Computers & Graphics (2022)

Contents lists available at ScienceDirect

Computers & Graphics

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

Underwater enhancement based on a self-learning strategy and attention mechanism for high-intensity regions

Claudio D. Mello Jr.[a,*], Bryan U. Moreira[a], Paulo J. D. O. Evald[a], Paulo J. L. Drews Jr.[a], Silvia S. C. Botelho[a]

[a] *Center for Computer Science, Federal University of Rio Grande*
Av. Italia - km8, Rio Grande, RS, 96203-900, Brazil

ARTICLE INFO

Article history:
Received August 4, 2022

Keywords: Self-supervised learning, underwater, enhancement.

ABSTRACT

Images acquired during underwater activities suffer from environmental properties of the water, such as turbidity and light attenuation. These phenomena cause color distortion, blurring, and contrast reduction. In addition, irregular ambient light distribution causes color channel unbalance and regions with high-intensity pixels. Recent works related to underwater image enhancement, and based on deep learning approaches, tackle the lack of paired datasets generating synthetic ground-truth. In this paper, we present a self-supervised learning methodology for underwater image enhancement based on deep learning that requires no paired datasets. The proposed method estimates the degradation present in underwater images. Besides, an autoencoder reconstructs this image, and its output image is degraded using the estimated degradation information. Therefore, the strategy replaces the output image with the degraded version in the loss function during the training phase. This procedure *misleads* the neural network that learns to compensate the additional degradation. As a result, the reconstructed image is an enhanced version of the input image. Also, the algorithm presents an attention module to reduce high-intensity areas generated in enhanced images by color channel unbalances and outlier regions. Furthermore, the proposed methodology requires no ground-truth. Besides, only real underwater images were used to train the neural network, and the results indicate the effectiveness of the method in terms of color preservation, color cast reduction, and contrast improvement.

© 2022 Elsevier B.V. All rights reserved.

1. Introduction

Modern underwater (UW) activities such as monitoring, inspection, maintenance, archaeology, and environmental research, involve the acquisition of video footage and images of objects, fauna and flora [1, 2]. The quality of perception of these objects in the scene depends on physical properties of water, ambient light, and depth [3]. Turbidity is defined by the particles of organic and inorganic materials in suspension in water. They cause scattering and absorption of light rays causing blurred, and dim images [4, 5]. The turbidity and distance of the objects from the camera define the intensity of these phenomena. The depth of the scene affects the color perception, and the light components with longer wavelength are the first to be attenuated [6]. The ambient light interacts with the particles in suspension in the water, and increases the turbidity perception in the image acquisition. Also, the nature of the material in suspension defines a brownish, greenish or blueish tonality or the *color cast* of the water [7].

⋆

*Corresponding author: Tel.: +55-053-3293-5106; fax: +55-053-3293-5105;
e-mail: claudio.mello@furg.br (Claudio D. Mello Jr.)

The UW image restoration methods describe the image using the Image Formation Model (IFM) [8, 5]. The IFM represents the scene captured by the camera at a given depth, considering water attenuation and ambient light. The UW light attenuation depends on geographic and environmental conditions of the water [9], and produces diversity in the image scenes. The variability of turbidity, ambient light, and color cast represent a relevant challenge to UW image improvement tasks, given that the attenuation coefficients, distance of the objects from the camera and background light are unknown in most of the real image acquisition situations.

Image enhancement proposals have explored the improvement of the contrast and color, focusing on pixel intensity redistribution via spatial-domain [10], transform-domain [11] or Convolutional Neural Network (CNN)-based image enhancement [12]. Recently, mixed approaches involving enhancement/restoration (dehazing) and IFM-Deep Learning (DL) based methods were presented in [13] and [14]. Also, strategies based on DL and multi-color spaces with adaptive frameworks were used in [15] and [16].

The improvement of the UW images based on DL is a difficult task, because the real reference image (*real* ground-truth) is, normally, unavailable. Therefore, it is common the authors to resort to synthesized reference images. The strategies used for synthesizing ground-truth images are physical model-based techniques [13] or more complex structures performing style transfer [17, 18], and transfer-domain based on the Generative Adversarial Networks (GANs) [19]. Recent approaches based on autoencoders use synthetically paired datasets, which was built and made available in [12, 14]. Besides, these datasets contain paired images from different water types, as well as additional information, such as: depth scene maps and water attenuation factors. Thereby, specific procedures and prior image processing are required in order to obtain this additional information.

In these restoration approaches, the quality of the resulting reference images depends on the ability of the synthesizing method. A good enough approximation of these synthesized images to the real non-degraded one's impacts on the resulting improved images. The dehazed reference image cannot be obtained in complex UW environments, unless standard color boards are taken into the UW scene [20]. The true intensity and hue of the colors of the objects in the UW scene are unknown in most cases. In the absence of real ground-truth, the quality of the results is defined mainly from the subjective analysis of visual perception. The above considerations show that UW image restoration is a challenging task, and an ill-posed problem.

On the other hand, a degradation action on the image can be more easily performed. The UW images tend to present some level of degradation that depends on the environmental conditions in which it were acquired. The degradation itself consists in available information, and it can be estimated and used to increase the degradation of the image [18, 21]. Unlike restoration, which implies an approximation of the real scene, the degradation task starts from a known condition, and the degraded image can even be evaluated in comparison to the original one.

Deep learning methods for UW image improvement are consistently represented by GAN-based proposals, particularly those involving the style-transfer and transfer-domain approaches. The assumption that some UW images present no degradation level is assumed in the transfer-domain methodologies [22, 18, 19]. However, the selection of the images is based on subjective analysis, focused on scenes with objects very close to the camera, and limited or absent background scenarios. This approach can lead to a reduced generalization ability of the method. Style-transfer approaches use outdoor or indoor images in order to train neural networks. Meanwhile, the chosen outdoor/indoor images should bear some feature-similarity to the UW images. This issue improves generalization ability and stability when training neural networks. Despite these issues, the methods also show complex network structures with moderate to high computational cost, requiring synthetic paired datasets, and elaborated training procedures [18] [23].

In this study, we propose a self-supervised learning strategy for UW image enhancement. The method is based on a single-image architecture using an autoencoder. The proposed algorithm does not require a ground-truth, and the basic conception comprises the teaching of a network to enhance images from a harsher penalty. This harsher penalty results from the replacement of the output image in the loss function with a synthetically more degraded version. The main assumption considers that a degradation method that captures the degradation nature, and intensifies it, tends to generate more realistic images than a restoration method. In addition, we present an attention mechanism to reduce saturated regions in the enhanced image. These regions surge in images with a strong unbalance of the color channel, and regions with high-intensity pixels related to the average value of the image.

The main contributions of this paper are:

1. We propose an effective enhancement method for UW images based on a small-sized neural network, a small dataset, and a quite simple algorithm;
2. We present an alternative approach to enhance natural UW images based on degradation content of the image using a self-supervised learning strategy. The algorithm uses the degradation estimated from the input image to *mislead* the training of an autoencoder to compensate for the degradation, without pre-processing of the image or synthesized ground-truths;
3. We propose an attention mechanism oriented to limit high intensity regions generated in the enhanced images when the original one presents strong unbalance in color channels and outlier pixels.

In the next sections, we will describe our method and results. In Section 2, the related works are discussed. Section 3 presents the framework for image enhancement and the degradation algorithm. In Section 4, the experimental results are presented, and Section 5 is dedicated to the conclusion.

2. Related works

In recent years, the enhancement of the UW images has experimented strong evolution [24, 20]. Most of the presented

works is related to deep learning methodologies. The main issues tackled by these works are color shift and contrast reduction on images.

The IFM-based enhancement method for UW images presented in [25] uses statistical models to estimate the background light from a manually annotated database, and histogram distributions of the UW images. The transmission and scene depth maps were obtained via a variant of the UW Dark Channel Prior (UDCP) [5], and it was applied to modify the color channel transmission map.

A method for adaptation of color and contrast enhancement based on digital image processing was presented in [26]. The algorithm used Gaussian and bilateral filtering to decompose the image into high and low frequency components. The minimization of the difference between the guided image obtained from the low frequency component, and the output image. The enhanced image is combined with the soft-thresholded high frequency component to compose the output image. The method presented effectiveness in the enhancement task. However, it showed significant sensitivity to the parameters for the minimization procedure. The previous selection of the parameters is performed empirically.

IFM-free enhancement methods do not require specialized UW conditions or scene depth information. In [4], an enhancement method based on the fusion technique was presented. The method fuses color corrected and contrast-enhanced versions of the input image to obtain the output image. Despite the effective results, some images show over-enhancement and unnatural colors. In [22], an improvement of the model shown in [27] was presented, where the Underwater GAN (UGAN) was trained to generate turbid images from clear ones. A synthetic dataset is produced, and used to train another GAN. During inference, the generator network predicts clear images from blurry images as input. These methods present unrealistic color correction due to the lack of true colors in the datasets for specific UW scenarios.

The study presented in [18] introduced the Underwater Denoising Autoencoder (UDAE) model based on the U-Net architecture [28]. Underwater images are presented to the CycleGAN generative model [29], for style-transfer between clean and distorted UW images, and to generate a paired UW image dataset. UDAE is trained to restore the image colors. This method depends on the diversity in nature and intensity of the degradation, present in the selected images.

In [19], a conditional GAN-based model (FunieGAN) conceived for real-time UW image enhancement was presented. The transfer-domain method developed in [22] was used to generate synthetic paired dataset. Then, this dataset was used to train the model. The authors reported a loss of effectiveness for enhancing texture-less and low-contrast images. Also, it was reported that the model is prone to instability during training.

In a different way, a hybrid approach for UW image restoration with edge-enhancement was presented in [7]. This method uses a CNN to estimate the IFM transmission map of UW images. In addition, a white balance strategy was developed to remove the color cast, and the non-subsampled Contourlet Transform was used to perform denoise and edge enhancement. The neural network was trained using clear UW images, and degraded versions obtained from IFM. However, this algorithm requires prior estimation of background light and filtering in a post-processing step.

A model based on UW scene priors was introduced in [21]. The IFM was used to generate synthetic UW images and a paired dataset was built. This resulting dataset contains images of several water types, as well as degradation levels, and it was used to train a lightweight neural network to enhance UW images and videos. However, the resulting images required post-processing in order to restore the dynamic range and colors of the images. Furthermore, in [14], an approach which combines CNN and IFM was developed. The method divides the image restoration process into two stages, the horizontal one, which embeds the IFM in the neural network, and the vertical one, which restores the image distortion caused by vertical absorption of light. This method requires prior processing of the images in order to estimate the physical parameters, and the image with vertical distortion.

In [12], it was presented a model that fuses different feature maps during image representation learning. A synergistic pooling mechanism was used to extract channel-wise attention maps to derive the locally weighted features. The model focused on features related to degraded patches in the UW image in order to improve these patches. Synthetic paired dataset were used to train neural network. The loss function presented a feature loss component, which required previous training of the neural network for feature mapping. In [30] was shown an unsupervised learning methodology based on an encoder-decoder architecture. The neural network presents a double encoder section, and the method estimates the degradation of the image to drive the training. The resulting images show a strong enhancement of the ambient light. Color restoration and dehazing action are limited. Finally, an encoder-decoder neural network integrating different color spaces was developed in [15]. The algorithm uses attention methodology to aim the features' extraction from the images in the RGB, HSV, and Lab color spaces. A transmission map was estimated and used to drive network learning to enhance UW images. The method showed effectiveness, and good performance in the enhancement task. The paired UIEB dataset [21] was used in the training of the neural network. Difficulties in enhancing images in limited lighting were reported by the authors.

3. Methodology

The perception of the intensities and hues of the real colors is unknown at higher depths, and under limited illumination conditions. In addition, water turbidity generates a loss of contrast and blurring. Methodologies for image improvement based on CNN and CNN-IFM use synthetic paired datasets to achieve or to complement network training.

We propose a single-image architecture for UW image enhancement. Our proposal explores the degradation content in the image to drive the enhancement task. The concept relies on the assumption that the UW image presents some level of degradation that can be estimated, and used to drive its reduction or

removal from the image. This synthetic degradation action will intensify the degradation level present in the original image. Then, this degraded image will be adopted as the output image of the neural network in the loss function during the training stage. This "trick" will cause a higher penalty in the loss function of a neural network during the training step. Therefore, the neural network will learn to remove the additional degradation by allowing the enhancement of the original image.

The proposed methodology uses a fully convolutional autoencoder to perform image enhancement from the described strategy that requires neither synthetic ground-truth nor priors of the image. The following subsection describes the conceptual architecture and the proposed algorithm.

3.1. Proposed framework for image enhancement

Essentially, an autoencoder learns to reconstruct its input information, and to present an identical version of the output. We propose to insert a degradation function D over the training stage that increases the distortion present in the output image of the autoencoder. The conceptual structure of the model is shown in Fig. 1.

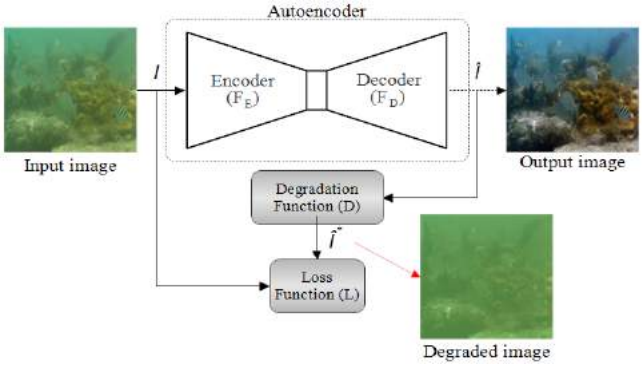

Fig. 1. Diagram of the conceptual structure of the proposed method.

The resulting image $\hat{I}^*$ from D is a degraded version of the output image $\hat{I}$. The degraded image $\hat{I}^*$ replaces the output image in the loss function. The neural network interprets that its output is not good enough, and learns to correct it. However, effective correction occurs if the degradation is feature-coherent in nature and intensity with those present in the input image.

The proposed methodology requires a lower level of supervision compared to peering approaches, since the input image acts as a reference image, not requiring a paired dataset.

3.2. Degradation Function

The distortion must be coherent in nature and intensity with the original image. According to Fig. 1, the degraded image $\hat{I}^*$ can be described by

$$\hat{I}^*(x,y) = \hat{I}(x,y) + \Delta(x,y), \tag{1}$$

where $\hat{I}$ represents the output image of the autoencoder, and Δ represents the imposed degradation over $\hat{I}$ by the degradation function D. Assuming that a non-distorted image I_c can be defined by removing the distortion Δ from a degraded image I^*; then, it follows

$$I_c(x,y) = I^*(x,y) - \Delta(x,y), \tag{2}$$

isolating Δ in (2), and using in the (1), generalizing with the representation of the color channel (k), we have:

$$\hat{I}^*_{(k)}(x,y) = \hat{I}_{(k)}(x,y) + I^*_{(k)}(x,y) - I_{c(k)}(x,y), \tag{3}$$

where $k \in \{R, G, B\}$.

The expression in (3) summarizes the degradation function. The image $\hat{I}^*_{(k)}$ corresponds to the degraded image resulting from the degradation function D in Fig. 1. The $I^*_{(k)}$ and $I_{c(k)}$ images are degraded and stretched histogram versions of the input image, respectively. Essentially, the difference between these images in (3) defines the degradation action. It implies that degradation added to the $\hat{I}_{(k)}$ is obtained from two opposite interpretations. First, the image $I^*_{(k)}$ must provide additional degradation-related features that will be added to $\hat{I}^*_{(k)}$.

Second, $I_{c(k)}$ must contain features that improve image quality, and show a path of decreasing degradation. At the same time, the degradation procedure increases the low quality-related features, and decreases the high quality-related features of the image. In the next sections will be presented the procedures to calculate these images. The calculations of these images use an image description inspired by the IFM, and are described in RGB space color.

3.3. Description of the image in the RGB color space

The image represented by the IFM has embedded elements that describe the image with its related features. An algorithm that uses this model needs to estimate the scene reflectance, water attenuation coefficient, and background light for the image description [9, 8].

The proposed algorithm imposes distortion in a coherent manner avoiding over or under degradation. However, the method does not perform any prior processing of the images, and the unique information available is the input image. We adopted an image description contextualized in the color space, but inspired by the physical model presented in the IFM. This description is presented in (4), and it is used in the estimation of the $I^*_{(k)}$, as shown in (3).

$$I_{(k)}(x,y) = I_{J(k)}(x,y)e^{-gd_{(k)}(x,y)} + (1 - e^{-gb_{(k)}(x,y)}).\Lambda_{(k)}, \tag{4}$$

where $I_{(k)}$ is an UW image, $I_{J(k)}$ represents the unknown UW scene without distortion, and $\Lambda_{(k)}$ is defined as the context luminosity of the image. The $gd_{(k)}$ and $gb_{(k)}$ parameters represent turbidity-distance factors related to the water attenuation of the light from the objects in the scene and the ambient light, respectively. These parameters are obtained from the input image in a pixel-wise context. In order to calculate them, we analyzed the characteristics of the UW images in color space under variable turbidity. To this end, we performed a study on the images from the TURBID Dataset [31]. This dataset contains three sets of real UW images with progressive turbidity: Milk, Deepblue,

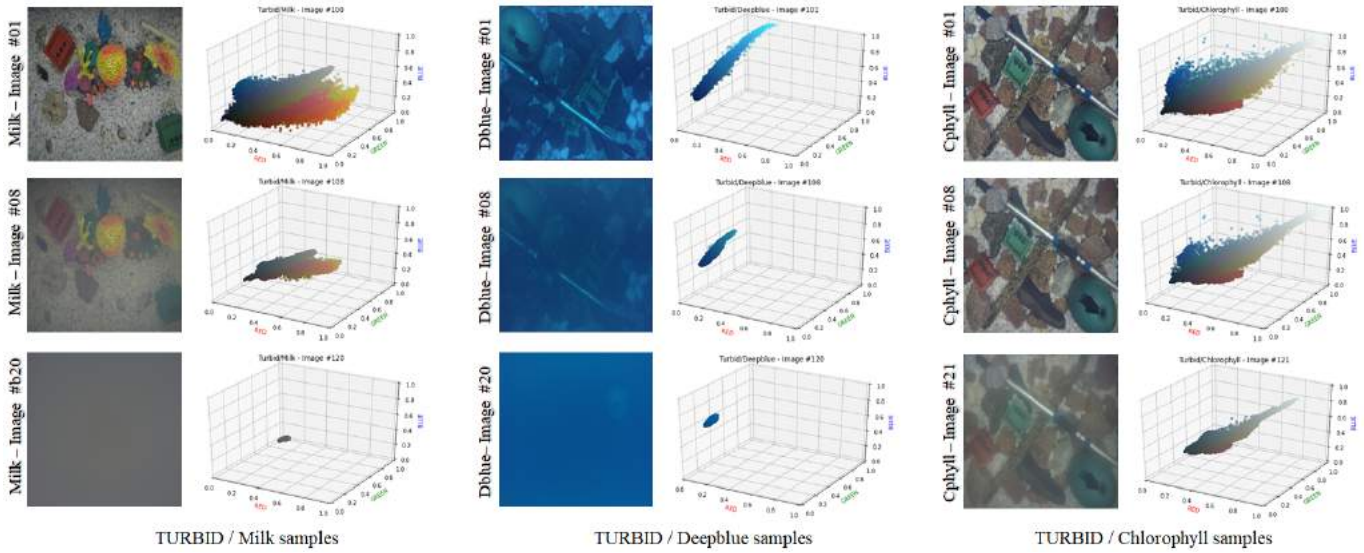

Fig. 2. Samples of the images from TURBID dataset, and respective pixel distribution in the RGB color space.

and Chlorophyll. The Milk subset has 20 images; the Deepblue and Chlorophyll subsets have 21 images.

In order to facilitate the analysis of the images, we define the *dynamic range* $b_{w(k)}$ of the image as the difference between the maximum and minimum values by color channel as defined in (5). The maximum value is represented by $max(\cdot)$, and the minimum value by $min(\cdot)$.

$$b_{w(k)} = max(I_{(k)}(x, y)) - min(I_{(k)}(x, y)). \quad (5)$$

3.4. Effects of the turbidity in UW images

Samples of the images from Milk, Deepblue, and Chlorophyll datasets are shown in Fig. 2. Also, it presents the distribution of the pixels in the RGB color space for each image. The first perception is image surface shrinkage in the color space with increasing turbidity. The *dynamic range* of the image decreases when the turbidity increases. Specifically, the maximum value is reduced due to forward scattering, and absorption phenomena generated by turbidity. In contrast, the minimum value increases due to the interaction between ambient light and turbidity. This effect increases the perception of turbidity in the image captured by the camera. In Fig. 2, the *dynamic range* decreases, and the scene converges to the background light, then turbidity perception dominates the image. This is perceptible in all sets of images but most evident in the Milk samples.

The behavior of the minimum and maximum values is shown in Fig. 3 for the TURBID dataset. The horizontal axis indicates images in ascending order of turbidity.

The Milk and Chlorophyll datasets show a balanced color cast and brown-greyish coloring. The effect on the maximum and minimum values presents a uniform variation related to turbidity. The Deepblue dataset presents a strong bluish color cast that accelerates the attenuation of the maximum values, mainly the red and green channels. The backscattering phenomenon is the main cause of the turbidity perception in UW images [8], and this effect is increased by the color cast.

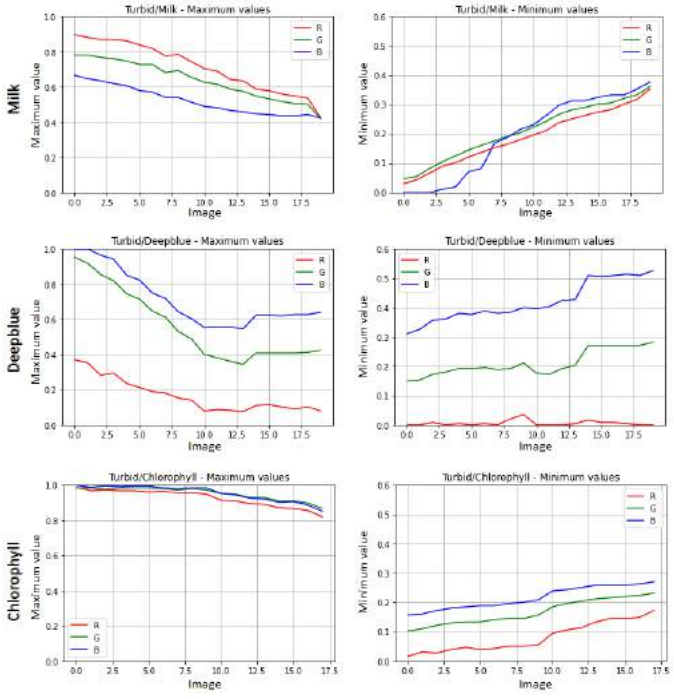

Fig. 3. Variation of the minimum and maximum values with turbidity for TURBID dataset.

3.5. Parameters for description of the image in the color space

The distorted image generated by degradation function, shown in (3), must have a narrower *dynamic range*. In other words, lower maximum and higher minimum values related to the input image (UW image). This conception is in accordance with the discussion above.

The degradation to be added to the output image of the autoencoder is described by the estimation of the $gd_{(k)}$, $gb_{(k)}$, and $\Lambda_{(k)}$ in (4). The perception of turbidity in the UW image is proportional to the pixel intensity. Thus, the degradation imposed on the pixel must be proportional to the pixel intensity.

We used the interaction between the unused interval available to the *dynamic range* [0, 1], and the information content inside the *dynamic range* ($b_{w(k)}$). Higher turbidity implies lower $b_{w(k)}$, and less information available in the image. Thus, the parameter $gd_{(k)}$ defines the turbidity effects over the reflectance of the objects in the scene. It is defined as indicated in (6). The term $[max(I_{(k)}) - I_{(k)}].[I_{(k)} - min(I_{(k)})]$ concentrates the degradation action within the *dynamic range*. It prevents the formation of saturated or zero-intensity regions in the degraded image, and loss of information. This parameter is described in a pixel-wise context, but the pixel coordinates (x,y) are omitted in (6) for clarity concerns.

$$gd_{(k)} = \frac{1 - b_{w(k)}}{b_{w(k)}}[max(I_{(k)}) - I_{(k)}].[I_{(k)} - min(I_{(k)})]. \quad (6)$$

Similarly, the parameter $gb_{(k)}$ is calculated from the interaction between the unused interval available to the *dynamic range* [0, 1], the interval $b_{w(k)}$ and the Context Luminosity information $[max(I_{(k)}) - I_{(k)}].[\Lambda_{(k)} - min(I_{(k)})]$ inside of the *dynamic range*. This concept is inspired by the revised image formation model, presented in [8], where the attenuation factor of the water has different dependencies in the scene and ambient light components in the revised physical model. Thus, $gb_{(k)}$ defines the backscattering effect due to ambient light, and its calculation is expressed in (7). The pixel coordinates (x,y) are also omitted for clarity.

$$gb_{(k)} = \frac{1 - b_{w(k)}}{b_{w(k)}}[max(I_{(k)}) - I_{(k)}].[\Lambda_{(k)} - min(I_{(k)})]. \quad (7)$$

The Context Luminosity ($\Lambda_{(k)}$) is estimated based on the background light definition presented in [8]. Considering the turbidity dominating the scene of an image, in this case, $gd_{(k)} \rightarrow \infty$, and there is no perception of the objects. This event corresponds to $b_w(k) \rightarrow 0$, and both color channel median and average values of the image converge to the center of the *dynamic range*. From the Gray World Theory [32], the average color in an image with a uniform color distribution can be calculated by the average intensity of the pixels. However, UW images can present irregular light distribution or *light spots* [33]. These phenomena are caused by turbidity, sunlight channeling, and artificial light sources that produce regions with high intensity pixels that bias the averaging. These *outlier* regions affect model performance resulting in images with incorrect ambient light and irregular color enhancement. Thus, the median per channel was adopted to calculate Context Luminosity. This option is due to the median value presenting more robustness to the outlier values than the average value. The Context Luminosity calculation is

$$\Lambda_{(k)} = median(I_{(k)}(x, y)). \quad (8)$$

3.6. Images for the degradation function

The degradation algorithm is summarized by (3), and the distortion imposed on the output image of the autoencoder $\hat{I}$ is obtained from the difference between $I^*_{(k)}$ and $I_{c(k)}$. These images comprise different features of distortion, and should produce a reduction in color and contrast along with increased turbidity and colorcast.

The $I^*_{(k)}$ image contains the turbidity and colorcast features added to the $\hat{I}_{(k)}$, and it is estimated in two steps using (4), as follows,

$$I^*_{(k)}(x, y) = I_{dbn(k)}(x, y).e^{-gd^*_{(k)}(x,y)} + (1 - e^{-gb^*_{(k)}(x,y)})\Lambda^*_{(k)}, \quad (9)$$

where

$$I_{dbn(k)}(x, y) = I_{(k)}(x, y).e^{-gd_{(k)}(x,y)} + (1 - e^{-gb_{(k)}(x,y)})\Lambda_{(k)}. \quad (10)$$

The image $I_{dbn(k)}$ is a degraded version of the input image $I_{(k)}$ and the parameters $gd_{(k)}$, $gb_{(k)}$ and $\Lambda_{(k)}$ are obtained using (6), (7) and (8), respectivaly. This image defines a new and increased context of degradation, allowing the calculation of the parameters $gd^*_{(k)}$, $gb^*_{(k)}$, and $\Lambda^*_{(k)}$. These parameters are used to calculate $I^*_{(k)}$ in (9), and they are also calculated by (6), (7), and (8), respectively, but using values from $I_{dbn(k)}$ context.

The image $I_{c(k)}$ is a histogram-stretched version of $I_{(k)}$, and it is defined in (11). Essentially, this histogram stretching produces increased contrast and pixel intensity by *dynamic range* expansion [34]. This operation is used in contrast adjustment in non-underwater images, and produces improvement in visual perception of the image. However, UW images tend to present reduced *dynamic range*, and the *stretching* can produce higher gaps between adjacent frequency sets in the histogram. This condition drives higher or excessive texturing effects, and loss of quality in visual perception. Moreover, images with outliers regions and strong unbalance of the color channels tend to present saturated regions when the histogram stretching operation is performed in order to calculate the $I_{c(k)}$. In our algorithm, this image is used as a distortion element for contrast and color, since it is subtracted from $\hat{I}_{(k)}$. Fig. 4 shows the block diagram of the degradation function and the autoencoder.

$$I_{c(k)}(x, y) = \frac{I_{(k)}(x, y) - min(I_{(k)}(x, y))}{b_{w(k)}}, \quad (11)$$

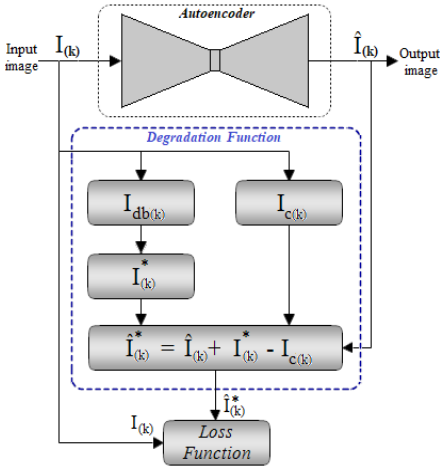

Fig. 4. Diagram of the method indicating the degradation function and the autoencoder.

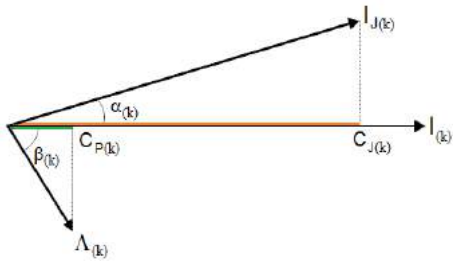

Fig. 5. Image representation described in (4) as a vector sum. The image I_k is obtained from the composition between vectors I_{Jk} and Λ_k. The C_{Jk} and C_{Pk} are the respective components in the I_k direction.

3.7. Loss function

The loss function adopted is composed of two terms. A term related to the scene radiance context, and another related to the entire image radiance. The scene component is dedicated to highlighting the objects in the scene, and the image component provides global information during the image reconstruction process by the autoencoder.

The scene radiance component is defined from the interpretation of (4) as a vector sum. Thus, the terms $I_{J(k)}(x,y)e^{-gd_{(k)}(x,y)}$ and $(1 - e^{-gb_{(k)}(x,y)})\Lambda_{(k)}$ compose a complementary action weighted by the exponential factor. We define the image as a vector resulting from these two terms. Assuming that $I_{(k)}$ is the outcome of the vector composition of the $I_{J(k)}$ and $\Lambda_{(k)}$, as shown in Fig. 5 and described as

$$I_{(k)} = I_{J(k)}cos(\alpha_{(k)}(x,y)) + \Lambda_{(k)}cos(\beta_{(k)}(x,y)), \tag{12}$$

where:

$$cos(\alpha_{(k)}(x,y)) = e^{-gd_{(k)}(x,y)}, \tag{13}$$

and

$$cos(\beta_{(k)}(x,y)) = (1 - e^{-gb_{(k)}(x,y)}). \tag{14}$$

In addition, all the quantities are defined in the RGB channels.

The components C_{Jk} and C_{Pk} of the I_k can be expressed by

$$C_{J(k)}(x,y) = I_{J(k)}(x,y).cos(\alpha_{(k)}(x,y)), \tag{15}$$

$$C_{P(k)}(x,y) = \Lambda_{(k)}.cos(\beta_{(k)}(x,y)), \tag{16}$$

and

$$I_{(k)}(x,y) = C_{J(k)}(x,y) + C_{P(k)}(x,y). \tag{17}$$

Expanding to the degraded image $\hat{I}^*$, it follows

$$\hat{I}^*_{(k)}(x,y) = \hat{C}^*_{J(k)}(x,y) + \hat{C}^*_{P(k)}(x,y), \tag{18}$$

the $C_{P(k)}$ and $\hat{C}^*_{P(k)}$ components are related to the environment light that reaches the camera. The $C_{J(k)}$ and $\hat{C}^*_{J(k)}$ components prioritize the information about objects in the scene. Essentially, the images $I_{(k)}$ and $\hat{I}^*_{(k)}$ are matched inside the loss function. Instead of using only the $I_{(k)}$ and $\hat{I}^*_{(k)}$ in the loss function, we also use a term composed by $C_{J(k)}$ and $\hat{C}^*_{J(k)}$ components as described by (19) and (20), respectively.

$$C_{J(k)}(x,y) = I_{(k)}(x,y) - C_{P(k)}(x,y), \tag{19}$$

and

$$\hat{C}^*_{J(k)}(x,y) = \hat{I}^*_{(k)}(x,y) - \hat{C}^*_{P(k)}(x,y). \tag{20}$$

The $C_{P(k)}$ and $\hat{C}^*_{P(k)}$ components are calculated using (16) with the respective parameters. The term of the loss function is obtained from the Mean Square Error (MSE) between the components, which is calculated as

$$\mathcal{L}_{sc} = MSE(C_{J(k)}(x,y), \hat{C}^*_{J(k)}(x,y)). \tag{21}$$

The term of the image radiance is composed by the MSE between images $I_{(k)}$ and $\hat{I_{(k)}}^*$ as indicated in (22). This term provides image context information avoiding the formation of artifacts in the output image due to loss unbalance, which appears when employing only scene components.

$$\mathcal{L}_{im} = MSE(I_{(k)}(x,y), \hat{I}^*_{(k)}(x,y)). \tag{22}$$

The scene and image radiance terms described in (21) and (22) are used to compose the loss function used in training the model, as indicated in (23). The constants c_1 and c_2 are used to weight the terms. They are indicated in the Experimental Results Section.

$$\mathcal{L} = c_1.\mathcal{L}_{sc} + c_2.\mathcal{L}_{im}, \tag{23}$$

where the c factors are weights for loss terms.

3.8. Attention for saturated regions in the image

The histogram stretching performed on images presenting strong unbalance of the color channel and outlier regions, simultaneously, tends to generate high-intensity or saturated areas in the images. The unbalance of the color channels can be caused by the nature of the scene itself or by the turbidity and color cast effects. The outlier regions are generated by the presence of pixels with an intensity that is much higher than the average of the image. They are due the irregular ambient light distribution. Another characteristic of these images is the much lower minimum values. Under these conditions, the histogram stretching operation generates high values for the high-intensity pixels in the normalization procedure. Thus, saturated areas are created in the histogram-stretched image. The content of this image is removed from the output image in the degradation function, including the saturated area. During the training phase, the autoencoder learns to compensate for the degradation. In this case, it generates compensation for the removed saturated areas, and these areas are formed in the output and enhanced images. Fig. 6 shows an example of presenting saturated and high-intensity areas. The red rectangle highlights the saturated area generated by the method. In order to overcome this drawback, we propose an attention mechanism that detects these outlier pixels and color channel unbalance. The algorithm was implemented in the degradation function, and compensates for these effects. It is discussed in the following section.

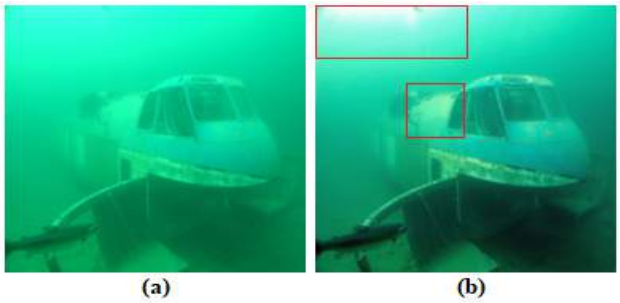

Fig. 6. Example of image with partial and strong incidence of the light source causing unbalance in color channels and outlier regions. The generated high-intensity areas in the output image are indicated inside the red rectangles. (a) Input image. (b) Output image.

In order to reduce these saturated regions, we propose an attention mechanism dedicated to treating these regions. The attention module is inserted in the degradation function as an additional term of degradation. The diagram in Fig. 7 shows the inserted attention module. This module maps the input image, and generates the attention image $I_{at(k)}$, containing regions of high intensity generated by the presence of outliers pixels and color channels unbalance. This image $I_{at(k)}$ is added to the output image in the degradation function, acting as additional degradation. The novel degradation function is

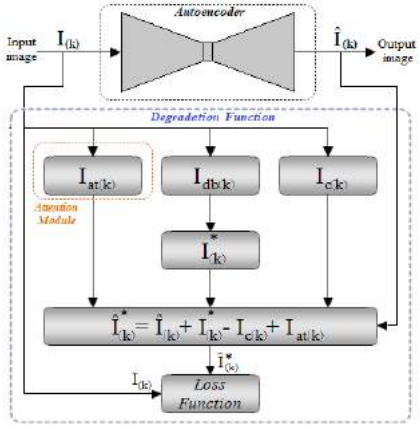

Fig. 7. Diagram of the method with the attention module inserted in the degradation function.

$$\hat{I}^{*}_{(k)}(x,y) = \hat{I}_{(k)}(x,y) + I^{*}_{(k)}(x,y) - I_{c(k)}(x,y) + I_{at(k)}(x,y), \quad (24)$$

The attention image $I_{at(k)}$ is a version with stretched histogram of the image $I_{a(k)}$, described in (26). The histogram stretching is performed using (11) with the parameters of (26). Thus, $I_{at(k)}$ is

$$I_{at(k)} = \frac{I_{a(k)} - min(I_{a(k)})}{max(I_{a(k)}) - min(I_{a(k)})}. \quad (25)$$

The image $I_{a(k)}$ is defined as

$$I_{a(k)} = g_{r(k)} \mid av(I_k) - av_g(I_{(k)}) \mid \frac{I_{x(k)}}{I_{(k)}}, \quad (26)$$

where $g_{r(k)}$ is a reinforcement factor calculated in (27), $av(I_{(k)})$ and $av_g(I_{(k)})$ are the average value per color channel and global average of the image, respectively. The detection of the unbalance of the color channels is performed by the term $\mid av(I_{(k)}) - av_g(I_{(k)}) \mid$ in (25). $I_{(k)}$ is the input image, and $I_{x(k)}$ is the image containing outlier pixels, which is calculated as indicated in (28). The pixels above of threshold represented by $max(I_{(k)})\frac{thr_{(k)}}{av(I_{(k)})}$ are considered outlier pixels. Thus, they are selected to compose $I_{x(k)}$.

$$g_{r(k)} = 1 + \frac{thr_{(k)}}{av(I_{(k)})}, \quad (27)$$

$$I_{x(k)} = \begin{cases} I_{(k)} - max(I_{(k)})\frac{thr_{(k)}}{av(I_{(k)})}, & I_{x(k)} \geqslant 0 \\ 0, & I_{x(k)} < 0, \end{cases} \quad (28)$$

where $max(I_k)$ correspond to the maximum value per color channel of $I_{(k)}$. The term $thr_{(k)}$ in (27) and (28) is defined in (29), and correspond to the threshold value for outlier detection. The presence of outliers in the image increases the distance between the average and the median values, because the average is strongly affected by data outliers. We explore this effect to detect the presence of outliers in the image $I_{(k)}$. The index (k) represents the color channel and $k \in \{R, G, B\}$.

$$thr_{(k)} = \begin{cases} av(I_{(k)}) - mdn(I_{(k)}), & thr_{(k)} \geqslant 0 \\ 0, & thr_{(k)} < 0, \end{cases} \quad (29)$$

where $mdn_{(k)}$ is the median value per color channel of the $I_{(k)}$.

4. Experimental results

The details of the implementation results and performance analysis of the method are described in this section. Also, we present a comparative study involving the proposed method, and other enhancement methodologies.

4.1. Autoencoder

In this work, we consider the enhancement task performed by the neural network as an image reconstruction task, driven by the degradation in a self-supervised approach. The neural network must be able to learn hidden structures in unlabeled data, representing them in latent features. A deep learning model based on autoencoder performs this trying to reconstruct its input [35]. The choice for an architecture based on autoencoder stems from this conception and the nature of the method. The objective was to implement a neural network with low computational cost and an easy approach for self-learning.

The implemented autoencoder is shown in Fig. 8. The structure of the neural network is a Fully Convolutional Network (FCN) with 170k trainable parameters. The downscaling steps are performed via *strided* convolutions, instead of pooling layers, in order to minimize information loss. The upscale in the decoder section is performed by Keras 2D upsampling layers, with size $(2, 2)$, interpolation by nearest neighbors and other arguments with default values. The activation functions are

ReLU, except in the final layers of the decoder that use Leaky-ReLU. The convolutional layers use L1 regularization with factors of the kernel, and bias set to 15×10^{-6} and 1.5×10^{-6}, respectively. Only the layers of the autoencoder have trainable parameters, and the degradation function was implemented as a non-trainable layer.

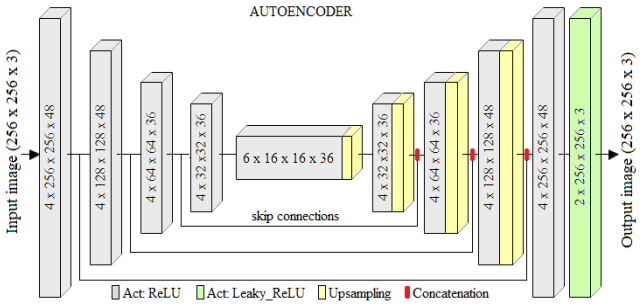

Fig. 8. Architecture of the implemented autoencoder.

4.2. Training details and dataset

The autoencoder was implemented in Keras/Tensorflow®, and it was initialized with Glorot Normal Algorithm. The optimizer is Adam with a learning rate equal to 0.0008, and other parameters were maintained in its default values. In the Leaky-ReLU layers, the alpha parameter is equal to 0.19. The training was performed in 200 epochs with a batch size equal to 6. The computer configuration is i5-6400 CPU, 32 MB RAM with a Titan X GPU. The image format is RGB $256 \times 256 \times 3$ scaled to the range $[0, 1]$.

As a future prospect for extending the application of the method is in embedded vision systems of Remotely Operated Vehicles (ROV). Processing time requirements for real time operations are an important issue in these systems. Thus, we evaluated the inference time of the trained model for the processing one image of $256 \times 256 \times 3$ in 0.25s, in a computer configuration i5-6400 CPU, 32 MB RAM, O.S. Ubuntu 18.04.5, Tensorflow vs.2.20, Keras vs.2.3.1. This results in a rate of 4 FPS, approximately. This value is quite low for real time applications. We expect that with optimization procedures favoring the operating system and model performances, this inference time can be reduced. However, more investigations are needed on this topic.

The dataset used to train the autoencoder contains 2200 real UW images. It was built prioritizing images with medium to high levels of turbidity, reduced color and contrast perception. Also, these present degradation related to shades, and intensities of color cast and ambient light. Images with very low degradation were also used, but not prioritized. We used 800 images from UIEBD database [23], 400 images from the SUIM database [19], 100 images from the RUIE dataset [36], and 900 UW images collected from the internet. These mentioned datasets were used in recent studies, related to the UW image enhancement [12, 14, 15, 19, 21]. The data were randomly split into 90% for training and validation, and 10% for test. Cross-fold validation was performed during the evaluation procedures to check for consistency in image reconstruction (autoencoder) with degradation function added to the pipeline. We used an additional dataset containing 90 images for evaluation in the comparative study described below in the respective section. These images were selected from the datasets cited above, and they were not used in training of the neural network. The c factors values in the loss function expression were empirically defined as $c_1 = 0.65$, and $c_2 = 0.35$.

4.3. Resulting images from the proposed method

The resulting images from the proposed method are obtained at the end of the training phase. Samples of the input, output and degraded images are indicated in Fig. 9. The objects in the scene show a natural aspect when compared to the input image. Visual perception points out contrast improvement and color recovery. Also, the images show color preservation furthering the color constancy properties ([32]). In the context of this work, this color preservation is an important issue, because no synthetic reference images are used. The proposal does not focus on changing the color perception of the image objects. In addition, the method presents good performance in turbidity and colorcast reduction. The reduction in turbidity perception is more evident in foreground regions as a result of the scene-related component in the loss function.

Fig. 9 shows scenes in distinct ambient light, color cast, and turbidity. Underwater images bring a myriad of complexities that pose challenging issues, and define the performance of a method under different aspects of evaluation. The depth affects color and ambient light in UW images. The type of water is related to the nature and quantity of particles in suspension in the water and impacts on contrast, color cast, and blurring. Essentially, the method considers these features as a result of the influence of depth and water type on image acquisition. Thus, the available information about ambient light and turbidity is estimated from the input image. These physical quantities are expressed and explored inside the degradation function as color information, driving the enhancement task. The resulting images are shown in Fig. 9 indicate that the methodology is able to enhance images in different UW environments. The model presents a good generalization ability, resulting from the dataset composition, and the nature of the methodology.

4.4. Comparative study

We performed a comparative evaluation of the method related to the consolidated methodologies oriented to UW enhancement and deep learning based methods, developed in [15, 21, 19, 30]. Also, we compared the proposed method to the works presented in [4] and [5], both based on digital image processing. In the comparative analysis, the methods are identified by "Li2021" [15], "Li2020" [21], "FUnIEGAN" [19], "Fusion" [4], "UDCP" [5], "Mello2021" [30], and the present proposal by "Ours".

4.4.1. Qualitative analysis

For evaluation concerns, we consider as quality criteria the ability of the method in recovery the color and contrast, reduction of the color cast and dehaze action. Also, the color constancy of the resulting images is considered through the color

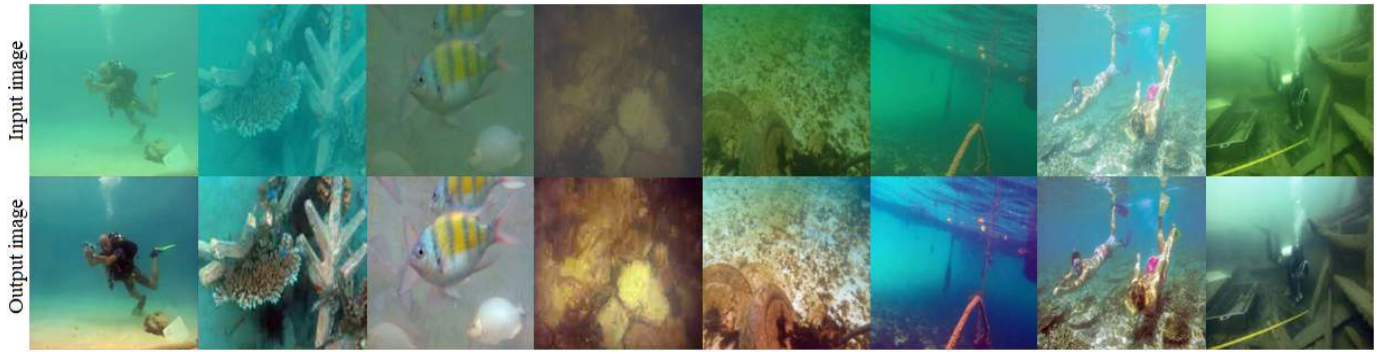

Fig. 9. Examples of enhanced images obtained from our method.

preservation and coherency between the input image (UW image) and the enhanced image. Fig. 10 shows samples of the resulting images obtained from the methods used in the comparative study. These images compose the dataset with 90 samples mentioned in Section 4.2, and are used in the comparative study only.

Most of the methods present an irregular enhancement action with strong color and color cast increasing. The UDCP showed effectiveness in the haze reduction, but with strong color distortion, and darkening in some images. The Fusion method presented an effective dehaze action and contrast recovery, but with strong texturing of the image, and some images showing excessive color enhancement. FUnIEGAN presented low performance dealing with turbidity, and color distortion seemingly associated with the excessive enhancement of the red channel. Li2020 showed intense enhancement of the red channel, and difficulty in leading with haze. Mello2021 presented irregular enhancement action in images with unbalanced ambient light, generating dark or very clear regions in the enhanced images. The method Li2021 and Ours had the best performances in the image improvement. Color preservation and recovery have similar intensities and dehaze action. However, the method Li2021 produced a more natural colorfulness in some images, while Ours showed the best improvement in contrast. Both methods presented good generalization ability.

In our method, the dehazing action is conditioned by the scene components of the loss function. These components drove the enhancement of the scene, and the haze reduction was concentrated in the foreground regions of the image. This scene enhancement provides an important property for UW application-oriented tasks, such as UW maintenance, inspection, and fauna and flora explorations. Turbidity effect reduction over wide, and long-range scenarios can be important to monitoring and inspection activities. In our context, large or restricted scenes are tackled similarly. In acquired images, with scenes showing objects far from the camera, the turbidity perception is more accentuated. The dehazing task performed by the method acts uniformly on the image, but in these regions it is less intense.

4.4.2. Quantitative analysis

The quantitative analysis of underwater images is a poorly defined issue, due to the lack of consensus on the effectiveness of existing metrics in representing the quality of these images [6]. Usually, the authors use an integrated set of the metrics containing the image quality metrics that require or not a reference image (ground-truth). The metrics that use reference or ground-truth images were developed to indoor/outdoor applications [20], but applied to UW image analysis, and they need synthesized paired datasets, built from the UW images. On the other hand, there are no-reference metrics proposed specifically for UW image quality analysis. We proceeded with the quantitative evaluation of the images, using the Underwater Image Color Evaluation (UCIQE) [37], Underwater Image Quality Metric (UIQM) [38], and CCF metric [39]. Also, we use the no-reference metric Natural image quality evaluator (NIQE) [40].

The UCIQE measures color degradation in the CIELab color space, based on standard deviation in chroma, average saturation, and difference between extreme values in luminance. The UIQM metric evaluates color degradation with measurements of colorfulness using the opponent color theory, and blurring degradation with measurements of edge sharpness and contrast. UCIQE, UIQM are a linear combination of weights obtained from a subjective test, and related to specific attributes, such as contrast and saturation, besides, luminance and chroma. Higher scores resulting from UCIQE and UIQM indicate higher image quality. Furthermore, CCF measures color degradation with a colorfulness index, blurring, and lack of visibility with a contrast measure and a foggy index. The weighting of these indexes is based on subjective tests, where images of a color chart are captured at different distances and turbidity. Higher values of CCF score indicate better image quality. Moreover, NIQE considers human vision sensitivity to high-contrast areas in images. It uses multivariate gaussian (MVG) to define the feature model of sensitive areas where larger values of the model parameter higher the quality of the image. A smaller score on NIQE indicates better perceptual quality.

Table 1 shows the UIQM, UCIQE, CCF, and NIQE scores for the compared methods. The best UCIQE score is UDCP, and the method Li2021 had the best performance in UIQM score. Besides, Fusion presented the best CCF score, while Ours had the best NIQE score.

The values presented in the Table 1 point out some issues related to the characteristics of the metrics. The UW image metrics focus on feature intensities like chroma and saturation (UCIQE), and colorfulness and contrast (UIQM), colorfulness,

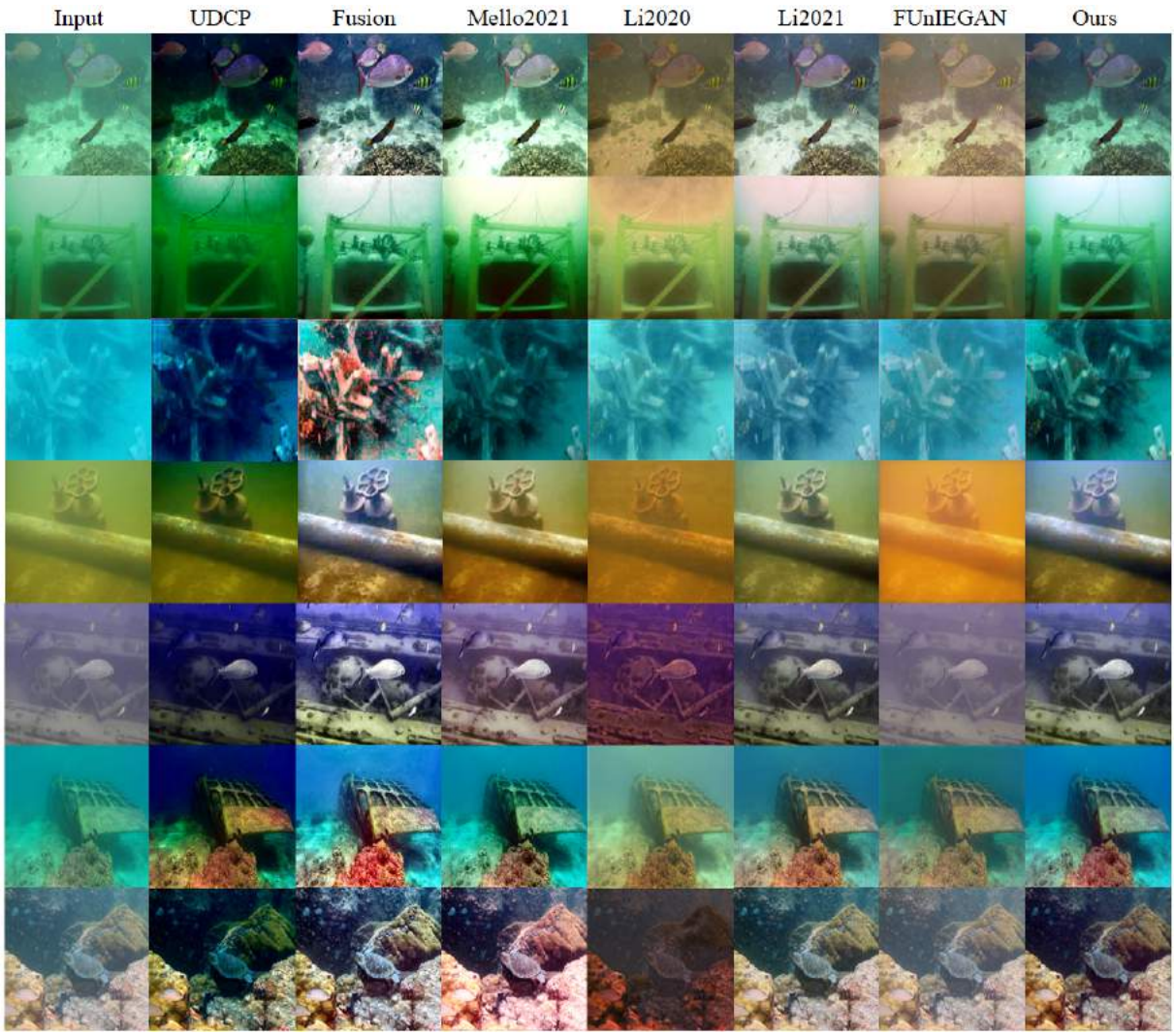

Fig. 10. Output images for the comparative study. Methods: UDCP [5], Fusion [4], Mello2021 [30], Li2020 [21], Li2021 [15], FUnIEGAN [19], Ours.

Table 1. Quantitative results for UIQM, UCIQE, CCF, and NIQE metrics. UDCP (Drews et al.,2016) [5], Fusion (Ancuti et al., 2012) [4], Mello2021 (Mello et al., 2021) [30], Li2020 (Li et al.,2020) [21], Li2021 (Li et al, 2021) [15], and FUnIEGAN (Islam, et al., 2019) [19].

Metric	UIQM ↑	UCIQE ↑	CCF ↑	NIQE ↓
UDCP	1.604	**6.915**	34.229	4.559
Fusion	2.623	5.568	**36.895**	5.875
Mello2021	2.189	4.934	27.803	4.786
Li2020	2.654	3.281	13.737	4.652
Li2021	**2.913**	3.593	19.824	4.830
FUnIEGAN	2.628	3.414	14.282	5.213
Ours	2.649	4.623	23.492	**4.415**

contrast and turbidity (CCF). They tend to produce high scores for images with high contrast and extreme chroma [20] ignoring human perception context [9]. The results shown in Table 1 do not present consistency with subjective evaluation of the images. The UCIQE and CCF scores suggest similar performance in the image assessment. The scores obtained by UDCP are much higher than other ones. They are resulting from excessive color improvement, and darkening perceptible in visual analysis. However, these distortions caused a loss of contrast and color diversity, resulting in the lowest UIQM score. The methods that did not score high on the UCIQE present high values for UIQM.

The resulting scores for most methods show uneven performance on the set of metrics. This condition makes a concise evaluation difficult. However, this can be mitigated when we consider the balance of the method scores. In this context, the methods Fusion, Li2021, and Ours presented the best performances. This approach is consistent with the visual perception of the images. These non-reference metrics have a sensibility to different features, and refer to distinct concepts of UW image quality. Underwater image enhancement quality metrics are an open issue, and they lack higher investigations.

4.5. Quantitative analysis using full-reference metrics

In order to provide a more concise quantitative evaluation of the method, we compared the methods using full-reference metrics often utilized to evaluate image quality. Specifically, Peak Signal-to-Noise Ratio (PSNR), Mean Square Error (MSE)

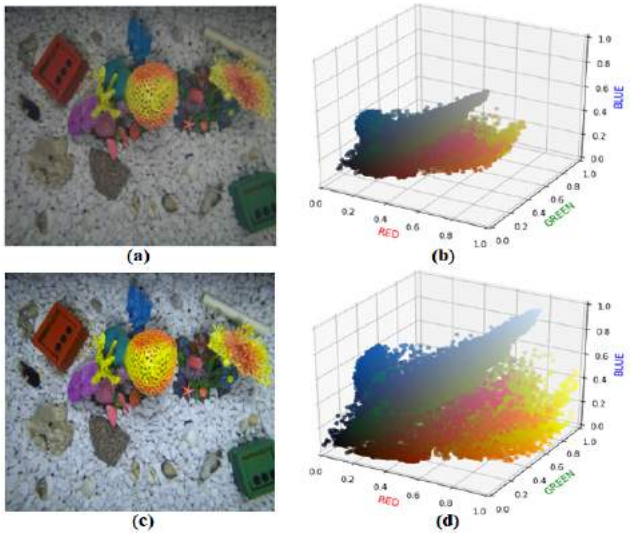

Fig. 11. The image "ref" of the Milk dataset and the respective stretched-histogram, and white-balanced version used as reference image in the quantitative evaluation with referenced metrics. (a) Original image from the Milk dataset (#ref), (b) Pixel distribution in the RGB color space of the original image, (c) Expanded dynamic and color balanced version of the image "ref", and (d) Pixel distribution in the RGB color space of the image with expanded dynamic and color balanced version.

[41], Structural Similarity (SSIM) [42], Gradient Magnitude Similarity Deviation (GMSD) [43], CIEDE2000 [44], Feature Similarity, and Feature Similarity with Chrominance (FSIM and FSIMc) [45]. The evaluation contexts of these metrics are pixel-wise (PSNR, MSE), structure-related (SSIM, GMSD), color (CIEDE2000), and features on visual perception (FSIM, FSIMc). However, all these metrics require a reference image. In this approach, we do not use reference images. Methodologies based on deep learning that use these images, synthesized them using a specific method, and proceed to the training of the their models. We consider that this approach can bias quantitative analysis. Since, a model trained using synthetic and improved images, restored with a specific method, can present good scores in referenced metrics, when using reference images obtained from the same specific method. Thus, we opted to use images from the Milk dataset provided by the Turbid dataset, instead of synthesizing reference images. The Milk dataset has 20 images presenting progressive turbidity, with the first image ("ref") acquired in water without turbidity. We used as reference image a stretched-histogram and white-balanced version of this image, shown in Fig. 11. This version allows improved color and contrast perception, and minimizes the darkened context of the original image, caused by the controlled illumination, used in the image acquisition procedures [31]. Fig. 11 (a) and (c) show the first image from the Milk dataset and the stretched-histogram, and white-balanced version, respectively. The distributions of the pixels in the RGB color space for both images are shown in Fig. 11 (b) and (d), respectively.

Fig. 12 shows samples of the resulting images from the methods for the images of the Turbid/Milk dataset. These images can be used to evaluate methods for progressive turbidity. From the visual perspective, the Fusion and Ours methods performed better under severe turbidity conditions, maintaining the color and scene structure perception. Table 2 shows the resulting scores for the image quality metrics that use the reference image.

The scores of the metrics related to error and noise (MSE and PSNR) show good performance of Ours, Fusion, and Mello2021. This is consistent with the subjective analysis of the images. These methods showed quite similar results, mainly in images with a low level of turbidity. Our proposal presented the best scores for the scene structure-related metrics (PSNR, MSE, SSIM), and for the feature perception on gray scale (FSIM). These results are important to the visual perception of UW images. The human vision system is less sensible to color changes than luminance changes [34]. The depth and water turbidity effects on the colors are increased under reduced ambient light. Under these conditions, scene perception is driven by luminance.

Furthermore, our method also presented the best score in the GMSD metric. Since that the gradient images are sensitive to image distortions, whereas different local structures in a distorted image suffer different degrees of degradation. GMSD captures image local quality, and the standard deviation of the image's gradient map is computed as the final image quality index.

The method Fusion presented better performance in color-related metrics (CIEDE2000, FSIMc); but, our method also had high scores in these metrics. Color recovery is an important issue in the enhancement task. Unlike the other deep learning-based methods, our proposal does not use paired datasets; hence, additional information provided by the synthetic ground-truth images is not available. Therefore, our method enhanced the color information present only in the input image. However, we did not consider this condition as a limitation, but a characteristic of the method, since the real intensity and hue of the colors in UW images are unknown.

From the above discussion, our method and Fusion showed the best scores for the referenced metrics. The results of this experiment agreed with the qualitative analysis, based on visual perception of the images, and they are consistent with the qualitative analysis for UW images, discussed in the previous subsection.

4.6. Performance of the method with attention module

In this section we evaluate the method with the proposed attention module inserted in the pipeline of the model. Fig. 13(a) shows samples of images presenting outlier regions (pixels). Fig. 13(b) shows output images of the method without attention module (baseline) as discussed in the sections above. The outlier regions can be observed in areas presenting high-intensity or color distortion in the images. As an example of this color distortion, the image showing the lion statue highlights an area in red. This area corresponds to an outlier region, generated in the histogram stretching procedure due to the unbalance of the color channels. The red channel in this image presents a narrow dynamic range, which produces a high normalization factor during the histogram stretching. In Fig. 13(c) are show the output images with attention module present in the pipeline of the model. The high-intensity areas are limited due to the attention module action. It is highlighted that there is a redistribution of the light over a wider area, and a smoothing in visual

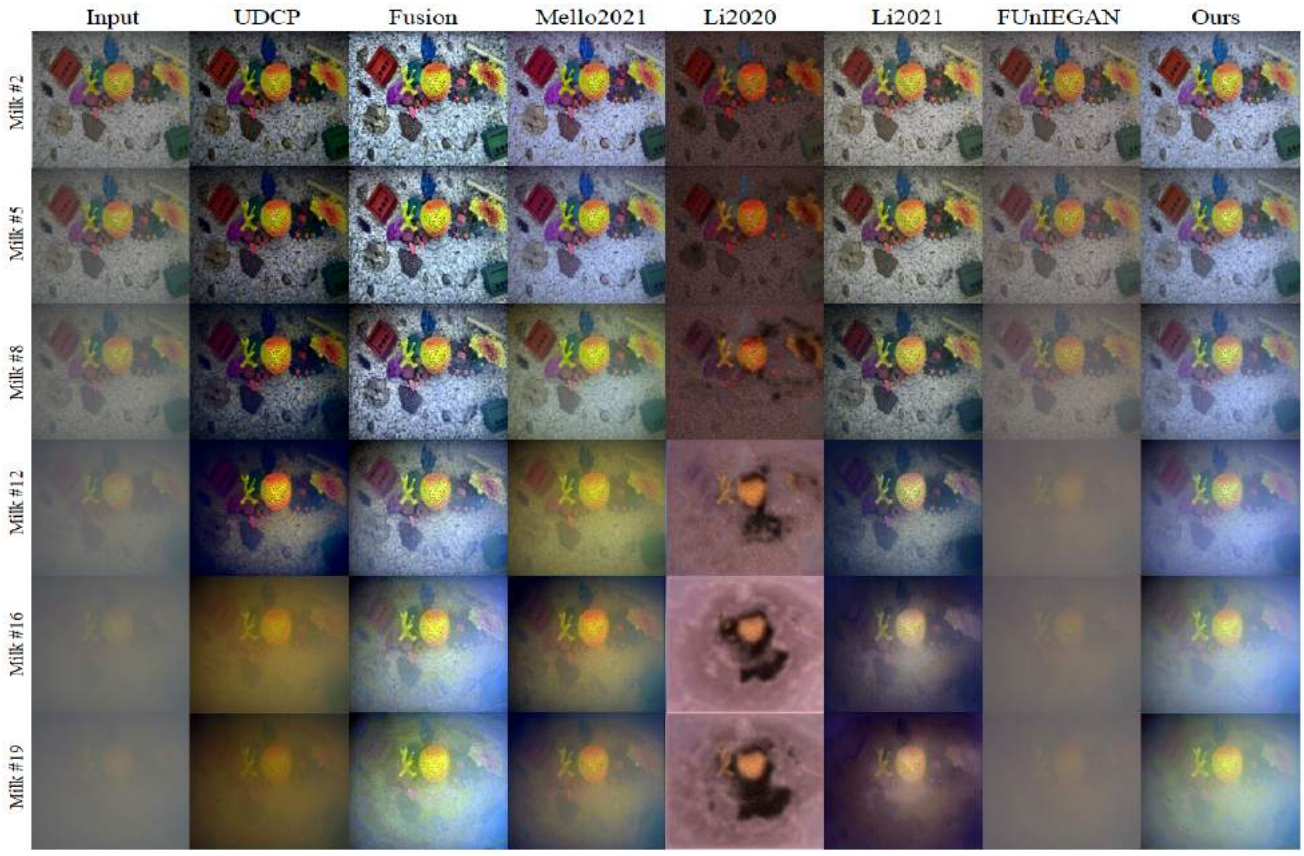

Fig. 12. Output images for the comparative study using images from the Turbid/Milk dataset. Methods UDCP [5], Fusion [4], Mello2021 [30], Li2020 [21], Li2021 [15], FUnIEGAN [19], Ours.

perception of these regions. In Fig. 13(d) are showed the images from the algorithm of the attention module. These images are obtained from (25). In the pipeline of the model, the degradation function uses these images to drive the autoencoder to reduce the high-intensity areas.

We evaluated the model with the attention module added to the pipeline, and compared it with the baseline model (without the attention module). Table 3 indicates the results for no-reference metrics. The performance of the model with attention module is similar to the baseline model. It shows slightly lower results for the UIQM, CCF, and NIQE metrics, and a higher score for the UCIQE metric. Similarly to the previous section, the performance related to the full-reference metrics was performed using the images of Turbid/Milk dataset, and the resulting scores are indicated in Table 4. The model with attention model shows performance is very close to the baseline model, and is in agreement with the results for no-reference metrics. In both cases, the differences in performance reflect the action of the attention module.

5. Conclusion

In this work, we presented an image enhancement proposal based on a deep learning approach. Our method uses real UW images only. The algorithm employs an image description, which is inspired by the IFM, but is color space contextualized. Our algorithm degrades properly the output image of the autoencoder, and replaces it in the loss function by the distorted image. This procedure *misleads* the training of the neural network, allowing it to learn how to correct the imposed distortion. The method requires no prior or ground-truth images. Our main assumption states that the natural degradation of an UW image can be synthetically increased more easily than removed from the image. The results showed the effective action of the method, related to color preservation and contrast improvement. Furthermore, we propose an attention mechanism to tackle saturated regions generated in images, presenting unbalance in color channels and strong outlier pixels. A comparative study was performed involving no and full-reference image quality metrics. In both cases, the proposed method provided good performance. However, the results from the no-reference metrics allowed a better assessment of the compared methods, when analyzed jointly. In future work, we intend to extend our methodology to UW applications, as well as different domains that require imagery denoising.

Acknowledgments

The authors would like to thank to National Council for Scientific and Technological Development (CNPq) and the Coordination for the Improvement of Higher Education Personnel (CAPES). In addition, the authors are grateful to Dr. Chongyi Li who kindly provided access to the images from the UIEBD database.

Table 2. Quantitative results for PSNR, MSE, SSIM, CIEDE2000 (CIE2K), FSIM, FSIMc, and GSMD metrics. The adopted reference image is the expanded *dynamic range* and white balanced version of the image 'ref' from Milk dataset, indicated in the Fig. 11(c). UDCP (Drews et al.,2016) [5], Fusion (Ancuti et al., 2012) [4], Mello2021 (Mello et al., 2021) [30], Li2020 (Li et al.,2020) [21], Li2021 (Li et al, 2021) [15], and FUnIEGAN (Islam, et al., 2019) [19]

Metric	PSNR ↑	MSE ↓	SSIM ↑	CIE2K ↓	FSIM ↑	FSIMc ↑	GMSD ↓
UDCP	10.17	0.099	0.492	29.15	0.972	0.749	$34e^{-6}$
Fusion	17.06	0.024	0.704	**14.20**	0.976	**0.812**	$26e^{-6}$
Mello2021	16.46	0.045	0.670	20.44	0.977	0.743	$24e^{-6}$
Li2020	11.23	0.083	0.481	29.72	0.926	0.645	$48e^{-6}$
Li2021	13.95	0.054	0.650	21.63	0.975	0.788	$26e^{-6}$
FUnIEGAN	14.52	0.036	0.551	18.52	0.973	0.620	$37e^{-6}$
Ours	**17.65**	**0.023**	**0.706**	15.32	**0.979**	0.763	$\mathbf{23}e^{-6}$

Fig. 13. Samples of the images presenting unbalance in color channels and outlier pixels treated by the attention module. (a) Input image; (b) Output image; (c) Output image with the attention module inserted in degradation function; (d) Image from the attention module algorithm.

Table 3. Quantitative results for the model (Mod) and the model with the attention module (Mod+Att). Results for no-reference metrics UIQM, UCIQE, CCF, and NIQE.

	Mod	Mod+Att
UIQM ↑	2.649	2.483
UCIQE ↑	4.623	5.050
CCF ↑	23.49	22.02
NIQE ↓	4.415	4.750

Table 4. Quantitative results for the model (Mod) and the model with the attention module (Mod+Att) for full-reference metrics: PSNR, MSE, SSIM, CIEDE2000 (CIE2K), FSIM, FSIMc, and GSMD. The adopted reference image is the expanded *dynamic range*, and white balanced version of the image 'ref' from Milk dataset, indicated in the Fig. 11(c).

	Mod	Mod+Att
PSNR ↑	17.65	17.24
MSE ↓	0.023	0.027
SSIM ↑	0.706	0.695
CIE2K ↓	15.32	16.83
FSIM ↑	0.979	0.978
FSIMc ↑	0.763	0.762
GSMD ↓	0.023	0.023

References

[1] Donaldson, JA, Drews-Jr, P, Bradley, M, Morgan, DL, Baker, R, Ebner, BC. Countering low visibility in video survey of an estuarine fish assemblage. Pacific Conservation Biology 2020;26(2):190–200.

[2] Drews-Jr, P, Hernández, E, Elfes, A, Nascimento, ER, Campos, M. Real-time monocular obstacle avoidance using underwater dark channel prior. In: IEEE/RSJ IROS. 2016, p. 4672–4677.

[3] Dos Santos, M, De Giacomo, GG, Drews-Jr, PLJ, Botelho, SSC. Matching color aerial images and underwater sonar images using deep learning for underwater localization. IEEE Robotics and Automation Letters 2020;5(4):6365–6370.

[4] Ancuti, C, Codruta, A, Haber, T, Bekaert, P. Enhancing underwater images and videos by fusion. IEEE/CVF CVPR 2012;:81–88.

[5] Drews-Jr, P, Nascimento, E, Botelho, S, Campos, M. Underwater depth estimation and image restoration based on single images. IEEE Computer Graphics and Applications 2016;36:24–35.

[6] Han, M, Lyu, Z, Qiu, T, Xu, M. A review on intelligence dehazing and color restoration for underwater images. IEEE Transactions on Systems, Man, and Cybernetics: Systems 2018;PP:1–13.

[7] Pan, Pw, Yuan, F, Cheng, E. De-scattering and edge-enhancement algorithms for underwater image restoration. Frontiers of Information Technology & Electronic Engineering 2019;20:862–871.

[8] Akkaynak, D, Treibitz, T. A revised underwater image formation model. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018, p. 6723–6732.

[9] Berman, D, Levy, D, Avidan, S, Treibitz, T. Underwater single image color restoration using haze-lines and a new quantitative dataset. IEEE Transactions on Pattern Analysis and Machine Intelligence 2021;43(8):2822–2837.

[10] Codruta, A, Ancuti, C, Vleeschouwer, C, Bekaert, P. Color balance and fusion for underwater image enhancement. IEEE TIP 2018;PP:1–1.

[11] Vasamsetti, S, Mittal, N, Neelapu, BC, Sardana, HK. Wavelet based perspective on variational enhancement technique for underwater imagery. Ocean Engineering 2017;141:88–100.

[12] Li, Y, Chen, R. Uda-net: Densely attention network for underwater image enhancement. IET Image Processing 2021;15.

[13] Dudhane, A, Hambarde, P, Patil, P, Murala, S. Deep underwater image restoration and beyond. IEEE Signal Processing Letters 2020;27:675–679.

[14] Lin, Y, Shen, L, Wang, Z, Wang, K, Zhang, X. Attenuation coefficient guided two-stage network for underwater image restoration. IEEE Signal Processing Letters 2020;PP:1–1.

[15] Li, C, Anwar, S, Hou, J, Cong, R, Guo, C, Ren, W. Underwater image enhancement via medium transmission-guided multi-color space embedding. IEEE Transactions on Image Processing 2021;30:4985–5000.

[16] Wang, Y, Guo, J, Gao, H, Yue, H. Uiec2-net: Cnn-based underwater image enhancement using two color space. Signal Processing: Image

Communication 2021;96:116250.
[17] Cho, Y, Jang, H, Malav, R, Pandey, G, Kim, A. Underwater image dehazing via unpaired image-to-image translation. International Journal of Control, Automation and Systems 2020;18:605–614.
[18] Hashisho, Y, Albadawi, M, Krause, T, von Lukas, UF. Underwater color restoration using u-net denoising autoencoder. In: 2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA). 2019, p. 117–122.
[19] Islam, MJ, Xia, Y, Sattar, J. Fast underwater image enhancement for improved visual perception. IEEE Robotics and Automation Letters 2020;5(2):3227–3234.
[20] Wang, Y, Song, W, Fortino, G, Qi, LZ, Zhang, W, Liotta, A. An experimental-based review of image enhancement and image restoration methods for underwater imaging. IEEE Access 2019;7:140233–140251.
[21] Li, C, Anwar, S, Porikli, F. Underwater scene prior inspired deep underwater image and video enhancement. Pattern Recognition 2020;98:107038.
[22] Fabbri, C, Islam, MJ, Sattar, J. Enhancing underwater imagery using generative adversarial networks. In: 2018 IEEE International Conference on Robotics and Automation (ICRA). 2018, p. 7159–7165.
[23] Li, C, Guo, C, Ren, W, Cong, R, Hou, J, Kwong, S, et al. An underwater image enhancement benchmark dataset and beyond. IEEE Transactions on Image Processing 2020;29:4376–4389.
[24] Fayaz, S, Parah, S, Qureshi, G, Kumar, V. Underwater image restoration: A state-of-the-art review. IET Image Processing 2020;15.
[25] Song, W, Wang, Y, Huang, D, Liotta, A, Perra, C. Enhancement of underwater images with statistical model of background light and optimization of transmission map. IEEE Transactions on Broadcasting 2020;66(1):153–169.
[26] Li, X, Hou, G, Li, K, Pan, Z. Enhancing underwater image via adaptive color and contrast enhancement, and denoising. Engineering Applications of Artificial Intelligence 2022;111:104759.
[27] Li, J, Skinner, KA, Eustice, RM, Johnson-Roberson, M. WaterGAN: Unsupervised generative network to enable real-time color correction of monocular underwater images. IEEE Robotics and Automation letters 2017;3(1):387–394.
[28] Ronneberger, O, Fischer, P, Brox, T. U-net: Convolutional networks for biomedical image segmentation. In: Navab, N, Hornegger, J, Wells, WM, Frangi, AF, editors. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Cham: Springer International Publishing. ISBN 978-3-319-24574-4; 2015, p. 234–241.
[29] Zhu, JY, Park, T, Isola, P, Efros, AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: 2017 IEEE International Conference on Computer Vision (ICCV). 2017, p. 2242–2251.
[30] Mello, C, Drews-Jr, P, Botelho, S. Degradation-driven underwater image enhancement. In: LARS 2021. 2021, p. 186–191.
[31] Duarte, A, Codevilla, F, Gaya, J, Botelho, S. A dataset to evaluate underwater image restoration methods. In: OCEANS 2016. 2016, p. 1–6.
[32] Ebner, M. Color Constancy. 1st ed.; Wiley Publishing; 2007. ISBN 0470058293.
[33] Jian, M, Liu, X, Luo, H, Lu, X, Yu, H, Junyu, D. Underwater image processing and analysis: A review. Signal Processing: Image Communication 2021;91.
[34] Burger, W, Burge, MJ. Digital Image Processing: An Algorithmic Introduction Using Java. 2nd ed.; Springer Publishing Company, Incorporated; 2016. ISBN 1447166833.
[35] Pretorius, A, Kroon, S, Kamper, H. Learning dynamics of linear denoising autoencoders. In: Dy, J, Krause, A, editors. Proceedings of the 35th International Conference on Machine Learning; vol. 80 of *Proceedings of Machine Learning Research*. 2018, p. 4141–4150.
[36] Liu, R, Fan, X, Zhu, M, Hou, M, Luo, Z. Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light. IEEE Transactions on Circuits and Systems for Video Technology 2020;PP:1–1.
[37] Yang, M, Sowmya, A. An underwater color image quality evaluation metric. IEEE Transactions on Image Processing 2015;24(12):6062–6071.
[38] Panetta, K, Gao, C. Human-visual-system-inspired underwater image quality measures. IEEE Journal of Oceanic Engineering 2015;41:1–11.
[39] Wang, Y, Li, N, Li, Z, Gu, Z, Zheng, H, Zheng, B, et al. An imaging-inspired no-reference underwater color image quality assessment metric. Computers & Electrical Engineering 2018;70:904–913.
[40] Mittal, A, Soundararajan, R, Bovik, AC. Making a "completely blind" image quality analyzer. IEEE Signal Processing Letters 2013;20(3):209–212.
[41] Steffens, C, Messias, L, Drews-Jr, P, Botelho, S. CNN Based image restoration: Adjusting ill-exposed srgb images in post-processing. Journal of Intelligent & Robotic Systems 2020;99(3-4):609–627.
[42] Wang, Z, Bovik, AC, Sheikh, HR, Simoncelli, EP. Image quality assessment: from error visibility to structural similarity. IEEE TIP 2004;13(4):600–612.
[43] Xue, W, Zhang, L, Mou, X, Bovik, A. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. IEEE TIP 2013;23.
[44] Sharma, G, Wu, W, Dalal, E. The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. Color Research & Application 2005;30:21 – 30.
[45] Zhang, L, Zhang, L, Mou, X. FSIM: A feature similarity index for image quality assessment. IEEE TIP 2011;20:2378 – 2386.