Neurocomputing 584 (2024) 127585

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

DICAM: Deep Inception and Channel-wise Attention Modules for underwater image enhancement

Hamidreza Farhadi Tolie [a], Jinchang Ren [a,*], Eyad Elyan [b]

[a] National Subsea Centre, Robert Gordon University, Aberdeen, AB21 0BH, UK
[b] School of Computing, Robert Gordon University, Aberdeen, AB10 7GJ, UK

ARTICLE INFO

Communicated by Z. Wang

Keywords:
Underwater image enhancement
Deep learning
Inception module
Channel-wise attention module

ABSTRACT

In underwater environments, imaging devices suffer from water turbidity, attenuation of lights, scattering, and particles, leading to low quality, poor contrast, and biased color images. This has led to great challenges for underwater condition monitoring and inspection using conventional vision techniques. In recent years, underwater image enhancement has attracted increasing attention due to its critical role in improving the performance of current computer vision tasks in underwater object detection and segmentation. As existing methods, built mainly from natural scenes, have performance limitations in improving the color richness and distributions we propose a novel deep learning-based approach namely Deep Inception and Channel-wise Attention Modules (DICAM) to enhance the quality, contrast, and color cast of the hazy underwater images. The proposed DICAM model enhances the quality of underwater images, considering both the proportional degradations and non-uniform color cast. Extensive experiments on two publicly available underwater image enhancement datasets have verified the superiority of our proposed model compared with several state-of-the-art conventional and deep learning-based methods in terms of full-reference and reference-free image quality assessment metrics. The source code of our DICAM model is available at https://github.com/hfarhaditolie/DICAM.

1. Introduction

With the fast growth in marine engineering and ecosystem developments toward Net-Zero, automatic exploration, protection, and monitoring of subsea resources have become an active topic in recent years. Underwater images and videos can provide promising information for many engineering and research tasks including but not limited to the condition monitoring of energy infrastructures, visual mapping of seabed [1,2], trash detection [3], or detection and classification of underwater objects and events [4,5] (e.g., fishes, species, pipeline failure). However, due to various noises introduced by water turbidity, attenuation of lights, and particles in the underwater world, raw Underwater Images (UIs) and videos suffer severely from visual distortions resulting from non-uniform color deviation and blurring effects, especially the low degree of quality, contrast, and brightness [6–8].

To address these problems and improve the visibility of UIs for better practical usage, various Underwater Image Enhancement (UIE) methods have been proposed [9,10]. UIE methods tend to obtain a clearer image by improving the contrast and color distribution whilst removing blurring effects. Early research focused on contrast enhancement methods such as Histogram Equalization (HE) and Contrast Limited Adaptive Histogram Equalization (CLAHE) [11] on UIs. However, the degradation of the captured UI is proportionally dependent on the distance between the object and camera [12]. For instance, as shown in Fig. 1.(a), the content of the highlighted area in the raw underwater image, which is far from the camera, is not clearly visible compared with the central areas, which are closer to the camera. However, in the enhanced version the visibility of the content is further improved. Therefore, the conventional enhancement approaches fail to properly enhance the UIs [13]. Hence, it is urgent to propose practical enhancement methods specially designed for UIs, where existing UIE methods, as detailed below, can be categorized into three groups, i.e. non-physical model-based, physical model-based, and deep learning-based methods.

1.1. Non-physical model-based methods

This category includes methods that focus on modifying the intensity values of image pixels for enhancement. Iqbal et al. proposed the Integrated Colour Model (ICM) [14], in which they enhance the image by first stretching the image's contrast in the Red Green Blue (RGB) color space and then stretching the image's saturation and

* Corresponding author.
E-mail addresses: h.farhadi-tolie@rgu.ac.uk (H.F. Tolie), j.ren@rgu.ac.uk (J. Ren), e.elyan@rgu.ac.uk (E. Elyan).

https://doi.org/10.1016/j.neucom.2024.127585
Received 11 September 2023; Received in revised form 19 January 2024; Accepted 19 March 2024
Available online 24 March 2024

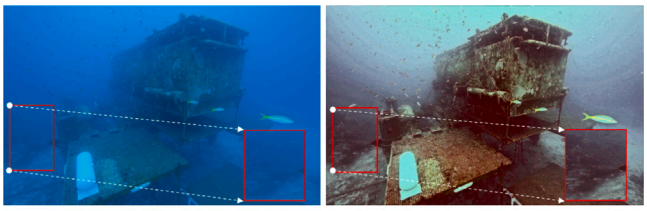

Fig. 1. Raw (a) and enhanced (b) underwater images taken from the UIEB dataset [13]. The raw image suffers from both proportional degradation and low color richness, where the content of the highlighted block is not clearly visible compared with the regions closer to the camera, while in the enhanced image (b) its visibility has improved.

intensity in the Hue Saturation Intensity (HSI) color space. Unsupervised Colour Correction Method (UCM) [15], proposed by Iqbal et al., applies contrast correction on the RGB image to increase the Red color and decrease the Blue color. Similar to ICM [14], it corrects the contrast of the saturation and intensity in HSI color space. In 2012. Ancuti et al. [16] proposed a multi-scale fusion strategy over contrast-enhanced and color-corrected images to produce the enhanced image. In [17], based on the human visual system, Fu et al. utilized a variational Retinex-based approach to decompose the reflectance and illumination from the color-corrected UI to generate the enhanced image. Inspired by [17], Zhang et al. [18] introduced the CIELAB (LAB)-color space Multi-Scale Retinex (LAB-MSR). Compared with the original Retinex, LAB-MSR [18] fuses bilateral and trilateral filters instead of the Gaussian filter in the CIELAB color space to enhance the raw UIs.

Moreover, Zhuang et al. [19] introduced an edge-preserving filtering Retinex algorithm, where guidance reflection and illumination are produced and fused with the guided image filtering for optimized enhancement. In Zhuang et al. [20], a Bayesian Retinex method was proposed to enhance the underwater images using the multi-order gradient priors. More recently, Zhou et al. [21] proposed an enhancement method based on the light scattering characteristic. They first categorize the color cast into five groups based on the average intensity values in RGB channels and then utilize the optical attenuation characteristic to compute the color information loss, followed by a multi-scene and a block-based histogram stretching approach to enhance the color and contrast of the underwater images.

These methods suffer from different noises, artifacts, and unpleasant color distortions due to their reliance on the observed data (i.e., pixel-level modifications) without any further consideration of the underwater environment's complexity and lighting conditions [22]. Also, as these methods rely on empirically set parameters, they lack a generalization ability for various underwater conditions.

1.2. Physical model-based methods

Physical model-based methods often utilize a mathematical model to describe the degradation of the image based on prior information and solve it as a reverse problem. Numerous methods used the Image Formation Model (IFM) [23,24] to describe an UI as a combination of a clear image and the background light weighed by a transmission map [25–28]. Moreover, several studies derived the transmission map based on prior information, such as Dark Channel Prior (DCP) [29]. Inspired by DCP, Drews Jr. et al. [30] proposed the Underwater Dark Channel Prior (UDCP) method, in which the blue and green channels of the raw RGB UI were used as the information source of underwater images. Peng et al. [24] proposed an IFM-based underwater image restoration method using an Image Blurriness and Light Absorption (IBLA) model, which utilized the UI's blurriness and light absorption to measure the background light, scene depth, and transmission maps rather than DCPs. Song et al. [31] proposed a scene depth estimation model based on the Underwater Light Attenuation Prior (ULAP) to correctly restore the true image. Zhou et al. [32] utilized the color-line model on a local scale, i.e. small patches, to recover their color line and estimating a transmission map to restore the UIs. Yang et al. [33] presented an UI restoration model based on lighting estimation of the local backscattering and the reflection-illumination decomposition to provide better edge restoration and colorfulness in recovered images. Liang et al. [34] introduced an UIE method to improve the low contrast and color cast of UIs, using a hierarchical searching technique to estimate the backscattered light and generalize the UDCP method toward generating a more robust transmission map.

Due to the ill-posed nature of the IFM problem, it requires different assumptions and priors (e.g., DCP) to estimate the transmission map. On the other hand, since the parameter estimation is complex, this has limited the performance of the physical model-based methods. In fact, most physical-based methods are time-consuming, visually unpleasing, and sensitive to different types of underwater images (e.g., oceanic or coastal) and the degradation level [22]. Thus it makes the physical model-based method highly challenging and complex.

1.3. Deep learning-based methods

In recent years, with the automatic and hierarchical feature extraction of deep learning models [35], which are invariant to the small changes of input data [36], several deep learning-based UIE methods have been developed. These include the Generative Adversarial Networks (GAN)-based, the Convolutional Neural Networks (CNN)-based, and the encoder–decoder-based neural networks. In 2019, Liu et al. [37] employed the residual networks to propose the Underwater ResNet (UResNet) and train it with their proposed Edge Difference Loss (EDL) alongside the well-known mean square error loss function. Later, Li et al. [13] introduced a CNN-based UIE method called Water-Net [13], which enhances the underwater images by using CNNs to extract features from the raw, gamma-corrected, histogram-equalized, and white-balanced underwater images before fusing them using their predicted confidence maps. Moreover, Islam et al. [38] proposed a conditional GAN-based method, namely Fast Underwater Image Enhancement Gan (FUnIE-GAN), with an U-NET [39] architecture-inspired generator network that utilizes the global content, color, local texture, and style information as a perceptual loss function to supervise the training of the model. In 2021, Wang et al. [40] proposed a CNN-based UIE method called UIEC^2-Net by incorporating two color spaces. UIEC^2-Net [40] employs the CNN architecture to extract features from the RGB and Hue Saturation Value (HSV) color spaces and combine them to produce an enhanced version of the raw UI. More recently, Sharman et al. [41], inspired by wavelength-based attributed deep networks, used convolutional layers and Convolutional Block Attention Modules (CBAM) in their proposed WaveNet method. WaveNet [41] separately processes the color channels of the raw images and fuses them in three steps by employing shortcut connections to preserve the extracted information from each channel at the previous step to construct the enhanced image.

Most deep learning-based UIE methods are either weakly supervised GAN-based or commonly used CNN architectures. However, due to the non-uniform attenuation of colors and proportional degradations of UIs, it is crucial to extract features with different ratios from each color channel. Moreover, as the red light disappears faster than the green and then the blue light, utilizing an adaptive weighting strategy to combine them to get the final enhanced image will ensure the accurate color correction of the method. Furthermore, since the light attenuation is non-uniform, the information loss will occur at different levels, and the existing objects and particles will also affect the attenuation rate. Thus, to address the aforementioned limitations and issues regarding underwater images, we propose to use a semi-adaptive feature size

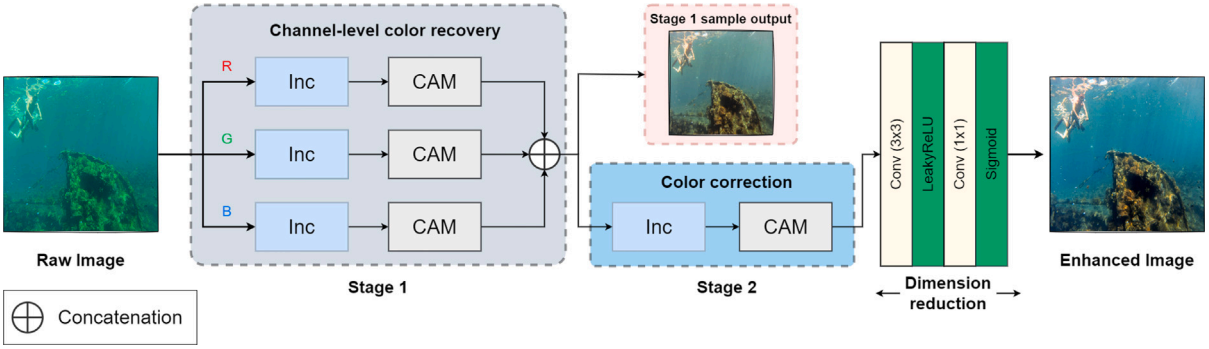

Fig. 2. Framework of the proposed DICAM model.

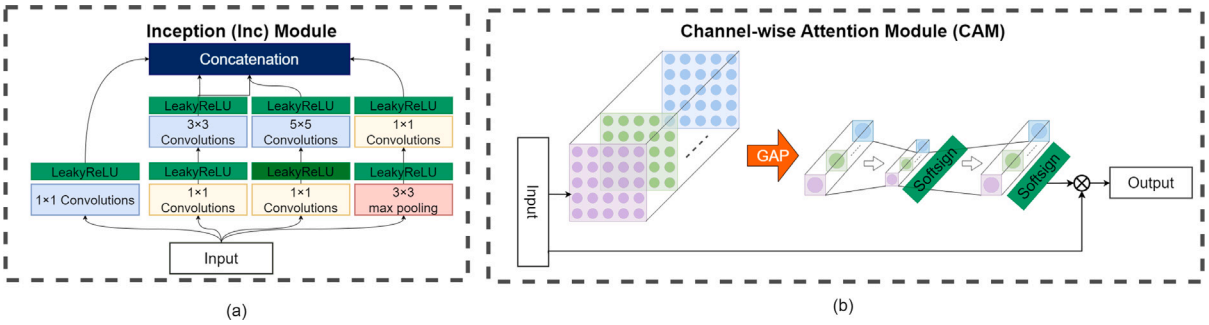

Fig. 3. Modules of the proposed DICAM model. From left to right: (a) Inception (Inc) module, (b) Channel-wise Attention Module (CAM).

to extract meaningful and effective features from raw underwater images.

The major contributions of our proposed approach can be highlighted as follows: (1) Multi-scale channel-wise feature extraction using an inception module to simultaneously quantify the color and distance-related proportional degradation, loss of color and content information, and color richness; (2) Adaptive fusion-based recovery and enhancement process, incorporating the Channel-wise Attention Module (CAM). Our approach enables us to generate high-quality enhanced images with a better color cast and richness, yielding a more visually pleasing and natural appearance through a dedicated color correction stage. Subsequent experiments have verified the superior performance of the proposed model in terms of image quality metrics, histogram comparison measures, and run-time.

2. The proposed method

Generally, underwater images suffer from two main defects: (1) proportional degradations and (2) non-uniform light attenuation leading to low visibility and color information loss. Proportional degradation, as illustrated in Fig. 1, mainly affects the visibility ratio of the content (e.g., objects, particles, etc.) in different regions of the image in a way that the content closer to the camera has better visibility than the distant ones, implying different degradation rates across various regions in underwater images. On the other hand, non-uniform light attenuation makes the majority of the captured UIs look bluish or greenish. Therefore, to address the aforementioned quality degradations and enhance the UIs accordingly, we designed a deep neural network architecture inspired by the inception [42] and attention modules [43] with three stages, namely, *Channel-level color recovery*, *Color correction*, and *Dimension reduction* shown in Fig. 2 as follows.

2.1. Channel-level color recovery

As mentioned earlier, underwater images suffer from non-uniform light attenuation as the color light dissolves at different rates after crossing the surface of the ocean. According to [44], the colors attenuate based on their wavelengths, where the red color dissolves at a significantly faster rate than the green and blue colors, respectively. Thus to measure the color information loss, we proposed to extract features from each color channel separately and then adaptively weigh them to construct the enhanced image.

In addition to the non-uniform light attenuation, underwater images suffer from proportional color degradation. To address this issue, a multi-scale feature extractor is employed to detect the color degradations at various scales and regions within the image. As the degree of color degradation is channel dependent, we extract features from each channel separately. By fusing together these channel-based features with an attention mechanism in a supervised learning process, our model gains the ability to discern the degree of information loss in the spatial domain of the image with respect to its color channel. This enables the model to assign appropriate weights to capture the amount of information loss from each color channel, thereby enhancing the overall quality of the results.

As shown in Fig. 2, to capture the proportional degradations from each color channel at different scales effectively, we have conducted the well-known Inception (Inc) module [42], shown in Fig. 3-(a), for feature extraction. As seen, Inc allows us to represent the input image with structural feature maps at different scales (i.e., 1×1 (pixel-wise), 3×3, and 5×5), in addition to its contour information obtained by the *Max-Pooling* layer. However, the degradation ratios of the $3 \times 3, 5 \times 5$, and contour information are not equal. Thus, inspired by [43], by using the Channel-wise Attention Module (CAM), illustrated in Fig. 3-(b), we have weighed the extracted feature maps. The proposed strategy can further improve the enhancement performance and content representation by considering both the color channel-level and various structural scale quality degradation.

Table 1
The inception module architecture in both the channel-level color recovery and color correction stages.

Kernel	Stride	Padding	Output channels	Output
1 × 1	1	0	64	64 × 256 × 256
3 × 3	1	1	64	64 × 256 × 256
5 × 5	1	2	64	64 × 256 × 256
Max-Pooling	1	1	64	64 × 256 × 256
Concatenation	–	–	–	256 × 256 × 256

2.2. Color correction

In addition to the proportional degradation, the distortion of colors, such as a bluish or greenish appearance, is further caused by the attenuation of the light. Hence, the extracted feature maps require adaptive weighting to recover the real color. After concatenating the extracted feature maps in the second stage, we extract features from the combined color feature maps to capture the degradation at a higher level and weigh them using the CAM. In this stage, CAM helps the model to retrieve the lost color information and do the color correction accordingly by using an adaptive weighting of the red, green, and blue channel features.

2.3. Dimension reduction

To generate a RGB output image, it is essential to reduce the dimensionality of the extracted feature maps. In our study, as we maintained the spatial resolution of the input image throughout the feature extraction process, the focus is solely on how to reduce the number of features. To achieve this whilst retaining the restored lost information, we have introduced a gradual approach for dimension reduction. Actually, the proposed DICAM model decreases the dimensionality of the resulting high-dimensional feature map by progressively reducing the number of channels. Subsequently, a *Sigmoid* activation function is utilized to ensure that the intensity values of the enhanced image fall within the range of [0, 1]. It is noteworthy that, as illustrated in Fig. 2, a 3 × 3 convolution kernel is used in the initial step of dimension reduction. This can help to safeguard the preservation of information while decreasing the number of feature maps in the output dimension.

2.4. The network architecture

Table 1 describes the architecture of our Inc module. The convolutional layers with a kernel size of 3 × 3 and 5 × 5 and a stride of 1 within the Inc module enable the proposed architecture to capture the proportional degradation in UIs. Each convolution layer here is followed by a *LeakyReLU* activation function. Also, we have set the number of filters in the convolutional layers to 64 in both channel-level color recovery and color correction stages. Utilizing strides of 1 and 64 filters will make the Inc modules' output to have a size of 256 × 256 × 256 (filters, (i.e., 4 × 64) × H × W, where H and W are the height and width of the input image and 4 is the number of feature sets being concatenated, i.e. $1 \times 1, 3 \times 3, 5 \times 5$, and *Max-Pooling*). With 64 filters in each layer, the output of the concatenation layer will have 256 filters.

Inspired by the Convolutional Block Attention Module (CBAM) [45] and the SQUEEZE-AND-EXCITATION proposed [43], we have proposed the CAM for fast and effective refinement of the extracted feature maps based on their global statistical information, i.e. average value. Technically, the CAM maps each input feature to a single coefficient. To this end, it uses the Global Average Pooling (GAP) to initialize the coefficients. Then, as seen in Fig. 3.(b), CAM utilizes the EXCITATION strategy proposed in [43] to map GAP to the weighting coefficients. Specifically, the $GAP \in \mathbb{R}^M$ is generated by shrinking each feature map through its spatial dimension, i.e. $H \times W$ as follows.

$$GAP_m = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F_m(i,j), m = \{1, 2, \dots, M\} \quad (1)$$

where F_m is the mth feature map extracted using the Inc module.

To remove the linearity of the obtained global information and have a multiple-channel weighting rather than an ad-hoc representation, we reduced the dimension of the global information with a ratio of r, i.e. which was empirically set to 4, and then increased the dimension to the original one. Formally the final coefficient vector is calculated as follows:

$$Coeff_m = \psi(C_2 \psi(C_1 GAP_m)) \quad (2)$$

where ψ indicated the $Softsign$ function, and the $C_1 \in \mathbb{R}^{\frac{M}{r} \times M}$ and $C_2 \in \mathbb{R}^{M \times \frac{M}{r}}$ declare the coefficient of the fully connected layers used for dimension reduction and increase namely, SQUEEZE-AND-EXCITATION.

The dimension reduction helps the CAM to learn non-linear weighting coefficients and the dimension increase returns the determined non-linear weight coefficients to the channel dimension. Also, unlike those in [43], after both reducing and increasing the dimension we utilized the $Softsign$ activation function to make the obtained weight coefficients in [−1, +1]. This helps the model to also learn negative weights for better adjustment of the obtained feature maps.

Utilizing the CAM in the first stage of our proposed architecture can guide the network to weigh those feature maps that are more compatible with the input color channel. In other words, it would tune the extracted feature maps at different scales from color channels to capture the proportional degradations effectively. On the other hand, using CAM in the color correction stage would force the model to put more weight on feature maps corresponding to color channels with a higher rate of information loss. The aforementioned feature maps refinement is done by multiplying the learned coefficients with the input feature maps as follows:

$$F_m = F_m \otimes Coeff_m, m = \{1, 2, \dots, M\} \quad (3)$$

Compared with the original CBAM [45], CAM has fewer parameters to learn, due to the removed spatial attention module and simplified refinement by just using the average pooling, which decreases the time for learning and inference of the model and increases the quality of the enhanced image. Also, by utilizing the $Softsign$ instead of the $ReLU$ and $Sigmoid$ functions proposed in [43], it allows the network to also consider the negative weights on the feature maps for color correction and refinement. The proposed negative weighting can help to suppress certain color channels allowing the model to better balance and refine the colors. The negative weighting strategy not only reduces the apparent blue and green colors but also uses them more freely and effectively to produce realistic colors, leading to improvement in the appearance and realism of the generated image.

2.5. Training procedure

To train our proposed DICAM model, we have incorporated three loss functions, namely, *L1*, *Structural SIMilarity* (SSIM) [46], and *Perceptual* loss [47]. We indicate the ground-truth UI as G and the generated UI as $\hat{G}$. The total loss function is defined as follows:

$$\mathcal{L} = \mathcal{L}_{l_1} + \mathcal{L}_{SSIM} + \mathcal{L}_p \quad (4)$$

In most image-to-image translation models, it is common to utilize the pixel-level differences such as Mean Squared Error (MSE) and *L1* loss functions for training [37,48]. Since the values of *L1* loss are more compatible with the SSIM loss function, we have used it both directly and in the perceptual loss function as below:

$$\mathcal{L}_{l_1} = \|\hat{G} - G\|_1 \quad (5)$$

Table 2
Performance comparison of our DICAM with conventional UIE methods over the testing subset of the UIEB dataset.

Method	(a) full-reference IQA				(b) reference-free IQA			
	SSIM ↑	PCQI ↑	PSNR ↑	MSE ↓	UIQM ↑	UCIQE ↑	MSE_UIQM ↓	MSE_UCIQE ↓
CLAHE	0.8673	**0.9266**	18.71	0.0183	2.57	0.5271	0.3219	0.0032
ICM	0.8397	0.7008	18.77	0.0188	2.26	0.5160	0.4721	0.0030
DCP	0.7037	0.6244	14.23	0.0476	1.73	0.4659	1.0880	0.0117
Fusion-based	**0.8916**	0.8868	**22.10**	**0.0119**	2.51	**0.5599**	0.2186	0.0019
UDCP	0.5300	0.5797	11.05	0.0902	1.46	0.4534	1.4961	0.0130
Retinex-based	0.8199	0.8079	17.68	0.0207	**2.76**	0.5473	**0.1462**	**0.0013**
IBLA	0.7086	0.7054	15.72	0.0423	1.49	0.5118	1.9894	0.0050
ULAP	0.7578	0.7493	16.32	0.0331	1.66	0.5259	1.2022	0.0037
DICAM	**0.9375**	**0.9007**	**24.43**	**0.0060**	**3.06**	**0.5547**	**0.1188**	**0.0012**

Although the pixel-level difference-based loss functions make the model achieve a higher Peak Signal-to-Noise Ratio (PSNR) [49], they all suffer from blurriness artifacts and thus make the enhanced image details too smooth. Therefore, to tackle this issue and take higher-level information like structure and contrast into account, we have also used the SSIM [46] and *Perceptual* loss [47] functions.

More precisely, to compute the structural and textural similarity between the generated and ground-truth images, we have incorporated the SSIM similarity score as follows:

$$SSIM(x) = \frac{2.\mu_m.\mu_n + c_1}{\mu_m^2 + \mu_n^2 + c_1} \cdot \frac{2.\sigma_{mn}^2 + c_2}{\sigma_m^2 + \sigma_n^2 + c_2} \tag{6}$$

where m and n are 11×11 patches from the G and $\hat{G}$ images around each pixel x, respectively, μ, σ, and σ_{mn} denote the mean, standard deviation, and covariance, c_1 and c_2 are small positive parameters to stabilize the division. Considering that the SSIM computes the similarity between images, to use it in a minimization problem in which we intend to minimize the dissimilarity, the SSIM loss function can be written as follows:

$$\mathcal{L}_{SSIM} = 1 - \frac{1}{N}\sum_{i=1}^{N} SSIM(x_i) \tag{7}$$

To retain the high-frequency information of the image, inspired by [40,41,47], we have utilized the *Perceptual* loss. Technically, each time the generated UI and ground-truth image are separately fed to the pre-trained VGG19 network over the ImageNet dataset [50], and we compute the *L1* norm between the *relu_4_3* features obtained from the VGG network. Using *Perceptual* loss will help the model to preserve the high-level information of the image by measuring how close the extracted *relu_4_3* features from the enhanced image are to the ground-truth image.

3. Experimental results

We evaluate the performance of the proposed DICAM underwater image enhancement method over two publicly available datasets and compare its performance with state-of-the-art non-physical model-, physical model-, and deep learning-based methods. The compared non-physical model and physical model-based methods are CLAHE [11], ICM [14], DCP [29], fusion-based method [16], UDCP [30], Retinex-based method [17], IBLA [24], and ULAP [31]. Moreover, FUnIE-GAN [38], Water-Net [13], UIEC^2-Net [40], and WaveNet [41] are the compared deep learning-based methods. To make a fair comparison, the results are reported over the same testing subsets of each dataset. Our extensive experiments demonstrate the superiority of the proposed DICAM method in terms of reference (full-reference) and reference-free image quality metrics, histogram comparison measures, and run-time.

3.1. Datasets

To evaluate the performance of our proposed method, we conducted experiments on two publicly available underwater image enhancement datasets: Underwater Image Enhancement Benchmark (UIEB) [13] and Enhancing Underwater Visual Perception (EUVP) [38]. UIEB contains 950 UIs, of which 890 are with their corresponding ground-truth images, and 60 have no reference images. EUVP includes three types of UIs, including underwater dark, ImageNet, and scenes. EUVP has 5550 paired dark, 3700 ImageNet, and 2185 underwater scene images for training purposes. For validation purposes, it contains 570 dark, 1270 ImageNet, and 130 underwater scene images. Also, it has 515 test images containing images from all three categories.

3.2. Implementation details

To train the proposed model over each of UIEB [13] and EUVP [38] datasets, we resize the input images to size 256×256. The proposed DICAM model is then trained using the ADAM optimizer with a learning rate of 0.0008. The batch size and the number of epochs are set to 5 and 120 for both UIEB and EUVP datasets. These parameters are set and tuned experimentally based on the best-obtained results. Also, we implemented the proposed architecture in Pytorch deep learning framework on Nvidia Quadro RTX 6000/8000 GPU.

3.3. Evaluation metrics

We have employed both the full-reference and reference-free Image Quality Assessment (IQA) metrics with histogram comparison indicators for a fair and robust comparison. Reference-based methods use both the generated and ground-truth images to assess the quality, while the reference-free methods only need the generated image to compute its quality. The representative methods are well-known reference-based SSIM, Patch-based Contrast Quality Index (PCQI) [51], PSNR, and MSE and reference-free Underwater Image Quality Measure (UIQM) [52] and Underwater Color Image Quality Evaluation (UCIQE) [53] methods. Following [40,41,54,55], we initially calculated the metrics mentioned above for each image within the testing subset. Subsequently, we reported the mean value of each metric for comparison. The higher SSIM, PCQI, PSNR, UIQM, and UCIQE values and a lower MSE value indicate better performance. Technically, SSIM measures the similarity between the reference (ground-truth) and the generated images in terms of the luminance, contrast, and structural components. PCQI locally estimates the contrast quality. PSNR and MSE represent image quality by calculating the image's content corruption level. On the other hand, UIQM and UCIQE are reference-free measures designed for UIs. UIQM measures the UI's colorfulness, sharpness, and contrast. UCIQE quantifies the non-uniform color cast, blurring, and low-contrast characteristics by linearly combining the chroma, saturation, and contrast components of UIs.

Moreover, to compare the histograms of the ground-truth and generated UIs, we first convert the UIs from the RGB to the HSV color space, then compute the Kullback–Leibler (KL) divergence distance and the Chi-squared [56] statistics over each channel to see how close the histograms of the generated and ground-truth images are. Lower KL divergence and Chi-squared statistics values indicate better performance. For all metrics, we highlight the best- and second-best results in red and blue, respectively.

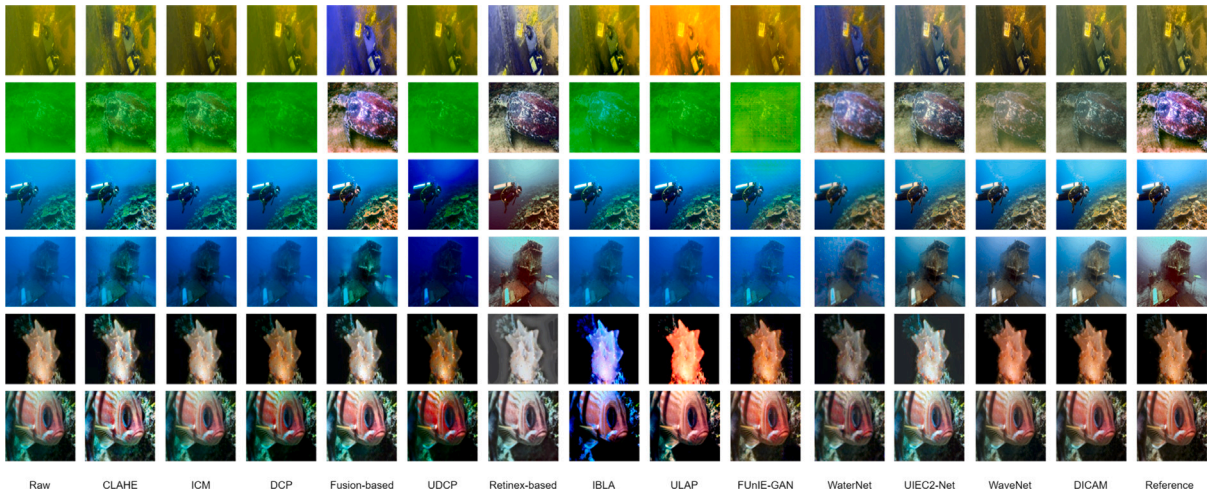

Fig. 4. Qualitative comparison of 6 underwater images taken from the UIEB and EUVP datasets. Images in the first four rows are from UIEB, and images in the fifth and sixth rows are from the EUVP dataset. From left to right: raw image, enhanced images of CLAHE, ICM, DCP, Fusion-based, UDCP, Retinex-based, IBLA, ULAP, FUnIE-GAN, Water-Net, UIEC^2-Net, WaveNet, DICAM, and the reference image.

Table 3
Performance comparison of our DICAM with deep learning-based UIE methods over the testing subset of the UIEB dataset.

Method	(a) full-reference IQA				(b) reference-free IQA			
	SSIM ↑	PCQI ↑	PSNR ↑	MSE ↓	UIQM ↑	UCIQE ↑	MSE_UIQM ↓	MSE_UCIQE ↓
FUnIE-GAN	0.8010	0.6256	18.14	0.0219	2.93	0.5084	0.2135	0.0034
Water-Net	0.8332	0.6194	20.69	0.0107	2.99	0.4726	0.1268	0.0033
UIEC^2-Net	**0.9215**	**0.8706**	**24.27**	**0.0054**	**3.10**	**0.5435**	0.1915	**0.0009**
WaveNet	0.9199	0.8491	23.30	0.0063	2.83	0.5406	**0.0892**	**0.0012**
DICAM	**0.9375**	**0.9007**	**24.43**	**0.0060**	**3.06**	**0.5547**	**0.1188**	**0.0012**

Table 4
Performance comparison of our DICAM with conventional UIE methods over the testing subset of the EUVP dataset.

Method	(a) full-reference IQA				(b) reference-free IQA			
	SSIM ↑	PCQI ↑	PSNR ↑	MSE ↓	UIQM ↑	UCIQE ↑	MSE_UIQM ↓	MSE_UCIQE ↓
CLAHE	0.8021	**0.8307**	17.41	0.0205	2.70	**0.5542**	**0.1071**	0.0029
ICM	0.8000	0.6982	**20.69**	**0.0110**	2.40	0.5317	0.1326	**0.0012**
DCP	0.7473	0.6434	17.58	0.0217	1.75	0.4668	0.9287	0.0027
Fusion-based	**0.8088**	**0.7924**	17.62	0.0204	2.56	**0.5705**	0.1220	0.005
UDCP	0.6197	0.5824	14.51	0.0452	1.64	0.4721	1.2137	0.0029
Retinex-based	0.7213	0.6864	15.96	0.0298	**2.92**	0.5461	0.2760	0.0033
IBLA	0.7573	0.7169	19.00	0.0225	1.74	0.5371	1.0899	0.0057
ULAP	0.7964	0.7439	19.62	0.0131	1.94	0.5373	0.7368	0.0015
DICAM	**0.9131**	0.7392	**25.13**	**0.0040**	**2.78**	0.5056	**0.0786**	**0.0003**

Table 5
Performance comparison of our DICAM with deep learning-based UIE methods over the testing subset of the EUVP dataset.

Method	(a) full-reference IQA				(b) reference-free IQA			
	SSIM ↑	PCQI ↑	PSNR ↑	MSE ↓	UIQM ↑	UCIQE ↑	MSE_UIQM ↓	MSE_UCIQE ↓
FUnIE-GAN	0.8718	0.7058	23.53	0.0061	2.70	**0.5265**	**0.0741**	0.0006
Water-Net	0.8048	**0.7662**	18.39	0.0186	2.84	0.4676	0.1495	0.0044
UIEC^2-Net	0.8310	**0.7520**	18.84	0.0159	**2.97**	**0.5518**	0.2481	0.0032
WaveNet	**0.8953**	0.6944	**24.67**	**0.0044**	**2.87**	0.5031	0.1296	**0.0005**
DICAM	**0.9131**	0.7392	**25.13**	**0.0040**	2.78	0.5056	**0.0786**	**0.0003**

3.4. Performance comparison

Table 2 reports the objective comparisons of the state-of-the-art non-physical model and physical model-based UIE methods and our proposed DICAM method over the testing subset of the UIEB dataset (800 images are randomly selected for training and 90 images are randomly selected for testing). According to the results, the proposed DICAM method yields the highest SSIM, PSNR, and UIQM values and the lowest MSE value. In terms of UCIQE, which measures the color cast, it has the second-best results. However, the subjective analyses show that the color cast of the images produced by DICAM is better than existing UIE methods. As shown in Fig. 4, compared with the enhanced images from the fusion-based method [16], which has the highest UCIQE score as shown in Table 2, the DICAM-generated enhanced images are much visually similar to the reference images with better color richness. Hence, it can be interpreted that UCIQE is not fully capable of quantifying the color distribution of UIs. Therefore, we separately computed the UIQM and UCIQE scores of both the reference and generated images and then reported their MSE in Table 2 to see how the UIQM and UCIQE scores of the generated images are close to their reference images. According to the obtained results, in terms of MSE of the UCIQE and UIQM, the proposed DICAM performs the

Table 6
Histogram comparisons of the Hue (H), Saturation (S), and Value (V) components of HSV color space in terms of the KL divergence and Chi-squared statistics on the testing subset of the UIEB dataset.

Method	(a) KL divergence			(b) Chi-squared		
	H-Channel ↓	S-Channel ↓	V-Channel ↓	H-Channel ↓	S-Channel ↓	V-Channel ↓
FUnIE-GAN	2.0447	1.3464	0.0024	0.4445	0.2888	**0.0005**
Water-Net	**0.9359**	**0.5233**	**0.0009**	**0.1727**	0.1396	**0.0005**
UIEC^2-Net	1.3825	0.5773	**0.0009**	0.2259	**0.1202**	**0.0005**
WaveNet	0.9879	0.5522	0.2914	0.2683	0.1423	**0.0078**
DICAM	**0.6703**	**0.3747**	**0.0014**	**0.1435**	**0.1064**	**0.0005**

Table 7
Histogram comparisons of the Hue (H), Saturation (S), and Value (V) components of HSV color space in terms of the KL divergence and Chi-squared statistics on the testing subset of the EUVP dataset.

Method	(a) KL divergence			(b) Chi-squared		
	H-Channel ↓	S-Channel ↓	V-Channel ↓	H-Channel ↓	S-Channel ↓	V-Channel ↓
FUnIE-GAN	**1.8400**	1.1268	0.0407	**0.1996**	**0.1356**	**0.0016**
Water-Net	6.1842	1.7728	0.0049	0.3726	0.2235	0.0023
UIEC^2-Net	7.9321	1.9995	**0.0047**	0.4298	0.2584	0.0023
WaveNet	2.2756	**0.9950**	0.0049	0.2878	0.2110	**0.0022**
DICAM	**2.0449**	**0.8058**	**0.0041**	**0.2526**	**0.1644**	**0.0016**

best, which means it is fully capable of enhancing the quality and color distribution of UIs. From Table 2, it is clear that the MSE_UIQM scores confirm the UIQM results. Therefore, it can be concluded that UIQM is more accurate than the UCIQE metric in assessing the quality of UIs.

Moreover, compared with the deep learning-based methods, as listed in Table 3, DICAM achieves the best results under SSIM, PCQI, PSNR, and UCIQE and second-best under MSE and UIQM metrics. Furthermore, as illustrated in Fig. 4, the DICAM-enhanced images are also better than the state-of-the-art deep learning-based UIE methods. It should be noted that we have used the trained model of FUnIE-GAN on the EUVP dataset to enhance the test images of the UIEB dataset.

Tables 4 and 5 present the quantitative comparison of our proposed DICAM model with the state-of-the-art non-physical model-, physical model-based, and deep learning-based methods on the EUVP dataset, respectively. Note that regarding the results of the UIEC^2-Net, due to computational complexity and a large number of training samples of the EUVP dataset, we could not train it over the training subset of the EUVP dataset. Hence, we used its publicly available trained model over a combination of the UIEB and a synthetic dataset called NYU-v2 RGB-D [57]. Regarding Water-Net, we could not train the model over the EUVP dataset, and since the model was only trained over UIEB, we used the same model to report the results on the EUVP dataset. It can be observed that the proposed DICAM model outperforms the traditional non-physical model- and physical model-based methods in terms of full-reference IQA metrics except the PCQI. Compared with deep models, it has the best results under the three compared full-reference IQA metrics. Based on the compared reference-free IQA metrics, DICAM has the second-best performance under the UIQM metric compared with the non-deep learning-based methods. For the UCIQE metric, DICAM achieves competitive results. For both the UIQM and UCIQE metrics, we again reported the mean MSE values for both metrics. From the results in Tables 4 and 5, it is clear that, like UIEB results, DICAM has a similar color cast to the reference images of the EUVP dataset as it yields the best MEAN_UCIQE results for both deep and non-deep learning-based methods. Also, for mean MSE values of UIQM (i.e., MEAN_UIQM), DICAM has the best and second-best results compared with non-deep and deep learning-based UIE methods, respectively.

3.5. Histogram comparison

As discussed earlier, despite the efforts made in objective quality evaluation of underwater images, existing IQA metrics, especially reference-free metrics, are not fully capable of assessing UIs' quality scores. As shown in Fig. 4 and obtained results in Tables 2 through 5, the predicted scores of reference-free IQA metrics are inconsistent with the visual perception of UIs. Therefore, in this study, we have incorporated the KL divergence and Chi-squared metrics for histogram comparison to evaluate the enhancement/color correction performance on the Hue, Saturation, and Value components of the HSV color space.

In Tables 6 and 7, we have compared and reported the histogram comparison results of deep learning-based methods on both UIEB and EUVP datasets. The obtained results in Table 6 show that DICAM has the best results in terms of both metrics for the Hue and Saturation components of UIEB images, and it has the second-best results for the Value component in terms of the KL divergence metric. In fact, in terms of KL divergence, DICAM improves the performance of the recently proposed WaveNet method on Hue, Saturation, and Hue by 31.70%, 34.98%, and 99.51%, respectively. Under the Chi-squared test, DICAM improves the WaveNet by 46.51%, 25.22%, and 93.58%, for Hue, Saturation, and Value. From Table 7, it can be seen that over the EUVP dataset, the proposed DICAM model's results are the best and at least the second-best for all three components and two metrics. This shows that compared with the FUnIE-GAN, which obtained the best results on Hue in terms of KL divergence and on Hue, Saturation, and Value components in terms of Chi-squared, DICAM is more consistent and stable. Furthermore, from Fig. 4, it can be also observed that the enhanced images of FUnIE-GAN suffer from blockiness distortions, which results in lower structural similarity with the reference images.

Table 8
Ablation study of the core modules of our DICAM enhancement network on the UIEB dataset.

Method	SSIM ↑	PCQI ↑	PSNR ↑	MSE ↓
No CAMs	0.9042	0.8262	20.27	0.0132
No color corrections	0.9313	0.8471	22.92	0.0074
Image-level Inception	0.9147	0.8509	23.17	0.0076
Image-level Inception w/o CAMs	0.8904	0.8205	20.49	0.0128
Inception + CBAM	0.9153	0.8289	23.78	0.0068
1 × 1 convolutions	0.9093	0.8282	22.78	0.0076
3 × 3 convolutions	0.9215	**0.8785**	24.38	0.0063
5 × 5 convolutions	0.9222	0.8748	24.01	0.0060
MaxPooling	0.8232	0.6300	21.75	0.0088
DICAM	**0.9364**	0.8705	**24.90**	**0.0058**

3.6. Ablation study

3.6.1. Component analysis

To further verify and analyze the effectiveness of the modules in the proposed architecture, ablation studies are conducted on the UIEB

Fig. 5. Three sample raw underwater images (a) and their enhanced versions (b)–(f) produced by different variations of the DICAM model, i.e (b) without the CAM modules, (c) without the color correction stage, (d) original DICAM with 60 epochs, (e) original DICAM with 120 epochs, and (f) the reference image.

Table 9
Cross-dataset evaluation results of our DICAM and deep learning-based UIE methods on the UIEB and EUVP datasets.

Method	(a) Training with UIEB				(b) Training with EUVP			
	SSIM ↑	PCQI ↑	PSNR ↑	MSE ↓	SSIM ↑	PCQI ↑	PSNR ↑	MSE ↓
FUnIE-GAN	–	–	–	–	0.8010	**0.6256**	18.14	0.0219
Water-Net	0.8048	**0.7662**	18.39	0.0186	–	–	–	–
UIEC^2-Net	**0.8310**	0.7520	**18.84**	**0.0159**	–	–	–	–
WaveNet	0.8235	**0.7939**	18.31	0.0176	**0.8541**	0.6218	**18.83**	**0.0166**
DICAM	**0.8390**	0.7516	**19.46**	**0.0143**	**0.8570**	**0.6613**	**18.75**	**0.0180**

dataset as it contains images with a complex and rich content of underwater sceneries and species. To this end, we removed the core modules (i.e., CAM module and color correction stage) of the DICAM network to analyze their role in performance results as follows: (a) All CAM modules are removed; (b) Color Correction stage is removed. In all compared variations of the DICAM model, we keep the Inc module for feature extraction with 60 training epochs. Given the resource constraints and the extended training time required for the proposed model, we opted to perform the ablation study using a reduced number of epochs, i.e. 60. Table 8, reports the performance results after removing the aforementioned modules in terms of the full-reference IQA methods to demonstrate the similarity of the produced enhanced images to their corresponding reference underwater images. The obtained results show that removing the CAM module can significantly decrease the performance of our DICAM method compared with the original version. Moreover, comparing the results of the DICAM version without the color correction stage, it seems that the color correction will only lead to a slightly dropped accuracy in comparison with the original DICAM version.

In addition to the previous analysis, to validate the effectiveness of the channel-level feature extraction, we have compared the results by applying the Inc module on the whole input image with and without the CAM module. The results in Table 8 demonstrate a lower performance when the features are extracted from the image-level, compared with our proposed DICAM with channel-level feature extraction. We have evaluated our DICAM model's performance by combining the inception module with the CBAM module to further verify the superiority of the proposed CAM module over the CBAM.

Also, to highlight the effectiveness of the Inc module, we evaluated the model's performance by only utilizing one of the branches within the module. The results indicate that the branch with 3×3 convolution layers achieves superior mean PCQI and PSNR values on the testing subset, while the branch with 5×5 convolution layers produces better mean SSIM and MSE values. However, it is noteworthy that their performance, especially concerning the SSIM metric, falls short of the DICAM with all branches used. This has verified the significance of combining these branches for improved feature extraction, leading to an overall better performance. This observation demonstrates the efficacy of multi-scale feature extraction in addressing the challenge of proportional color degradation in this context.

Furthermore, two sample images enhanced by the DICAM variations are illustrated in Fig. 5. As seen, the produced images from DICAM with all modules have the highest color richness, clearest visibility, and a more natural-looking appearance. Note that increasing the number of training epochs from 60 to 120 can significantly improve the color cast/richness in the image. As the color richness of the enhanced images is better than the reference image, it shows the effectiveness of the color correction stage and the generalization ability of the proposed DICAM model. Based on the obtained UCIQE scores as highlighted in Fig. 5, the generated enhanced images via the variations of the DICAM model achieve better performance compared with the raw image, demonstrating its capability to improve the color richness. Note that, although the enhanced images of the second and third samples in (d) seem to have better UCIQE scores, they are not visually as pleasant as the generated images using the model trained with 120 epochs. This is due to the limitations associated with the current underwater image quality metrics. Overall, the ablation study verifies the effectiveness of utilizing the CAM and color correction stages, which are essential for the adaptive weighting of the color channels and their extracted feature maps.

Table 10
Performance of the ablation study of the loss functions on the UIEB dataset.

$\mathcal{L}_{MAE}$	$\mathcal{L}_{SSIM}$	$\mathcal{L}_p$	SSIM	PCQI	PSNR	MSE
✓	×	×	0.9200	**0.8764**	24.03	0.0061
✓	✓	×	0.9317	0.8583	24.02	0.0062
✓	✓	✓	**0.9364**	0.8705	**24.90**	**0.0058**

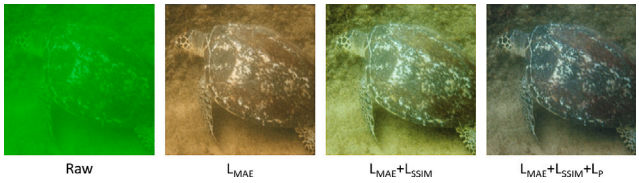

Fig. 6. Visual results of ablation study with different loss functions.

3.6.2. Loss function

To validate the efficacy of the incorporated loss functions, we assessed the performance of our method through training over 60 epochs, employing the following combinations of loss functions: (a) $\mathcal{L}_{MAE}$, (b) $\mathcal{L}_{MAE} + \mathcal{L}_{SSIM}$, and (c) $\mathcal{L}_{MAE} + \mathcal{L}_{SSIM} + \mathcal{L}_p$. It is crucial to note that excluding $\mathcal{L}_{MAE}$ adversely affects the model's performance, emphasizing its indispensability in our training procedure. Thus, we performed various training by combining the other loss functions with $\mathcal{L}_{MAE}$. Table 10 presents the results, indicating that the incorporation of $\mathcal{L}_{SSIM}$ and $\mathcal{L}_p$ (perceptual VGG loss) leads to a slight decrease in performance according to the PCQI metric. However, this trade-off is deemed necessary to enhance the visual perception of the model. Specifically, $\mathcal{L}_{SSIM}$ contributes to a 1.27% improvement in the SSIM metric, while the combination of $\mathcal{L}_p$ with the other two results in an additional 0.50% improvement in SSIM.

Beyond the detailed quantitative analysis provided earlier, we have visually demonstrated the impact of the specified combination of loss functions on a sample image and its enhanced version in Fig. 6. Evidently, the proposed combination balances color distribution and enhances the contrast of the image.

3.7. Cross-dataset evaluation

To validate the generalizability of the proposed DICAM model, we have employed the commonly used cross-dataset technique. We trained the DICAM and compared peer models on one dataset, then tested their performance on the other. Table 9 compares the results of the cross-dataset evaluation of four deep learning-based UIE models. As the trained models of FUnIE-GAN on the UIEB dataset, and UIEC^2-Net and Water-Net on the EUVP are unavailable, we could not report their corresponding cross-dataset performance. For the rest of the models, The column of *Training with UIEB* shows their testing results on the EUVP dataset when their models are trained on the UIEB. Also, The column of *Training with EUVP* reports the test results on the UIEB dataset when the models are trained on the EUVP dataset. The best and second-best results are highlighted in red and blue, respectively. Based on the results, DICAM has superior performance for seven of the eight compared indices, which demonstrates its generalizability and stability.

3.8. Run-time comparison

To evaluate the computational run-time of the DICAM model, we applied each method on the testing subset of the UIEB dataset, which has 90 test images, and reported the average run-time (i.e., divide the total run-time by the total number of test images). Table 11 lists the average run-time of the proposed and four compared deep learning-based UIE models. According to the results, the DICAM model obtains the second-best run-time, whereas the FUnIE-GAN method has the lowest run-time. However, in terms of IQA metrics, as discussed earlier, DICAM outperforms the method proposed in FUnIE-GAN.

Table 11
Average run-time comparison of DICAM and four deep learning-based UIE methods on the testing subset of the UIEB dataset.

Method	FUnIE-GAN	Water-Net	UIEC^2-Net	WaveNet	DICAM
Time (s)	**0.0014**	0.4839	0.0440	0.0573	**0.0248**
FLOPS (Billion (G))	3.591G	71.42G	26.16G	72.59G	53.13G

In addition to reporting the computational run-time, we have provided insights into the computational complexity of each model measured in terms of the Floating Point Operations Per Second (FLOPS). FLOPS offers an estimation of the floating-point operations required for the forward pass through the network. According to our analysis, the FUnIE-GAN model exhibits a lower FLOPS count, translating to faster run-time performance. Although the DICAM model has a significantly higher FLOPS count, it still secures the second-best in the compared models. This is attributed to the relatively lightweight nature of the incorporated modules, particularly when comparing Channel-wise Attention Module (CAM) to Convolutional Block Attention Module (CBAM).

4. Conclusion

In this paper, we proposed a novel UIE model, called DICAM, by addressing the proportional degradations and non-uniform color cast. To this end, we first use inception modules over each color channel to extract feature maps on three scales, then we weigh the extracted feature maps using the introduced CAM to capture the importance of degradations occurring in different ratios. Next, to refine the color distribution, we combine the extracted feature maps and apply the CAM to improve the color richness of the image. Obtained full-reference IQA (i.e., SSIM, PSNR, MSE) and measurements demonstrate the superior, effective, and accurate performance of the proposed DICAM model to produce enhanced images similar to the ground-truths on both datasets. Also, in terms of no-reference IQA, histogram comparisons, and run-time we got competitive results. In the future, potential works could be to develop models which can not only enhance the degradations caused by natural artifacts such as light attenuation but also can consider the distortions that occur during storing and transmission of images. Additionally, leveraging regional relationships [58,59] in information retrieval, coupled with content-oriented enhancement [60], can contribute to the production of improved images. This enhancement, in turn, has the potential to increase the accuracy of object detection methods. Moreover, it is vital to develop reference-free metrics to evaluate the quality of underwater images and compare the performance of the compared methods.

CRediT authorship contribution statement

Hamidreza Farhadi Tolie: Conceptualization, Data curation, Investigation, Methodology, Software, Visualization, Writing – original draft. **Jinchang Ren:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing – review & editing. **Eyad Elyan:** Investigation, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

We have only used the publicly available data to verify our findings.

Acknowledgments

This work is partially supported by the SeaSense project, funded by the Net Zero Technology Centre, UK.

References

[1] Y. Rzhanov, L. Linnett, R. Forbes, Underwater video mosaicing for seabed mapping, in: Proceedings 2000 International Conference on Image Processing (Cat. No.00CH37101), Vol. 1, 2000, pp. 224–227 vol.1, http://dx.doi.org/10.1109/ICIP.2000.900935.

[2] J. Tegdan, S. Ekehaug, I.M. Hansen, L.M.S. Aas, K.J. Steen, R. Pettersen, F. Beuchel, L. Camus, Underwater hyperspectral imaging for environmental mapping and monitoring of seabed habitats, in: OCEANS 2015 - Genova, 2015, pp. 1–6, http://dx.doi.org/10.1109/OCEANS-Genova.2015.7271703.

[3] M. Fulton, J. Hong, M.J. Islam, J. Sattar, Robotic detection of marine litter using deep visual detection models, in: 2019 International Conference on Robotics and Automation, ICRA, 2019, pp. 5752–5758, http://dx.doi.org/10.1109/ICRA.2019.8793975.

[4] Y. Xu, Y. Zhang, H. Wang, X. Liu, Underwater image classification using deep convolutional neural networks and data augmentation, in: 2017 IEEE International Conference on Signal Processing, Communications and Computing, ICSPCC, 2017, pp. 1–5, http://dx.doi.org/10.1109/ICSPCC.2017.8242527.

[5] M. Rajasekar, A.C. Aruldoss, M.A. Bennet, A novel method to detect corrosion in underwater infrastructure using an image processing, ARPN J. Eng. Appl. Sci. 13 (2018) 2556–2561.

[6] F. Ferreira, D. Machado, G. Ferri, S. Dugelay, J. Potter, Underwater optical and acoustic imaging: A time for fusion? a brief overview of the state-of-the-art, in: OCEANS 2016 MTS/IEEE Monterey, 2016, pp. 1–6, http://dx.doi.org/10.1109/OCEANS.2016.7761354.

[7] D. Berman, D. Levy, S. Avidan, T. Treibitz, Underwater single image color restoration using haze-lines and a new quantitative dataset, IEEE Trans. Pattern Anal. Mach. Intell. 43 (8) (2021) 2822–2837, http://dx.doi.org/10.1109/TPAMI.2020.2977624.

[8] G. Sequeira, V. Mekkalki, J. Prabhu, S. Borkar, M. Desai, Hybrid approach for underwater image restoration and enhancement, in: 2021 International Conference on Emerging Smart Computing and Informatics, ESCI, 2021, pp. 427–432, http://dx.doi.org/10.1109/ESCI50559.2021.9397058.

[9] R. Schettini, S. Corchs, Underwater image processing: State of the art of restoration and image enhancement methods, EURASIP J. Adv. Signal Process. 2010 (2010).

[10] S. Anwar, C. Li, Diving deeper into underwater image enhancement: A survey, Signal Process., Image Commun. 89 (2020) 115978, http://dx.doi.org/10.1016/j.image.2020.115978.

[11] K. Zuiderveld, Contrast limited adaptive histogram equalization, in: Graphics Gems IV, Academic Press Professional, Inc., USA, 1994, pp. 474–485.

[12] C.-Y. Li, J.-C. Guo, R.-M. Cong, Y.-W. Pang, B. Wang, Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior, IEEE Trans. Image Process. 25 (12) (2016) 5664–5677, http://dx.doi.org/10.1109/TIP.2016.2612882.

[13] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, D. Tao, An underwater image enhancement benchmark dataset and beyond, IEEE Trans. Image Process. 29 (2020) 4376–4389, http://dx.doi.org/10.1109/TIP.2019.2955241.

[14] K. Iqbal, R.A. Salam, A. Osman, A.Z. Talib, Underwater image enhancement using an integrated colour model., IAENG Int. J. Comput. Sci. 34 (2) (2007).

[15] K. Iqbal, M. Odetayo, A. James, R.A. Salam, A.Z.H. Talib, Enhancing the low quality images using unsupervised colour correction method, in: 2010 IEEE International Conference on Systems, Man and Cybernetics, 2010, pp. 1703–1709, http://dx.doi.org/10.1109/ICSMC.2010.5642311.

[16] C. Ancuti, C.O. Ancuti, T. Haber, P. Bekaert, Enhancing underwater images and videos by fusion, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 81–88, http://dx.doi.org/10.1109/CVPR.2012.6247661.

[17] X. Fu, P. Zhuang, Y. Huang, Y. Liao, X.-P. Zhang, X. Ding, A retinex-based enhancing approach for single underwater image, in: 2014 IEEE International Conference on Image Processing, ICIP, 2014, pp. 4572–4576, http://dx.doi.org/10.1109/ICIP.2014.7025927.

[18] S. Zhang, T. Wang, J. Dong, H. Yu, Underwater image enhancement via extended multi-scale retinex, Neurocomputing 245 (2017) 1–9, http://dx.doi.org/10.1016/j.neucom.2017.03.029.

[19] P. Zhuang, X. Ding, Underwater image enhancement using an edge-preserving filtering retinex algorithm, Multimedia Tools Appl. 79 (2020) 17257–17277.

[20] P. Zhuang, C. Li, J. Wu, Bayesian retinex underwater image enhancement, Eng. Appl. Artif. Intell. 101 (2021) 104171, http://dx.doi.org/10.1016/j.engappai.2021.104171.

[21] J. Zhou, X. Wei, J. Shi, W. Chu, W. Zhang, Underwater image enhancement method with light scattering characteristics, Comput. Electr. Eng. 100 (2022) 107898, http://dx.doi.org/10.1016/j.compeleceng.2022.107898.

[22] R. Liu, X. Fan, M. Zhu, M. Hou, Z. Luo, Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light, IEEE Trans. Circuits Syst. Video Technol. 30 (12) (2020) 4861–4875, http://dx.doi.org/10.1109/TCSVT.2019.2963772.

[23] R. Fattal, Single image dehazing, ACM Trans. Graph. 27 (3) (2008) 1–9, http://dx.doi.org/10.1145/1360612.1360671.

[24] Y.-T. Peng, P.C. Cosman, Underwater image restoration based on image blurriness and light absorption, IEEE Trans. Image Process. 26 (4) (2017) 1579–1594, http://dx.doi.org/10.1109/TIP.2017.2663846.

[25] L. Chao, M. Wang, Removal of water scattering, in: 2010 2nd International Conference on Computer Engineering and Technology, Vol. 2, 2010, pp. V2–35–V2–39, http://dx.doi.org/10.1109/ICCET.2010.5485339.

[26] H.-Y. Yang, P.-Y. Chen, C.-C. Huang, Y.-Z. Zhuang, Y.-H. Shiau, Low complexity underwater image enhancement based on dark channel prior, in: 2011 Second International Conference on Innovations in Bio-Inspired Computing and Applications, 2011, pp. 17–20, http://dx.doi.org/10.1109/IBICA.2011.9.

[27] H. Wen, Y. Tian, T. Huang, W. Gao, Single underwater image enhancement with a new optical model, in: 2013 IEEE International Symposium on Circuits and Systems, ISCAS, 2013, pp. 753–756, http://dx.doi.org/10.1109/ISCAS.2013.6571956.

[28] A. Galdran, D. Pardo, A. Picón, A. Alvarez-Gila, Automatic red-channel underwater image restoration, J. Vis. Commun. Image Represent. 26 (2015) 132–145, http://dx.doi.org/10.1016/j.jvcir.2014.11.006.

[29] K. He, J. Sun, X. Tang, Single image haze removal using dark channel prior, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 1956–1963, http://dx.doi.org/10.1109/CVPR.2009.5206515.

[30] P. Drews, E. Nascimento, F. Moraes, S. Botelho, M. Campos, Transmission estimation in underwater single images, in: 2013 IEEE International Conference on Computer Vision Workshops, 2013, pp. 825–830, http://dx.doi.org/10.1109/ICCVW.2013.113.

[31] W. Song, Y. Wang, D. Huang, D. Tjondronegoro, A rapid scene depth estimation model based on underwater light attenuation prior for underwater image restoration, in: R. Hong, W.-H. Cheng, T. Yamasaki, M. Wang, C.-W. Ngo (Eds.), Advances in Multimedia Information Processing – PCM 2018, Springer International Publishing, Cham, 2018, pp. 678–688.

[32] Y. Zhou, Q. Wu, K. Yan, L. Feng, W. Xiang, Underwater image restoration using color-line model, IEEE Trans. Circuits Syst. Video Technol. 29 (3) (2019) 907–911, http://dx.doi.org/10.1109/TCSVT.2018.2884615.

[33] M. Yang, A. Sowmya, Z. Wei, B. Zheng, Offshore underwater image restoration using reflection-decomposition-based transmission map estimation, IEEE J. Ocean. Eng. 45 (2) (2020) 521–533, http://dx.doi.org/10.1109/JOE.2018.2886093.

[34] Z. Liang, X. Ding, Y. Wang, X. Yan, X. Fu, GUDCP: Generalization of underwater dark channel prior for underwater image restoration, IEEE Trans. Circuits Syst. Video Technol. 32 (7) (2022) 4879–4884, http://dx.doi.org/10.1109/TCSVT.2021.3114230.

[35] M. Gao, F. Zheng, J.J. Yu, C. Shan, G. Ding, J. Han, Deep learning for video object segmentation: a review, Artif. Intell. Rev. 56 (1) (2023) 457–531.

[36] P. Ghamisi, B. Höfle, X.X. Zhu, Hyperspectral and LiDAR data fusion using extinction profiles and deep convolutional neural network, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10 (6) (2017) 3011–3024, http://dx.doi.org/10.1109/JSTARS.2016.2634863.

[37] P. Liu, G. Wang, H. Qi, C. Zhang, H. Zheng, Z. Yu, Underwater image enhancement with a deep residual framework, IEEE Access 7 (2019) 94614–94629, http://dx.doi.org/10.1109/ACCESS.2019.2928976.

[38] M.J. Islam, Y. Xia, J. Sattar, Fast underwater image enhancement for improved visual perception, IEEE Robot. Autom. Lett. 5 (2) (2020) 3227–3234, http://dx.doi.org/10.1109/LRA.2020.2974710.

[39] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: N. Navab, J. Hornegger, W.M. Wells, A.F. Frangi (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Springer International Publishing, Cham, 2015, pp. 234–241.

[40] Y. Wang, J. Guo, H. Gao, H. Yue, UIEC^2-Net: CNN-based underwater image enhancement using two color space, Signal Process., Image Commun. 96 (2021) 116250, http://dx.doi.org/10.1016/j.image.2021.116250.

[41] P.K. Sharma, I. Bisht, A. Sur, Wavelength-based attributed deep neural network for underwater image restoration, ACM Trans. Multimedia Comput. Commun. Appl. (2022) http://dx.doi.org/10.1145/3511021, Just Accepted.

[42] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2015, pp. 1–9, http://dx.doi.org/10.1109/CVPR.2015.7298594.

[43] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.

[44] S. Raimondo, C. Silvia, Underwater image processing: State of the art of restoration and image enhancement methods, Eurasip J. Adv. Signal Process. 2010 (7460252) (2010) http://dx.doi.org/10.1155/2010/746052.

[45] S. Woo, J. Park, J.-Y. Lee, I.S. Kweon, CBAM: Convolutional block attention module, in: V. Ferrari, M. Hebert, C. Sminchisescu, Y. Weiss (Eds.), Computer Vision – ECCV 2018, Springer International Publishing, Cham, 2018, pp. 3–19.
[46] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[47] J. Johnson, A. Alahi, L. Fei-Fei, Perceptual losses for real-time style transfer and super-resolution, in: B. Leibe, J. Matas, N. Sebe, M. Welling (Eds.), Computer Vision – ECCV 2016, Springer International Publishing, Cham, 2016, pp. 694–711.
[48] P. Kumar, S. Priyankar, J.A. Sur, Scale-aware conditional generative adversarial network for image dehazing, in: 2020 IEEE Winter Conference on Applications of Computer Vision, WACV, 2020, pp. 2344–2354, http://dx.doi.org/10.1109/WACV45572.2020.9093528.
[49] I. Avcibas, B. Sankur, K. Sayood, Statistical evaluation of image quality measures, J. Electron. Imaging 11 (2) (2002) 206–223.
[50] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, ImageNet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248–255, http://dx.doi.org/10.1109/CVPR.2009.5206848.
[51] S. Wang, K. Ma, H. Yeganeh, Z. Wang, W. Lin, A patch-structure representation method for quality assessment of contrast changed images, IEEE Signal Process. Lett. 22 (12) (2015) 2387–2390, http://dx.doi.org/10.1109/LSP.2015.2487369.
[52] K. Panetta, C. Gao, S. Agaian, Human-visual-system-inspired underwater image quality measures, IEEE J. Ocean. Eng. 41 (3) (2016) 541–551, http://dx.doi.org/10.1109/JOE.2015.2469915.
[53] M. Yang, A. Sowmya, An underwater color image quality evaluation metric, IEEE Trans. Image Process. 24 (12) (2015) 6062–6071, http://dx.doi.org/10.1109/TIP.2015.2491020.
[54] K. Yan, L. Liang, Z. Zheng, G. Wang, Y. Yang, Medium transmission map matters for learning to restore real-world underwater images, Appl. Sci. 12 (11) (2022).
[55] A. Pipara, U. Oza, S. Mandal, Underwater image color correction using ensemble colorization network, in: 2021 IEEE/CVF International Conference on Computer Vision Workshops, ICCVW, 2021, pp. 2011–2020, http://dx.doi.org/10.1109/ICCVW54120.2021.00228.
[56] Y. Rubner, C. Tomasi, L.J. Guibas, The earth mover's distance as a metric for image retrieval, Int. J. Comput. Vis. 10 (2000) 99–121.
[57] N. Silberman, D. Hoiem, P. Kohli, R. Fergus, Indoor segmentation and support inference from RGBD images, in: A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, C. Schmid (Eds.), Computer Vision – ECCV 2012, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 746–760.
[58] Y. Liu, D. Zhang, Q. Zhang, J. Han, Part-object relational visual saliency, IEEE Trans. Pattern Anal. Mach. Intell. 44 (7) (2021) 3688–3704.
[59] Z. Shao, J. Han, D. Marnerides, K. Debattista, Region-object relation-aware dense captioning via transformer, IEEE Trans. Neural Netw. Learn. Syst. (2022).
[60] Z. Shao, J. Han, K. Debattista, Y. Pang, Textual context-aware dense captioning with diverse words, IEEE Trans. Multimed. (2023).

Hamidreza Farhadi Tolie is currently a Ph.D. student at National Subsea Centre, Robert Gordon University, UK. He got his B.Sc. degree in Information Technology, in 2018, and his M.Sc. degree in Computer Science, in 2021, both from the Institute for Advanced Studies in Basic Sciences (IASBS). His research interests include, image quality assessment, image enhancement, image processing, and computer vision.

Jinchang Ren (Senior Member, IEEE) received the BEng in Computer Software in 1992, MEng in Image Processing and Pattern Recognition in 1997 and DEng in Computer Vision in 2000, all from Northwestern Polytechnical University, Xi'an, China. He also received a Ph.D. degree in electronic imaging from the University of Bradford, Bradford, U.K., in 2009. He is currently a Professor of computing science and Transparent Ocean Lead in National Subsea Centre, Robert Gordon University, Aberdeen, U.K. He has published over 380 research papers. His research interests include hyperspectral imaging, image processing, computer vision, big data analytics, and machine learning.

Eyad Elyan received the degree in computer science from Al-Quds University, Palestine, in 1999, the M.Sc. degree in software engineering from Bradford University, U.K., in 2004, and the Ph.D. degree from Bradford University, for his work on 3D facial modeling and recognition, in 2008. He is currently a Professor in machine learning and computer vision with the School of Computing, Robert Gordon University. His research interests include machine learning, deep learning, applied computer vision, and ensemble learning. He is a Fellow Member of the British Higher Education Academy. He works with the Scotland Data Laboratory, Innovation Centre Ambassador.