IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 19 December 2024, accepted 20 January 2025, date of publication 29 January 2025, date of current version 4 February 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3536173

RESEARCH ARTICLE

Joint Optimization in Underwater Image Enhancement: A Training Framework Integrating Pixel-Level and Physical-Channel Techniques

OZAN DEMİR[1,2], METIN AKTAS[1], AND ENDER M. EKSIOGLU[2], (Member, IEEE)
[1]Defense Systems Technologies Division, Aselsan Inc., 06200 Ankara, Türkiye
[2]Department of Electronics and Communication Engineering, Istanbul Technical University, 34469 Istanbul, Türkiye

Corresponding author: Ozan Demir (ozandemir@aselsan.com.tr)

This work was supported by Aselsan Inc.

ABSTRACT In recent years, with the increasing interest in marine research, the need to collect and process clear underwater optical images has become crucial. However, underwater images suffer from the absorption and scattering effects of the environment. In this paper, we propose Hybrid Underwater Image Enhancement Network (HUWIE-Net), a novel deep learning-based underwater image enhancement framework consisting of three distinct sections, which include an Image-to-Image Module, a Physics-Informed Module and a Fusion Module. The training methodology of HUWIE-Net is designed to jointly optimize both pixel-level-based and physical-channel-based enhancement modules. In this framework, while Image-to-Image Module is used for color correction in pixel level, Physics-Informed Module is used for dehazing by exploiting the underwater image formation model which defines the deformations in the underwater light propagation channel. We also propose to use the joint loss function for both Image-to-Image Module and Physics-Informed Module to enforce the joint optimization for better underwater image enhancement performance. The results of experiments conducted with real-world underwater images show that the proposed model achieves improved performance compared to state-of-the-art methods. The code for the newly developed HUWIE-Net is available at https://github.com/UIE-Lab/HUWIE-Net.

INDEX TERMS Underwater image enhancement, deep learning, underwater image formation model, dark channel prior, physics-informed deep network, joint optimization.

I. INTRODUCTION

Interest in underwater world has increased exponentially in recent years for various reasons, especially marine resource research, military applications and ecological studies with advances in robotics and imaging technologies. With this increasing interest, the need for collecting and processing clear underwater optical images becomes crucial. However, underwater images suffer from various disturbing factors, primarily the absorption and scattering of light [1].

Absorption is the process by which light energy is captured by the medium it passes through. This phenomenon is wavelength-dependent, and longer wavelengths such as red and orange is absorbed more rapidly than shorter wavelengths such as blue and green. Wavelength-dependent light absorption causes issues in underwater imaging, including color distortion, contrast loss, and brightness reduction. Due to the increasing absorption at longer wavelengths, captured images tend to appear bluish or greenish as the water depth increases. On the other hand, scattering is defined as a change in the direction of light caused by its interaction with particles suspended in water. In underwater imaging, scattering is considered a twofold phenomenon: first, the

The associate editor coordinating the review of this manuscript and approving it for publication was Yougan Chen.

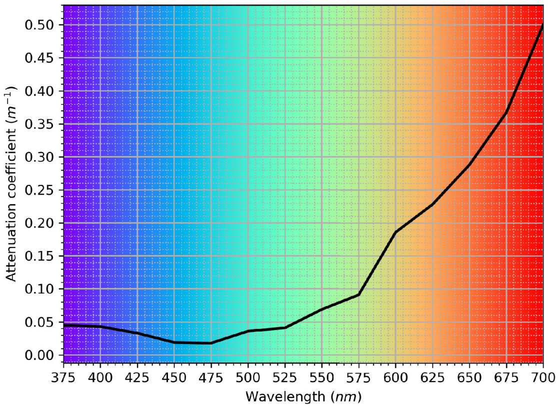


FIGURE 1. The wavelength-dependent attenuation coefficient (p_λ) for pure water [3].

scattering of light from the scene before it reaches the camera; second, the scattering of light by suspended particles in the medium, which reaches the camera and produces undesired effects. Scattering causes contrast loss, a foggy appearance, blurriness, and a loss of details in images. Alongside these considerations since the light intensity decreases as the depth increases, artificial light sources are needed. Although this artificial light increases the visible distance, it makes scattering caused by particles suspended in water a more serious problem [2]. In addition, the problem becomes even more complicated because the diversity of water types can also cause differences in their effects on the absorption and scattering of light [3].

The mentioned disturbing factors cause color distortion, haze, blurring, and decreased contrast and brightness in underwater images. To solve these problems in underwater images, conventional computer vision methods [5], [13], physics-based methods [7], [8], [9], [10], [12], hybrid methods [4], [6], and deep learning-based methods [1], [14], [15], [16], [17], [18], [19], [20] have been developed.

The remaining of the paper is organized as follows. Section II discusses studies in the literature on underwater image enhancement. Section III describes underwater image formation, underwater dark channel prior and proposed method. Section IV provides details of experimental results, visual outputs, and metric measurements. The paper is concluded with Section V.

II. RELATED WORK

Several conventional computer vision based methods and deep learning based methods are proposed for solving underwater image enhancement problem. Conventional methods are generally the application of in-air image enhancement methods on underwater images. On the other side, as in many areas, deep learning approaches with higher representation levels consisting of non-linear modules have become a milestone in the field of underwater image enhancement. It can be stated that deep learning-based approaches stand out due to their improved generalization performance compared to conventional methods.

A prior-based method for haze removal was adapted for underwater use by [10] as an underwater version of the earlier work in [11]. Following this, the Wavelength Compensation and Image Dehazing (WCID) algorithm, which performs color correction and dehazing, was introduced in [9]. Subsequently, a contrast-limited adaptive histogram equalization-based method called mixture Contrast Limited Adaptive Histogram Equalization (CLAHE-mix) was proposed in [5]. In [8], a technique was developed that includes dehazing of blue and green channels along with color correction for the red channel. Around the same period, [7] proposed color correction and visibility restoration algorithms; however, the study noted a limitation where the global background estimation algorithm required input images to contain background areas. Further advancements were seen in [13], where a method was proposed combining white balance, gamma correction, and sharpening operations to enhance underwater images. Later, a physics-based method for estimating transmission maps via light attenuation prior was introduced in [12]. In [4], an underwater image dehazing and contrast enhancement algorithm was proposed. It was noted that the proposed method could not sufficiently improve colors in low-light conditions, as it only considers the distance between the object and the camera, ignoring the distance of object to the water surface. Later, [6] introduced a hybrid approach that combined a color correction method with a learning-based dehazing technique.

The development of deep learning-based approaches for underwater image enhancement began with Underwater Image Enhancement Network (UIE-Net), a Convolutional Neural Network (CNN) based model proposed in [18], which consists of two sub-modules: Color Correction Network (CC-Net), outputting color correction coefficients, and Haze Removal Network (HR-Net), enhancing contrast. Building on these early advances, [16] introduced a Pixel-to-Pixel CNN model with an encoder-decoder architecture. Similarly, [15] proposed Underwater Image Restoration Network (UIR-Net), which includes two independent networks: Transmission Map Network (TM-Net) for estimating the transmission map and Background Light Network (BL-Net) for determining background light, both taking the underwater image formation model as a reference. Subsequently, [19] introduced Water-Net, a model that processes a raw image along with three images derived from histogram equalization, gamma correction, and white balance transformations. Water-Net outputs an enhanced image by performing element-wise multiplication between the learned confidence maps and these processed outputs. Further advancements were seen in [17], where a model incorporating a nuisance classifier, designed to handle various water types, was proposed alongside encoder-decoder modules. In [14], a CNN-based image-to-image deep learning model was introduced, trained with synthetic data to improve underwater images. Innovation continued with Underwater Image Enhancement Convolution

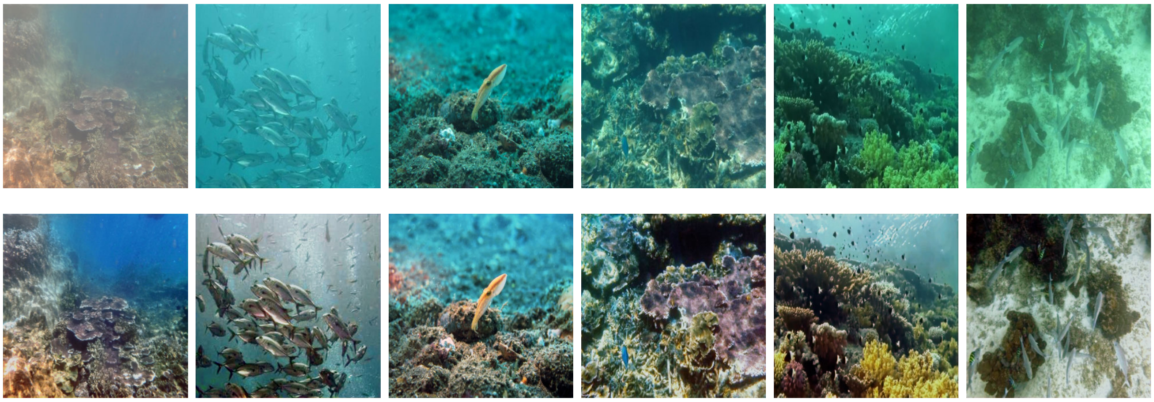


FIGURE 2. Top row: real-world underwater images [19], bottom row: the corresponding enhanced images by HUWIE-Net.

Neural Network using 2 Color Space (UIEC^2-Net) [1], which combines an RGB pixel-level block for denoising and color correction, an HSV global adjustment block for brightness and saturation enhancement, and an attention map block that integrates the effects of the first two. Recently, [20] presented Zero-UIE, a model that estimates curve parameters directly without needing reference images; these parameters, along with background light, are integrated into the underwater image formation model to produce an enhanced output image.

In this paper, a deep learning-based underwater image enhancement method has been proposed. Figure-2 shows samples of real-world underwater image enhancement using the proposed HUWIE-Net. Our main contributions can be summarized as follows. We propose a training framework that jointly optimizes pixel-level-based and physical-channel-based underwater image enhancement methods. In this framework, while Image-to-Image Module (I2IM) is used for color correction in pixel level, Physics-Informed Module (PIM) is used for dehazing by exploiting the underwater image formation model which defines the deformations in the underwater light propagation channel. In PIM the parameters of the underwater image formation model are estimated using the deep network of the PIM and the Underwater Dark Channel Prior [10]. To the best of our knowledge, PIM is the first deep learning model that estimates Underwater Image Formation Model (UIFM) parameters by using Underwater Dark Channel Prior. We propose to use joint loss function for both I2IM and PIM to enforce the joint optimization for better underwater image enhancement performance. HUWIE-Net is shown to have better generalization performance and results compared to state-of-the-art methods by experimental studies.

III. METHODOLOGY

In this section, the Underwater Image Formation Model, which forms the basis of the PIM, is first explained, followed by the Underwater Dark Channel Prior. Finally, detailed information about the proposed model is provided.

A. UNDERWATER IMAGE FORMATION MODEL

Underwater image formation models take into account the propagation of light in underwater. Developing these models is challenging because the absorbing and scattering properties of the transmission medium between the light source, camera and scene are complex. According to the Jaffe-McGlamery model [21], [29], which is a widely used model for underwater image formation, it is stated that the underwater image can be expressed as the sum of three components, i.e., the direct component E_d, the forward scattering component E_f and the backscattering component E_b. E_d is the light reflected from the scene and reaches the camera directly. E_f is the light reflected from the scene but scattered on the way to the camera. E_b is the light that reaches the camera by reflecting directly from particles suspended in the water [2], [22]. Due to image deformations dominated by backscattering [2], [10], [31], [32] and the generally close distances between the camera and the scene [23], the forward scattering component can be neglected in the underwater image formation model. Therefore, the total irradiance energy is expressed as the sum of the direct component and the backscattering component as given in Equation-1.

$$E_T = E_d + E_b. \tag{1}$$

The total irradiance energy E_T, as given in Equation-1, depends on E_d and E_b. E_d is the direct component reflected from the scene and captured by the camera. As the underwater path length of E_d increases (i.e., as the distance between the camera and the scene increases) and as the water absorption and scattering coefficients (a_λ and b_λ) increase, E_d weakens. If E_d becomes insufficient for detection by the sensors, the resulting images are classified as power-limited [21]. On the other hand, E_b is the backscattering component caused by light reflected from suspended particles within the water volume inside the camera's field of view. Increased scattering effects raise the energy of E_b, resulting in contrast-limited images [21].

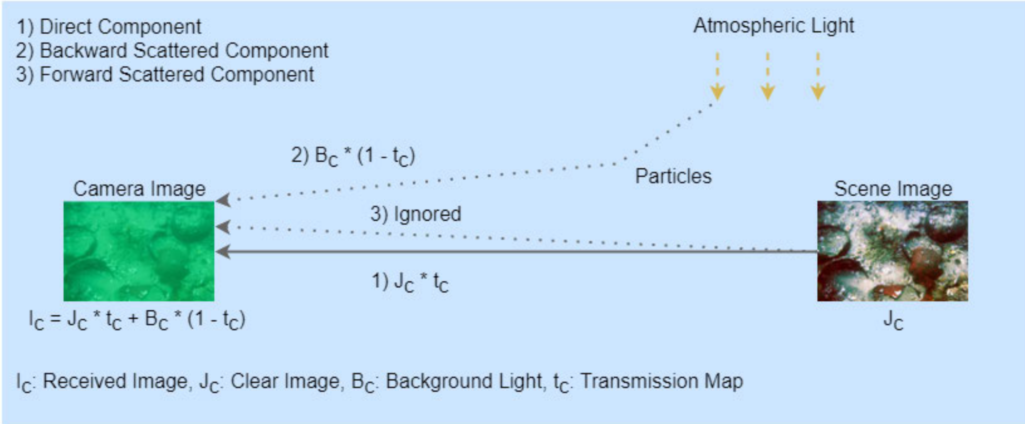


FIGURE 3. Visualization of underwater image formation model.

The irradiance E at a distance d from an initial position 0 can be mathematically modeled as [23]:

$$E(d, \lambda) = E(0, \lambda)e^{-a_\lambda d}e^{-b_\lambda d}. \tag{2}$$

Here, d denotes the path traveled by light, λ represents the wavelength, and a_λ and b_λ represent the absorption and scattering coefficients of the medium, respectively. Based on Equation-2, the direct component E_d changes with a distance is expressed as follows:

$$E_d = J_c e^{-a_\lambda d(x)}e^{-b_\lambda d(x)}. \tag{3}$$

Here, $c \in \{R, G, B\}$, $J_c \in \mathbb{R}^{3\times H\times W}$ represents the clear image, $d(x)$ is the distance between objects and the camera.

$$t_c = e^{-a_\lambda d(x)}e^{-b_\lambda d(x)} = e^{-p_\lambda d(x)}. \tag{4}$$

Here, $t_c \in \mathbb{R}^{3\times H\times W}$ represents the transmission map, attenuation coefficient, p_λ (Figure-1), denotes the sum of a_λ and b_λ [2], [22]. The backward scattering component E_b are given below [2]:

$$E_b = B_c(1 - e^{-p_\lambda d}) = B_c(1 - t_c). \tag{5}$$

Here, $B_c \in \mathbb{R}^{3\times 1\times 1}$ denotes the background light. The underwater image formation model can be expressed as follows (Figure-3) [2], [9], [23]:

$$I_c = J_c t_c + B_c(1 - t_c) \tag{6}$$

where $I_c \in \mathbb{R}^{3\times H\times W}$ is received image. For restoring the clear image J_c, the background light and the transmission map should be estimated first.

B. UNDERWATER DARK CHANNEL PRIOR

Dark Channel Prior (DCP) is a statistical approach that is proposed by accepting as a priori that outdoor images contain dark channels, that is, the lowest pixel values in the patchs of image are zero. DCP is based on the analysis of 5000 random images which results 75 percent of the dark channels have pixel values of 0 and 90 percent have pixel values below 25 in [0, 255] image space. Shadows, vividly colored objects with one or two dominant channels and dark-colored objects in outdoor images are stated as the main reasons for the low density of DCP. Dark channel $J_{dark} \in \mathbb{R}^{1\times H\times W}$ of an image which is stated to tend to be zero, is obtained by performing minimum pooling filter to minimum channel values of the image, and is defined as follows [11]:

$$J_{dark}(x) = \min_{y\in\Omega(x)} \left(\min_{c\in\{R,G,B\}} J_c(y)\right), \tag{7}$$

where J_c is a channel of image J, x represents image pixel and $\Omega(x)$ is the image patch centered at x.

Underwater Dark Channel Prior (UDCP), which is developed by adapting DCP, is based on the fact that underwater environments have wavelength-dependent absorption properties. While calculating the dark channel, J_{UDCP}, of the underwater image, only G and B color channels from RGB color channels are used. Therefore, $J_{UDCP} \in \mathbb{R}^{1\times H\times W}$ is given as follows [10]:

$$J_{UDCP}(x) = \min_{y\in\Omega(x)} \left(\min_{c\in\{G,B\}} J^c(y)\right). \tag{8}$$

The basic purpose of calculating the dark channel of an image is to estimate both the transmission map and the background light, as in [10] and [11]. In our approach, the transmission map is estimated through the PIM, while the background light is estimated using the dark channel. For background light estimation, 0.1% of the pixels with large values of the J_{UDCP} of the image are selected [10], [11]. The largest values in the RGB color channels of the corresponding pixels of the underwater image are assigned as background light for each channel separately.

After estimating the transmission map $t_c(x)$ from the model output and the background light B_c, the enhanced image of the PIM, $\hat{J}_c^{PIM}(x)$, can be determined according to Equation-6 as follows:

$$\hat{J}_c^{PIM}(x) = \frac{I_c(x) - B_c}{t_c(x) + \epsilon} + B_c. \tag{9}$$

where ϵ is a small value, i.e., 10^{-5} added to the denominator for numerical stability.

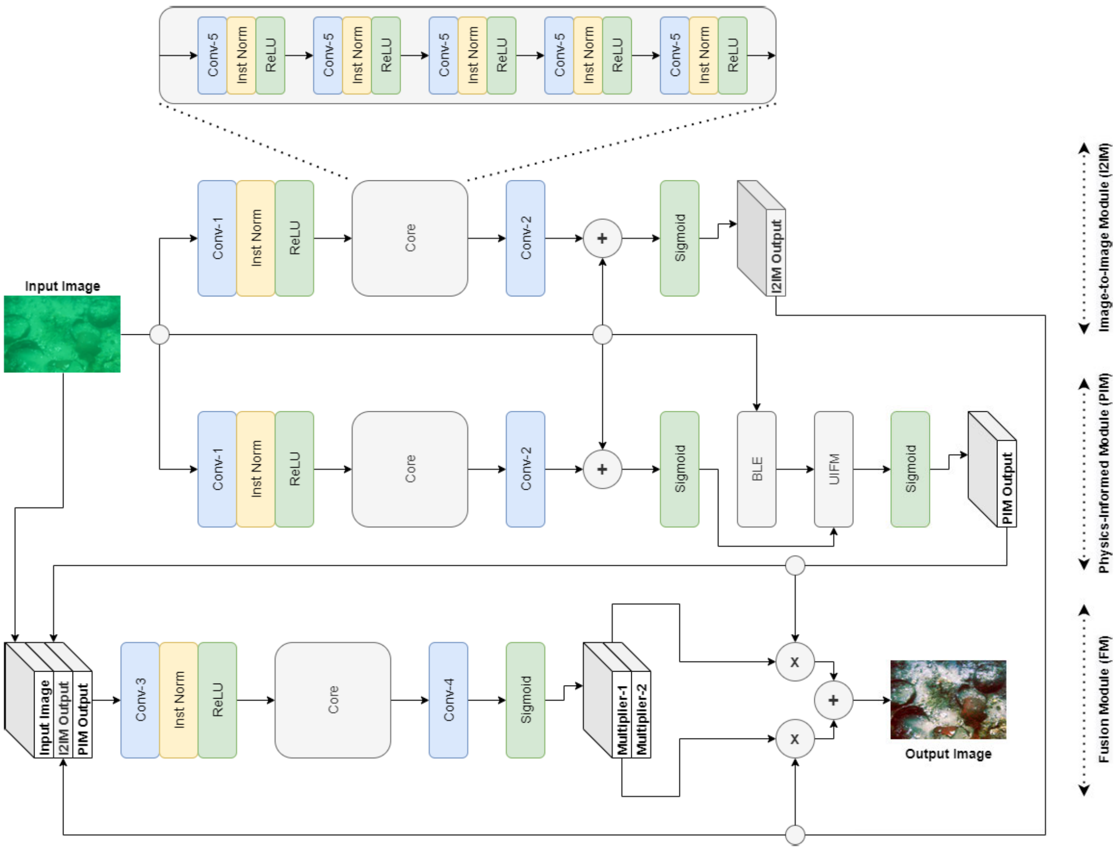


FIGURE 4. An overview of the HUWIE-Net architecture: Image-to-Image Module, Physics-Informed Module, Fusion Module, BLE: Background Light Estimation, UIFM: Underwater Image Formation Model. Note: None of the convolutional layers in the model perform weight sharing.

C. PROPOSED MODEL

This section provides detailed information about the proposed deep learning model. The model, HUWIE-Net, integrates a pixel-level-based model and a physical-channel-based model to leverage the strengths of both approaches. HUWIE-Net includes the I2IM with the task of color correction against light absorption and the PIM with the task of dehazing against light scattering. Two different approaches are integrated into the CNN model. The I2IM gets the received underwater image as input and outputs the estimated clear image with the same size as input image. The PIM gets the captured underwater image and the estimated background light and outputs the estimated clear image by exploring UIFM. I2IM and PIM module outputs are jointly used in Fusion Module (FM) to estimate the final clear image. The overall architecture is called as HUWIE-Net and illustrated in Figure-4.

The details of all convolution layers used in HUWIE-Net are given in Table-1. These convolutional layers are independent of each other, even if they share the same name, with no weight sharing involved. The absorption coefficients in the underwater environment depend on the light wavelength. These varying absorption coefficients lead to the idea of performing separate normalization to all channels of the image. After the convolution layers, there is an Inst Norm (instance normalization) layer that normalizes each channel separately according to its own statistical data [24]. Following the Inst Norm layers, the non-linear activation function Rectified Linear Unit (ReLU) is employed. The core blocks play a critical role in deep feature extraction and transformation within the Image-to-Image Module (I2IM), Physics Informed Module (PIM), and Fusion Module (FM). The core blocks employed in all three modules are identical with different weights for each module. These blocks further refine the features extracted by the previous convolutional layers, enabling the capture of complex environmental relationships inherent to underwater images, such as color distortions and scattering effects. In the I2IM, the core block focuses on enhancing colors and preserving structural details, while in the PIM, it leverages physical model to learn the details of

TABLE 1. Parameters of HUWIE-Net convolution layers.

Layer	In Chan.	Out Chan.	Kernel Size	Stride	Padding
Conv1	3	64	3x3	1x1	1x1
Conv2	64	3	1x1	1x1	0x0
Conv3	9	64	3x3	1x1	1x1
Conv4	64	6	1x1	1x1	0x0
Conv5	64	64	3x3	1x1	1x1

the transmission map and improve dehazing performance. In the FM, the core block integrates and combines the features of the I2IM and PIM. In the I2IM, after the core block the input image is added to the output of Conv-2 as a residual connection inspired by [26]. Similar to normalizing the input image in the range [0, 1], the sigmoid function is used in the last layer of I2IM to ensure that the output image is also in this range.

PIM is a module of HUWIE-Net based on UIFM using UDCP [10]. In PIM, a deep network learns and estimates the transmission map ($t_c(x)$) of underwater images. To keep the 3-channel transmission map in the range of [0, 1], a sigmoid activation function is used in the network. The background light estimated from the dark channel of the input images, and the transmission map are substituted into Equation-9 to obtain the enhanced underwater image.

The other module of HUWIE-Net, FM, consists of convolution, instance normalization, ReLU and sigmoid activation functions layers. FM generates the final enhanced image by effectively fusing the I2IM and PIM outputs. FM input is 9-channels data obtained by concatenating input image, I2IM and PIM outputs. FM outputs 6-channels of data, 3 of which are the multiplier of the output of I2IM and 3 of which are the multiplier of the output of PIM. The final enhanced image is generated by multiplying the outputs of I2IM and PIM by the multipliers from FM, and then summing the results.

HUWIE-Net is trained using a total loss function of l_{L1} and l_{SSIM}, which are obtained from the output images of network and the reference images from the dataset. The total loss is given below:

$$l_T = l_{L1} + l_{SSIM}. \tag{10}$$

l_{L1} is defined as the mean of the sum of the absolute differences between the pixel values of the reference image, $J(x)$, and the output image, $\hat{J}(x)$,

$$l_{L1} = \frac{1}{n}\sum_{i=1}^{n} |J(x) - \hat{J}(x)|, \tag{11}$$

n is the number of pixels of the images. l_{SSIM} is defined as follows:

$$l_{SSIM} = 1 - S, \tag{12}$$

where Structural Similarity Index Measure (SSIM), denoted as S, is obtained by multiplying the luminance term, the contrast term and the structural terms as a measure of the textural and structural similarity between the reference image and the output image. The formula for SSIM, is given below [28]:

$$S(J, \hat{J}) = \frac{(2\mu_J\mu_{\hat{J}} + C_1)(2\sigma_{J\hat{J}} + C_2)}{(\mu_J^2 + \mu_{\hat{J}}^2 + C_1)(\sigma_J^2 + \sigma_{\hat{J}}^2 + C_2)}, \tag{13}$$

where μ_J and $\mu_{\hat{J}}$ are the mean values of the patches of reference and output images, σ_J and $\sigma_{\hat{J}}$ are their standard deviations, $\sigma_{J\hat{J}}$ is their covariance, $C_1 = 0.1^2$, and $C_2 = 0.3^2$.

TABLE 2. Implementation details of HUWIE-Net.

Model	HUWIE-Net
Loss function	l_{L1}, l_{SSIM}
Training data	Random 800 images of UIEB dataset
Testing data	Remaining 90 images of UIEB dataset
Epoch	50
Batch size	8
Optimizer	Adam
Learning rate	1e-3
LR scheduling	Constant
Transforming	Resize input images to size 320 x 320 Random flip horizontally with probability 0.5
Platform	CPU: Intel Core i9-9900K 3.6 GHz GPU: Nvidia RTX 2080 Ti RAM: 32 GB
Framework	PyTorch

TABLE 3. Mean and variance of MSE for the outputs of the methods on real-world underwater images.

Method	μ_{MSE}	σ_{MSE}
Raw images	1285.344	1021.837
HE	2008.058	1520.444
WB	781.278	719.953
UDCP	4069.233	2314.884
ULAP	1847.690	1518.420
UWCNN	893.065	842.898
UIEC^2-Net	429.073	436.038
HUWIE-Net	392.725	344.875

l_{SSIM} is one of the widely used loss function, though it is not as commonly employed as l_{L1} [1], [14], [27]. The SSIM value is in the range of [0, 1], and a value of 1 means that the reference image and the output image are exactly the same.

IV. EXPERIMENTAL RESULTS

In this section, the performance results of the proposed HUWIE-Net model are presented. HUWIE-Net is compared with traditional and state-of-the-art methods using both qualitative and quantitative evaluations. In this context, visual outputs and metric calculations are provided for all methods. Histogram Equalization (HE), White Balance (WB), Underwater Dark Channel Prior (UDCP) [10], Underwater Light Attenuation Prior (ULAP) [12], Underwater Image Enhancement Convolution Neural Network (UWCNN) [14], Underwater Image Enhancement Convolution Neural Network using 2 Color Space (UIEC^2-Net) [1] and the proposed HUWIE-Net are the underwater image enhancement methods compared. Real-world underwater images from Underwater

TABLE 4. Mean and variance of SSIM for the outputs of the methods on real-world underwater images.

Method	μ_{SSIM}	σ_{SSIM}
Raw images	0.815	0.111
HE	0.783	0.113
WB	0.869	0.071
UDCP	0.596	0.135
ULAP	0.762	0.127
UWCNN	0.892	0.063
UIEC^2-Net	0.919	0.059
HUWIE-Net	0.922	0.050

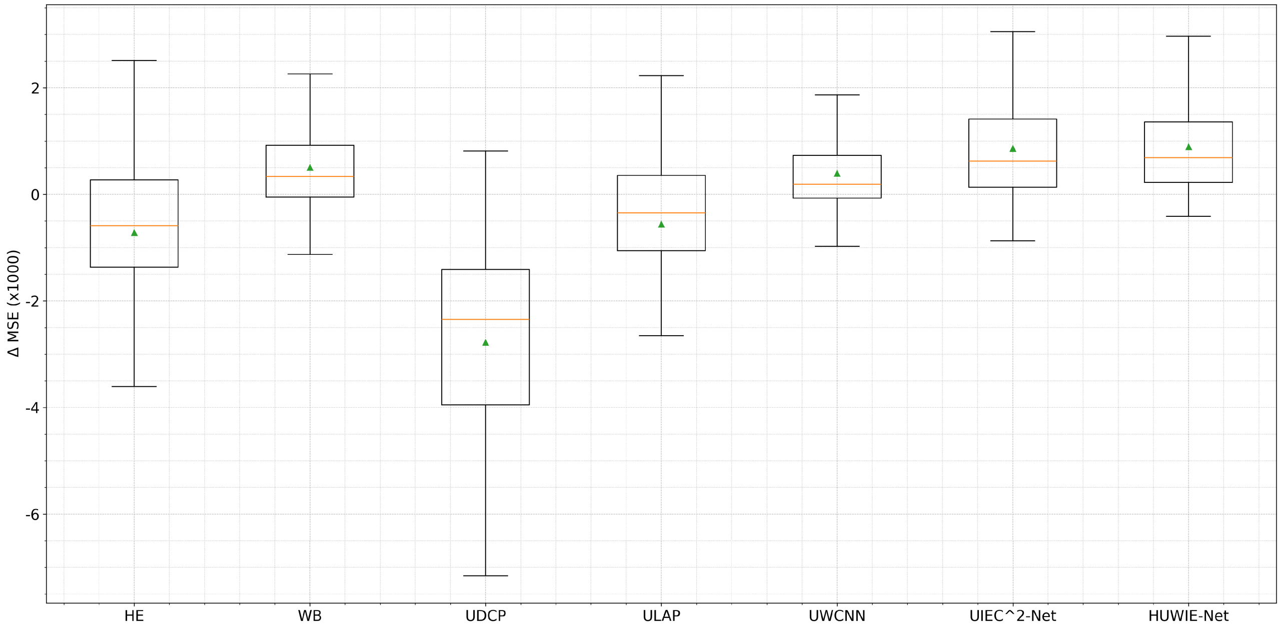


FIGURE 5. Box plot of ΔMSE for method outputs on real-world underwater images. Green arrow indicates the mean value and orange line represents the median. Higher values and a narrower distribution for ΔMSE indicate better performance.

TABLE 5. Box plot data of ΔMSE: Q1 (First Quartile), Median, Q3 (Third Quartile), LW (Lower Whisker), UW (Upper Whisker).

Method	Q1	Median	Q3	LW	UW
HE	-1368.3	-588.8	270.1	-3609.8	2508.8
WB	-50.8	331.8	921.6	-1128.6	2258.4
UDCP	-3951.6	-2348.5	-1412.2	-7160.2	812.1
ULAP	-1061.5	-350.9	353.6	-2654.1	2225.4
UWCNN	-74.0	186.7	728.5	-975.3	1867.5
UIEC^2-Net	131.2	623.9	1412.0	-874.5	3050.3
HUWIE-Net	219.9	686.1	1358.0	-415.6	2966.8

TABLE 6. Box plot data of ΔSSIM: Q1 (First Quartile), Median, Q3 (Third Quartile), LW (Lower Whisker), UW (Upper Whisker).

Method	Q1	Median	Q3	LW	UW
HE	-0.0554	0.0437	0.1271	-0.2963	0.3364
WB	-0.0860	-0.0348	0.0085	-0.2135	0.1451
UDCP	0.1299	0.2208	0.3043	-0.1167	0.5173
ULAP	-0.0108	0.0488	0.1103	-0.1855	0.2718
UWCNN	-0.1265	-0.0687	-0.0119	-0.2833	0.0595
UIEC^2-Net	-0.1833	-0.0877	-0.0208	-0.3726	0.1226
HUWIE-Net	-0.1725	-0.0862	-0.0263	-0.3809	0.0986

Image Enhancement Benchmark (UIEB) dataset [19] are used for training and testing. The images from UIEB dataset, which consist of various scenes such as coral, fish, diving, marine life, rocks, wreckage, and sculptures, can be categorized as bluish, greenish, shallow, and low-illuminated images. It can be argued that this dataset, derived from images captured across various scenes and conditions, provides a certain degree of generalization. Therefore, tests performed with this dataset provide information about the generalization performance of the models. In order to create a fair comparison environment, deep learning-based models are trained using exactly the same training data instead of pre-trained models. 800 images randomly selected from 890 images in the UIEB dataset are used for training, and the remaining 90 images are used for testing. Publicly available codes were useful when developing our code [14],[1] [1].[2] Implementation details are given in Table-2.

[1]https://li-chongyi.github.io/proj_underwater_image_synthesis.html
[2]https://github.com/BIGWangYuDong/UWEnhancement

A. EVALUATION ON REAL-WORLD UNDERWATER IMAGES

To evaluate the level of enhancement, commonly used metrics are calculated using the outputs of underwater image enhancement methods and reference images: Mean Squared Error (MSE) and Structural Similarity Index (SSIM) [27]. A low MSE value indicates that the pixel values of the reference image and the output image are close to each other. SSIM, which is in the range of [0, 1] and higher values indicate higher levels of similarity, is a measure of the textural and structural similarity between the reference image and the output image.

The MSE and SSIM of test images are calculated using corresponding reference images. Similarly, after enhancement, metrics of model outputs are calculated and means and standard deviations of the metrics are given in Table-3 and Table-4. In the tables, the result of the method showing better performance is red colored, while the results of the second ranked method are blue colored. According to the μ_{MSE}, it can be interpreted that the proposed model enhances underwater images with pixel values closest to the reference image. Based on these values, it is observed

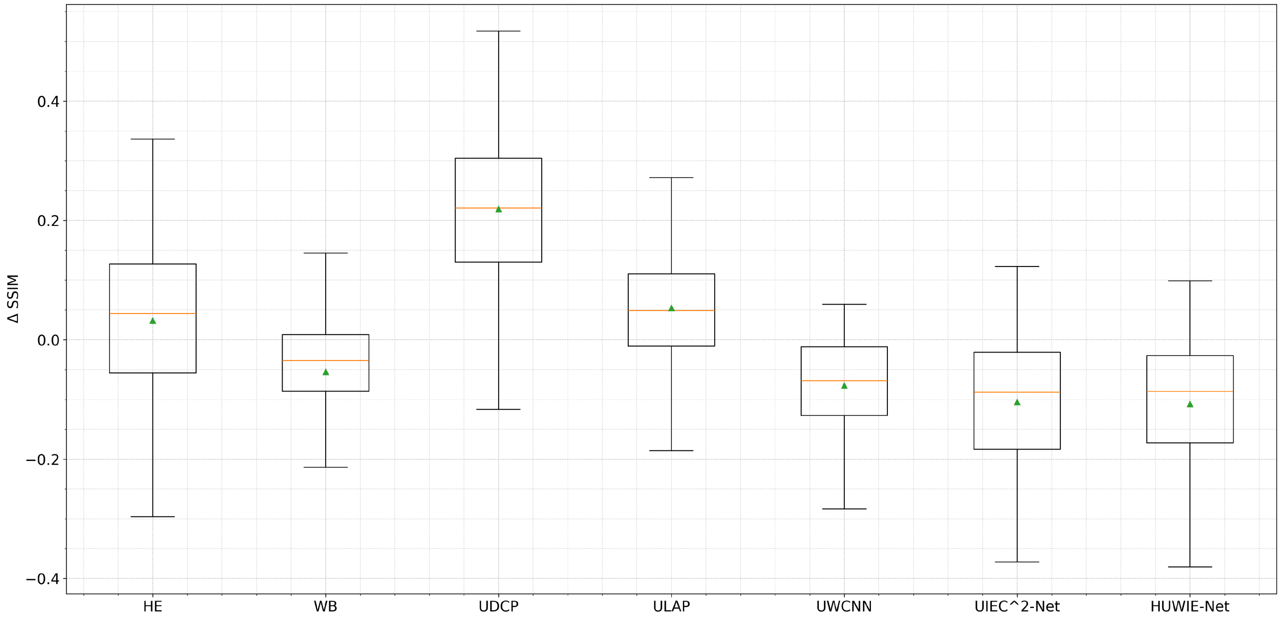


FIGURE 6. Box plot of ΔSSIM for method outputs on real-world underwater images. Green arrow indicates the mean value and orange line represents the median. Lower values and a narrower distribution for ΔSSIM indicate better performance.

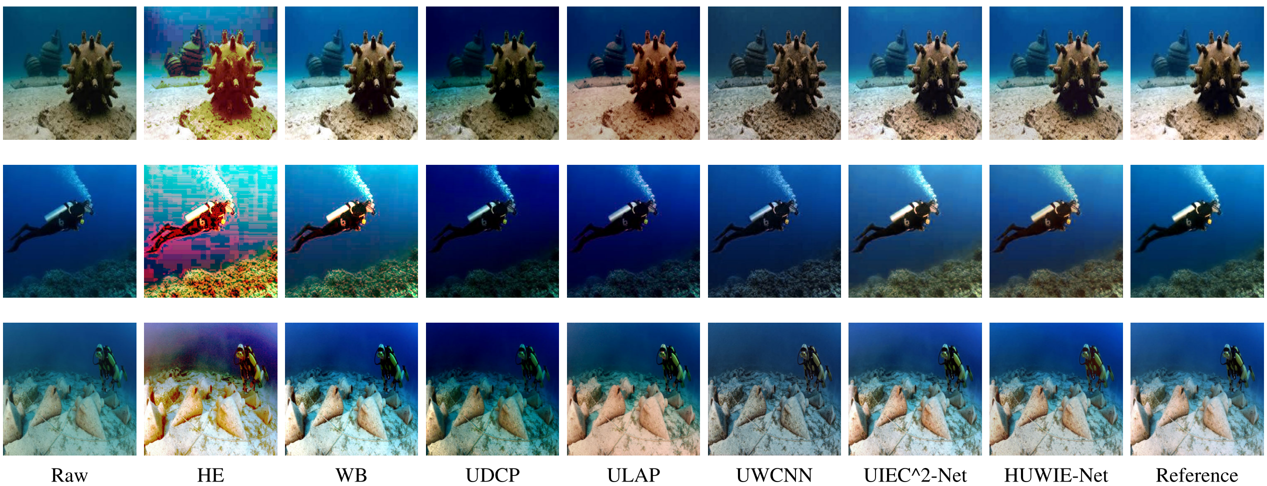


FIGURE 7. Sample outputs of the methods on bluish images, in order: raw, HE, WB, UDCP [10], ULAP [12], UWCNN [14], UIEC^2-Net [1], HUWIE-Net, and reference.

that HUWIE-Net enhances images with pixel values closer to the reference images on average. The σ_{MSE} given in the table provides information about the stability of the models. HUWIE-Net with the smallest σ_{MSE} has better stability performance. When evaluated according to SSIM in Table-4, the μ_{SSIM} of HUWIE-Net indicates that the output images of the model have more textural and structural similarity with the reference images. Similarly, HUWIE-Net stands out in terms of textural and structural similarity to the reference image. In terms of the SSIM, HUWIE-Net also demonstrates better stability performance. UIEC^2-Net achieves the second-best performance in both metrics.

It can be said that existing underwater image enhancement methods focus on improving low-quality images but do not address preserving the quality of images taken under good conditions or those that are already optimal and cannot be further improved. Although the average metric results of the methods are shared, the quality of some input images is degraded. The distribution of metric values of the method outputs can provide this information. The mean squared error calculated using the raw image (I_i) and the corresponding reference image (J_i) is defined as MSE_{I_i}, while the mean squared error calculated using the enhanced image ($\hat{J}_i$) and the corresponding reference image is defined as $\text{MSE}_{\hat{J}_i}$. The difference between MSE_{I_i} and $\text{MSE}_{\hat{J}_i}$ is expressed as ΔMSE_i.

FIGURE 8. Sample outputs of the methods on greenish images, in order: raw, HE, WB, UDCP [10], ULAP [12], UWCNN [14], UIEC^2-Net [1], HUWIE-Net, and reference.

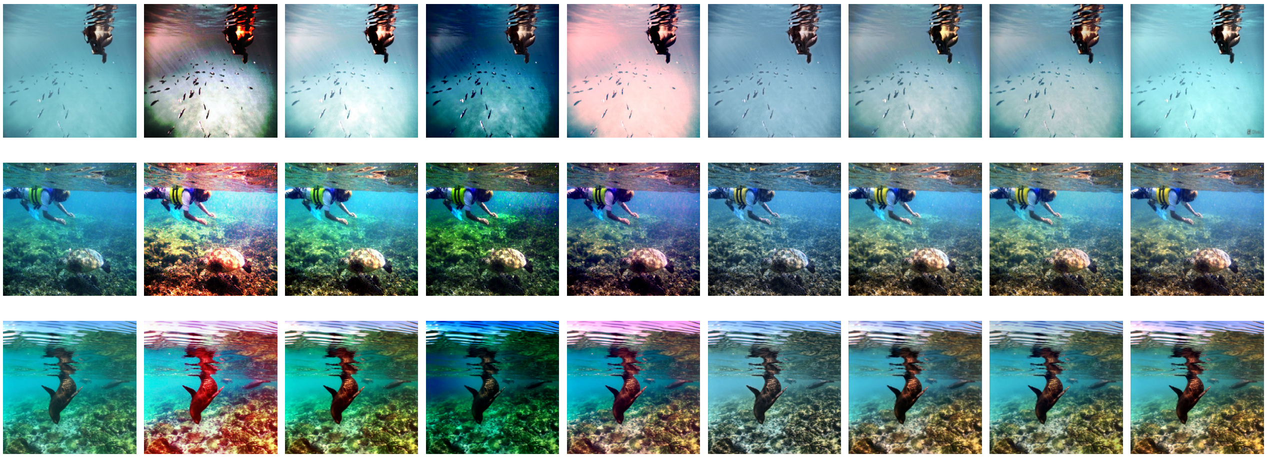


FIGURE 9. Sample outputs of the methods on shallow-water images, in order: raw, HE, WB, UDCP [10], ULAP [12], UWCNN [14], UIEC^2-Net [1], HUWIE-Net, and reference.

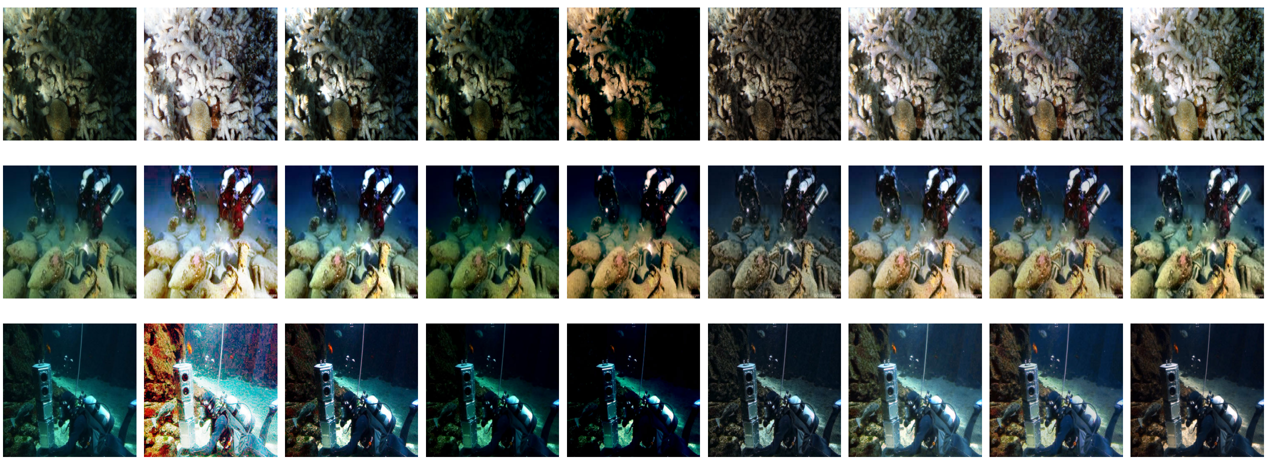
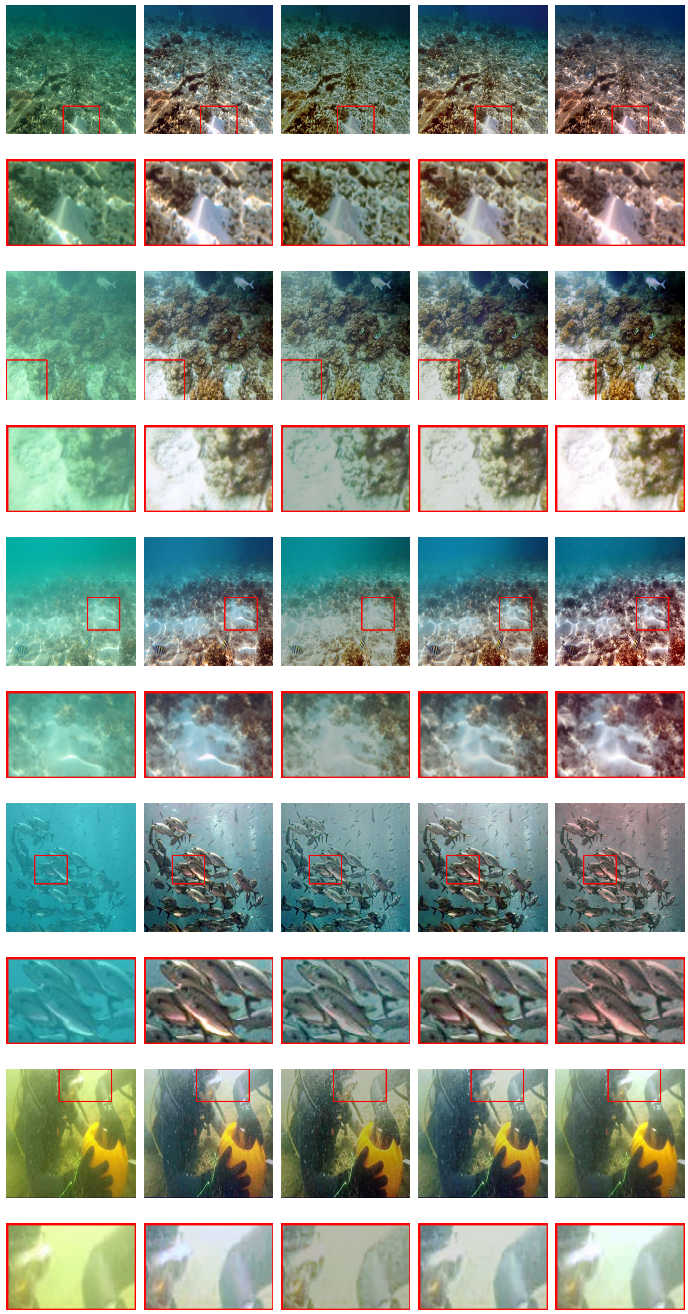


FIGURE 10. Sample outputs of the methods on low-illuminated images, in order: raw, HE, WB, UDCP [10], ULAP [12], UWCNN [14], UIEC^2-Net [1], HUWIE-Net, and reference.



Raw I2IM PIM HUWIE-Net Reference

FIGURE 11. Sample outputs of the methods for ablation study, in order: raw, I2IM, PIM, HUWIE-Net, and reference.

The ΔMSE_i values are calculated for each test images, as shown below:

$$\Delta\text{MSE}_i = \text{MSE}_{I_i} - \text{MSE}_{\hat{J}_i} \tag{14}$$

Here, i denotes the index of the test image. ΔMSE arrays, consisting of ΔMSE_i values, are created for each method, i.e., $\Delta\text{MSE}_{UIEC\text{^}2-Net}$, $\Delta\text{MSE}_{HUWIE-Net}$. Figure-5 presents a box plot showing the distribution of the ΔMSE arrays. The statistical data for this box plot are provided in Table-5. Higher ΔMSE values and a narrower distribution indicate better performance. A box plot is a chart that visualizes the distribution of data by highlighting five summary statistics: the lower whisker, first quartile, median, third quartile, and upper whisker [33]. This plot serves as a useful tool for comparing the methods. Since the medians of the ΔMSE for HE, UDCP, and ULAP are below 0, it indicates that while these methods improve some test images, they degrade the quality of the majority. Among the conventional methods, WB demonstrates more effective compared to the others. Deep learning-based methods improve the quality of the majority of test images. Figure-5 highlights the generalization problem as one of the key challenges faced by underwater image enhancement methods. HUWIE-Net shows a higher median (686.1 vs 623.9) compared to UIEC^2-Net indicating better average performance. The range between LW and UW for HUWIE-Net (−415.6 to 2966.8) is narrower than that of UIEC^2-Net (-874.5 to 3050.3), indicating better stability in performance. The ΔSSIM box plot, which is presented in Figure-6, shows similar results to the ΔMSE box plot. The values of this box plot are provided in Table-6. Lower values of ΔSSIM and a narrower distribution indicate better performance. Although HUWIE-Net and UIEC^2-Net show competitive results in average performance, the range between LW (-0.3809) and UW (0.0986) for HUWIE-Net is narrower than that of UIEC^2-Net (−0.3726 to 0.1226), indicating that HUWIE-Net is more stable. Other methods, particularly UDCP, WB, and HE, show significantly weaker performance in preserving structural and textural similarity.

To compare the outputs of the methods, test images are categorized as bluish (Figure-7), greenish (Figure-8), shallow water (Figure-9), and low-illumination underwater images (Figure-10), following the approach used in other studies in the literature. These figures present samples of the input images and the corresponding outputs of the methods. The second image in Figure-7 and the third image in Figure-9 illustrate examples of HE over-enhancing the red channel, which has the largest absorption coefficient, while the third image in Figure-8 is an example of contrast improvement. WB, a color correction approach that adjusts the average of each color channel relative to a reference, is effective for images with a dominant color channel such as greenish or bluish, but suffers from poor generalization performance. ULAP is considered to be scene-dependent because it improves color quality in some images while degrading it in others. UDCP generally performs poorly. Although UWCNN removes haze from the images, the contrast remains low, as shown in the images in Figure-8. While UIEC^2-Net shows good performance, some color artifacts are present in certain areas, such as the third image in Figure-8 (reddish tone in rocks) and the third image in Figure-9. The enhanced images of HUWIE-Net, which demonstrates better performance in color correction, dehazing, and generalization, are presented in the figures.

TABLE 7. Ablation study: mean and variance of MSE and SSIM for the outputs of the methods on real-world underwater images.

Method	μ_{MSE}	σ_{MSE}	μ_{SSIM}	σ_{SSIM}
Received images	1285.344	1021.837	0.815	0.111
I2IM	431.658	420.965	0.918	0.059
PIM	1012.689	754.080	0.868	0.068
HUWIE-Net	392.725	344.875	0.922	0.050

B. ABLATION STUDY

Ablation study is performed by training and testing I2IM and PIM modules separately. According to the MSE and SSIM values in Table-7, HUWIE-Net shows better performance, followed by I2IM and PIM. The color correction performance of I2IM and the dehazing performance of PIM can be observed from Figure-11. Although modules using similar deep network structures improve image quality, the significantly better results obtained with I2IM compared to PIM are considered as a clear evidence of the limitations of the underwater image formation model [2], [9], [23]. It can be observed that fusing I2IM with PIM results in improved color tones at certain points. This demonstrates the effectiveness of joint optimization compared to optimizing I2IM and PIM separately, as highlighted in Figure-11. At the point marked in the second and the fifth image in Figure-11, PIM effectively improves the over-brightness problem of I2IM. It is even possible to say that the HUWIE-Net outputs are more natural than the reference image. Ablation studies show that both PIM and I2IM independently enhance raw images. Together with these two modules, the FM module forms HUWIE-Net, which provides better results.

C. FUTURE WORK

Future work could focus on improving PIM's performance metrics, particularly in generating higher-contrast images. Since the colors in PIM outputs are relatively dull, incorporating a loss function based on the transmission map output of PIM could be beneficial. Additionally, the development of a more advanced underwater image formation model, including an investigation into the effects of neglecting E_f, could be explored. Expanding the dataset may also be considered as part of future research.

V. CONCLUSION

In this paper, we have proposed the HUWIE-Net underwater image enhancement framework. This model is designed to jointly optimize both pixel-level-based and physical-channel-based enhancement methods. In this framework, the Image-to-Image Module is employed for the purpose of pixel-level color correction of degraded underwater images, while the Physics-Informed Module is employed for dehazing based on the underwater image formation model. A joint loss function is used to both modules to enforce joint optimization, leading to improved underwater image enhancement performance. Using both qualitative and quantitative evaluations, HUWIE-Net is compared with both traditional and state-of-the-art methods. Experiments with real-world underwater images demonstrate the effectiveness of our method through both qualitative and quantitative evaluations. Additionally, the distributions of the enhancement levels for each method are presented using box plots. According to these plots, the proposed method demonstrates better stability performance.

ACKNOWLEDGMENT

The authors would like to thank Aselsan Inc., for funding this study.

REFERENCES

[1] Y. Wang, J. Guo, H. Gao, and H. Yue, "UIEC^2-Net: CNN-based underwater image enhancement using two color space," *Signal Process., Image Commun.*, vol. 96, Aug. 2021, Art. no. 116250.

[2] M. Yang, J. Hu, C. Li, G. Rohde, Y. Du, and K. Hu, "An in-depth survey of underwater image enhancement and restoration," *IEEE Access*, vol. 7, pp. 123638–123657, 2019.

[3] N. G. Jerlov, *Marine Optics*. Amsterdam, The Netherlands: Elsevier, 1976.

[4] C.-Y. Li, J.-C. Guo, R.-M. Cong, Y.-W. Pang, and B. Wang, "Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior," *IEEE Trans. Image Process.*, vol. 25, no. 12, pp. 5664–5677, Dec. 2016.

[5] M. S. Hitam, E. A. Awalludin, W. N. J. H. W. Yussof, and Z. Bachok, "Mixture contrast limited adaptive histogram equalization for underwater image enhancement," in *Proc. Int. Conf. Comput. Appl. Technol. (ICCAT)*, Jan. 2013, pp. 1–5.

[6] C. Li, J. Guo, C. Guo, R. Cong, and J. Gong, "A hybrid method for underwater image correction," *Pattern Recognit. Lett.*, vol. 94, pp. 62–67, Jul. 2017, doi: 10.1016/j.patrec.2017.05.023.

[7] C. Li, J. Guo, B. Wang, R. Cong, Y. Zhang, and J. Wang, "Single underwater image enhancement based on color cast removal and visibility restoration," *J. Electron. Imag.*, vol. 25, no. 3, Jun. 2016, Art. no. 033012.

[8] C. Li, J. Quo, Y. Pang, S. Chen, and J. Wang, "Single underwater image restoration by blue-green channels dehazing and red channel correction," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 1731–1735, doi: 10.1109/ICASSP.2016.7471973.

[9] J. Y. Chiang and Y.-C. Chen, "Underwater image enhancement by wavelength compensation and dehazing," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1756–1769, Apr. 2012, doi: 10.1109/TIP.2011.2179666.

[10] P. Drews, E. Nascimento, F. Moraes, S. Botelho, and M. Campos, "Transmission estimation in underwater single images," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Dec. 2013, pp. 825–830.

[11] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011, doi: 10.1109/TPAMI.2010.168.

[12] W. Song, Y. Wang, D. Huang, and D. Tjondronegoro, "A rapid scene depth estimation model based on underwater light attenuation prior for underwater image restoration," in *Proc. Adv. Multimedia Inf. Process.–PCM*, R. Hong, W.-H. Cheng, T. Yamasaki, M. Wang, and C.-W. Ngo, Eds., Cham, Switzerland: Springer, Jan. 2018, pp. 678–688.

[13] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and P. Bekaert, "Color balance and fusion for underwater image enhancement," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 379–393, Jan. 2018, doi: 10.1109/TIP.2017.2759252.

[14] C. Li, S. Anwar, and F. Porikli, "Underwater scene prior inspired deep underwater image and video enhancement," *Pattern Recognit.*, vol. 98, Feb. 2020, Art. no. 107038.

[15] K. Cao, Y.-T. Peng, and P. C. Cosman, "Underwater image restoration using deep networks to estimate background light and scene depth," in *Proc. IEEE Southwest Symp. Image Anal. Interpretation (SSIAI)*, Apr. 2018, pp. 1–4, doi: 10.1109/SSIAI.2018.8470347.

[16] X. Sun, L. Liu, Q. Li, J. Dong, E. Lima, and R. Yin, "Deep pixel-to-pixel network for underwater image enhancement and restoration," *IET Image Process.*, vol. 13, no. 3, pp. 469–474, Feb. 2019.

[17] P. Uplavikar, Z. Wu, and Z. Wang, "All-in-One underwater image enhancement using domain-adversarial learning," in *Proc. CVPR Workshops*, Jan. 2019, pp. 1–8.

[18] Y. Wang, J. Zhang, Y. Cao, and Z. Wang, "A deep CNN method for underwater image enhancement," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 1382–1386.

[19] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao, "An underwater image enhancement benchmark dataset and beyond," *IEEE Trans. Image Process.*, vol. 29, pp. 4376–4389, 2020.

[20] Y. Huang, F. Yuan, F. Xiao, J. Lu, and E. Cheng, "Underwater image enhancement based on zero-reference deep network," *IEEE J. Ocean. Eng.*, vol. 48, no. 3, pp. 903–924, Apr. 2023.

[21] J. S. Jaffe, "Computer modeling and the design of optimal underwater imaging systems," *IEEE J. Ocean. Eng.*, vol. 15, no. 2, pp. 101–111, Apr. 1990.

[22] R. Schettini and S. Corchs, "Underwater image processing: State of the art of restoration and image enhancement methods," *EURASIP J. Adv. Signal Process.*, vol. 2010, no. 1, pp. 1–14, Dec. 2010.

[23] G. Hou, X. Zhao, Z. Pan, H. Yang, L. Tan, and J. Li, "Benchmarking underwater image enhancement and restoration, and beyond," *IEEE Access*, vol. 8, pp. 122078–122091, 2020.

[24] PyTorch Documentation. (2024). *InstanceNorm2d*. Accessed: Sep. 9, 2024. [Online]. Available: https://pytorch.org/docs/stable/generated/torch.nn.InstanceNorm2d.html

[25] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[27] S. Anwar and C. Li, "Diving deeper into underwater image enhancement: A survey," *Signal Process., Image Commun.*, vol. 89, Nov. 2020, Art. no. 115978.

[28] MathWorks Documentation. (2024). *SSIM*. Accessed: Sep. 9, 2024. [Online]. Available: https://www.mathworks.com/help/images/ref/ssim.html

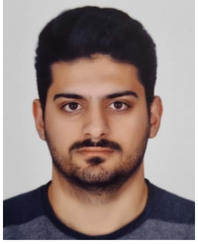
[29] B. McGlamery, "A computer model for underwater camera systems," *Proc. SPIE*, vol. 208, pp. 221–231, Mar. 1980.

[30] *PyTorch*. Accessed: Sep. 9, 2024. [Online]. Available: https://pytorch.org/

[31] D. Akkaynak and T. Treibitz, "A revised underwater image formation model," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6723–6732.

[32] Y. Y. Schechner and N. Karpel, "Clear underwater vision," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., (CVPR)*, vol. 1, Jun. 2004, pp. 536–543.

[33] B. Iglewicz, "Summarizing data with boxplots," in *International Encyclopedia of Statistical Science*, M. Lovric, Ed., Berlin, Germany: Springer, 2011, pp. 1572–1575, doi: 10.1007/978-3-642-04898-2_582.



OZAN DEMİR received the B.Sc. degree in electronics and communication engineering from Yildiz Technical University, Istanbul, Türkiye, in 2016, and the M.Sc. degree in electrical and electronics engineering from Gazi University, Ankara, Türkiye, in 2019. He is currently pursuing the Ph.D. degree in electronics and communication engineering with Istanbul Technical University, Istanbul. Since 2016, he has been working as a Lead Engineer with Defense Systems Technologies Division, Aselsan Inc. His research interests include computer vision, deep learning, and underwater acoustic systems.

METIN AKTAS received the B.Sc. degree from the Department of Electrical and Electronics Engineering, Dokuz Eylul University, Izmir, Türkiye, in 2002, and the M.Sc. and Ph.D. degrees from the Department of Electronics Engineering, Middle East Techinal University, Ankara, Türkiye, in 2004 and 2012, respectively. Since 2005, he has been working as a Senior Chief Engineer with Aselsan Inc. From 2005 to 2014, he worked as a Avionic Software Engineer at Microelectronics, Guidance, and Electro-Optics Division, Aselsan, where he is currently working with Defense Systems Technologies Division. His research interests include array signal processing, computer vision, machine learning, deep learning, remote sensing, and security systems.

ENDER M. EKSIOGLU (Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from the University of Michigan, Ann Arbor, in 1997 and 1999, respectively, and the Ph.D. degree from Istanbul Technical University (ITU), in 2005. He is currently working as a Professor with the Department of Electronics and Communication Engineering, ITU. His current research interests include deep learning for imaging problems, sparsity-based signal processing, and their applications.