
<td>7.3312</td>
<td>1.0143</td>
<td>28.9752</td>
<td></td>
</tr>
<tr>
<td>MLLE [74]</td>
<td>18.107</td>
<td>19.4733</td>
<td>0.7985</td>
<td>0.9105</td>
<td>0.6044</td>
<td>1.6287</td>
<td>4.8271</td>
<td>1.0124</td>
<td>36.2740</td>
<td></td>
</tr>
<tr>
<td>IBIA [54]</td>
<td>15.561</td>
<td>17.7756</td>
<td>0.7390</td>
<td>0.6949</td>
<td>0.6020</td>
<td>1.4397</td>
<td>7.3312</td>
<td>1.0143</td>
<td>28.9752</td>
<td></td>
</tr>
<tr>
<td>TOPAL [36]</td>
<td>20.587</td>
<td>22.6706</td>
<td>0.8674</td>
<td>0.7148</td>
<td>0.5841</td>
<td>1.2209</td>
<td>5.0093</td>
<td>1.0321</td>
<td>21.2175</td>
<td></td>
</tr>
<tr>
<td>UDCP [14]</td>
<td>11.928</td>
<td>12.7288</td>
<td>0.6441</td>
<td>0.6117</td>
<td>0.5956</td>
<td>1.6477</td>
<td>9.5089</td>
<td>1.1915</td>
<td>36.9312</td>
<td></td>
</tr>
<tr>
<td>Water-Net [40]</td>
<td>18.700</td>
<td>19.7733</td>
<td>0.8674</td>
<td>0.6949</td>
<td>0.6020</td>
<td>1.4397</td>
<td>7.3312</td>
<td>1.0143</td>
<td>28.9752</td>
<td></td>
</tr>
<tr>
<td>MLLE [74]</td>
<td>18.107</td>
<td>19.4733</td>
<td>0.7985</td>
<td>0.9105</td>
<td>0.6044</td>
<td>1.6287</td>
<td>4.8271</td>
<td>1.0124</td>
<td>36.2740</td>
<td></td>
</tr>
<tr>
<td>IBIA [54]</td>
<td>15.561</td>
<td>17.7756</td>
<td>0.7390</td>
<td>0.6949</td>
<td>0.6020</td>
<td>1.4397</td>
<td>7.3312</td>
<td>1.0143</td>
<td>28.9752</td>
<td></td>
</tr>
<tr>
<td>TOPAL [36]</td>
<td>20.587</td>
<td>22.6706</td>
<td>0.8674</td>
<td>0.7148</td>
<td>0.5841</td>
<td>1.2209</td>
<td>5.0093</td>
<td>1.0321</td>
<td>21.2175</td>
<td></td>
</tr>
<tr>
<td>UDCP [14]</td>
<td>11.928</td>
<td>12.7288</td>
<td>0.6441</td>
<td>0.6117</td>
<td>0.5956</td>
<td>1.6477</td>
<td>9.5089</td>
<td>1.1915</td>
<td>36.9312</td>
<td></td>
</tr>
<tr>
<td>Water-Net [40]</td>
<td>18.700</td>
<td>19.7733</td>
<td>0.8674</td>
<td>0.6949</td>
<td>0.6020</td>
<td>1.4397</td>
<td>7.3312</td>
<td>1.0143</td>
<td>28.9752</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Image Quality Metrics of various UIE methods on the UIEB and EUVP datasets.

**Figure 4.** Stage-wise outputs of the proposed framework.

results, summarized in Table 1, clearly demonstrate the superiority of our approach over all previous methods on both datasets. While our model might not secure the highest ranking in certain non-reference-based metrics, it’s noteworthy to observe that techniques achieving high non-reference metrics tend to yield low reference-based metrics, particularly in terms of SSIM and PCQI. This topic has been elaborated upon in the supplementary material, alongside enhancements in metrics attributed to the photofinishing step. Additionally, stage-wise results are shown in Figure 4. Overall, these quantitative evaluations establish the superiority of our proposed framework over previous works.

**Qualitative results** We further demonstrate the effectiveness of our framework by showcasing sample results from two datasets: UIEB and EUVP (see Figures 1 and 6 respectively). Additionally, we evaluate the generalization

<table>
<thead>
<tr>
<th>L_{c}</th>
<th>L_{s}</th>
<th>PSNR (↑)</th>
<th>SSIM (↑)</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>✓</td>
<td>20.7987</td>
<td>0.8299</td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>20.8560</td>
<td>0.8319</td>
</tr>
<tr>
<td>✓ ✓</td>
<td>✓</td>
<td>20.8671</td>
<td>0.8320</td>
</tr>
<tr>
<td>✓ ✓ ✓</td>
<td>✓</td>
<td>21.1390</td>
<td>0.8686</td>
</tr>
</tbody>
</table>

Table 2. Ablation study of loss components on UIEB dataset
Ablation Study To assess the significance of each loss component in our proposed approach, we conducted a comprehensive ablation study. The results, presented in Figure 5, highlight the consistent improvement in model performance with the addition of each loss term. Additionally, we analyze the impact of each extra component on quality metrics, as summarized in Table 2. We observe the PSNR and SSIM metrics improve significantly as each individual loss term is incorporated, indicating superior image quality. Overall, the ablation study provides valuable insights into the efficacy of our proposed loss. Additionally, we conducted ablation studies by altering the intrinsic parameters of the model and evaluating its robustness against image and depth degradations. Further details are provided in the supplementary material.

Complexity Analysis We evaluate our method’s complexity in FLOPs, parameters, and runtime (Table 3). We observe that our model achieves a significant improvement in run-time, even for high-resolution underwater images. In terms of parameters and FLOPs, the proposed model is extremely lightweight, at only 0.55M parameters.

5. Conclusion In this paper, we have presented an effective framework for underwater image enhancement that combines a physics-based UIFM with a deep image enhancement approach based on the retinex model. Our proposed PhISH-Net framework achieves real-time processing of high-resolution underwater images using a lightweight neural network and a bilateral-grid-based upsampler. By introducing a novel Wideband Attenuation prior, we account for the optical characteristics and color attenuation properties of water while also following the retinex model. Extensive experiments on two underwater image datasets demonstrate the superior performance of our method compared to state-of-the-art techniques. Additionally, our approach showcases cross-dataset generalization capability. Overall, our contributions lie in the integration of physics-based modeling, deep image enhancement, and efficient processing, offering a compelling solution for the underwater image enhancement task.
References


[60] Nan Wang, Yabin Zhou, Fenglei Han, Haitao Zhu, and Jingzheng Yao. Uwgan: Underwater gan for real-world underwater color restoration and dehazing. 2019. 3, 6, 7, 8


