

L²UWE: A Framework for the Efficient Enhancement of Low-Light Underwater Images Using Local Contrast and Multi-Scale Fusion

Tunai Porto Marques, Alexandra Branzan Albu

University of Victoria

Victoria, British Columbia, Canada

tunaip@uvic.ca, aalbu@uvic.ca

Abstract

Images captured underwater often suffer from sub-optimal illumination settings that can hide important visual features, reducing their quality. We present a novel single-image low-light underwater image enhancer, L²UWE, that builds on our observation that an efficient model of atmospheric lighting can be derived from local contrast information. We create two distinct models and generate two enhanced images from them: one that highlights finer details, the other focused on darkness removal. A multi-scale fusion process is employed to combine these images while emphasizing regions of higher luminance, saliency and local contrast. We demonstrate the performance of L²UWE by using seven metrics to test it against seven state-of-the-art enhancement methods specific to underwater and low-light scenes.

1. Introduction

The study of underwater sites is important for the environmental monitoring field, as it provides valuable insight regarding their rich ecosystems and geological structures. Sensors placed used underwater measure a host of properties, such as physical (e.g., temperature, pressure, conductivity) and biological (e.g., levels of chlorophyll and oxygen); visual data are captured by cameras, and acoustic data are collected with hydrophones. Cabled ocean observatories have installed and maintained various sensor layouts that allow for the continuous monitoring of underwater regions over extended time series.

Cabled ocean observatories have captured thousands of hours of underwater videos from fixed and mobile platforms [39]. The manual interpretation of these data requires a prohibitive amount of time, highlighting the necessity for semi- and fully-automated methods for the annotation of marine imagery. Mallet *et al.* [33] show that the use of underwater video cameras and their associated software in marine ecol-

ogy has considerably grown in the last six decades. The efficient interpretation of underwater images requires that they maintain a certain level of *quality* (i.e., contrast, sharpness and color fidelity), which is often not possible.

Underwater seafloor cameras often can not count on sunlight, prompting the use of artificial means of illumination. These artificial sources are not able to uniformly illuminate the entirety of a scene, and are typically employed in groups. These different and non-uniform lighting setups call for the use of specialized models of atmospheric lighting in image enhancement efforts. Low levels of lighting reduce the quality of visual data because contrast, color and sharpness are deteriorated, making it difficult to detect important features such as edges and textures. Additional challenges posed to underwater image's quality are related to physical properties of the water medium: *absorption*, which selectively reduces the energy of the propagated light based on its wavelength, and *scattering*. The combined effect of these degradation factors results in images with dark regions, low contrast and hazy appearance.

The proposed single-image system, L²UWE, offers a novel methodology for the enhancement of low-light underwater images. Its novelty is based on the observation that low-light scenes present particular local illumination profiles that can be efficiently inferred by considering local levels of contrast. We propose two contrast-guided atmospheric illumination models that can harvest the advantages of underwater darkness removal systems such as [39], while preserving colors and enhancing important visual features (e.g., edges, textures). By combining these outputs via a multi-scale fusion process [7] we highlight regions of high contrast, saliency and color saturation on the final result.

The performance of L²UWE is compared to that of five underwater-specific image enhancers [10, 16, 19, 20, 39], as well as two low-light-specific enhancers [22, 49] using the OceanDark dataset [39]. Seven distinct metrics (i.e., UIQM [37], PCQI [45], FADE [17], GCF [34], *r* and *e*-score [23], SURF features [9]) show that L²UWE outperforms the compared methods, achieving a significant en-

hancement in overall image visibility (by reducing low-light regions) and emphasizing the image features (e.g., edges, corners, textures).

2. Related work

Sub-section 2.1 discusses works in three areas of relevance to the development of L²UWE: underwater image enhancement, aerial image dehazing and low-light image enhancement. The sub-section that follows, 2.2, summarizes the single image dehazing framework of He *et al.* [24], the multi-scale fusion approach of Ancuti and Ancuti [7], and the contrast-guided low-light underwater enhancement system of Marques *et al.* [39].

2.1. Background

Underwater image enhancement. Some early approaches that attempted the enhancement of underwater images include: the color correction scheme of Chambah *et al.* [14] that aimed for a better detection of fish, the work of Iqbal *et al.* [27], which focused in adjusting the contrast, saturation and intensity to boost images' quality, and the method of Hitam *et al.* [26], where the equalization of the histogram from underwater images is proposed as a means to achieve enhancement. More recently, Ancuti *et al.* [6] introduced a popular framework that derived two inputs (a color-corrected and a contrast-enhanced version of the degraded image), as well as four weight maps that guaranteed that a fusion of the inputs generated would have good sharpness, contrast and color distribution.

Multiple underwater image processing methods [15, 47, 5, 8, 38, 19, 10] make use of aerial dehazing techniques, given that the issues that plague hazy images (absorption and scattering) create outputs that are similar to those captured underwater. Fu *et al.* [20] proposed a framework based on the Retinex model that enhances underwater images by calculating their detail and brightness, as well as performing color correction. Berman *et al.* [10] used the color attenuation and different models of water types to create a single underwater image enhancer. Cho and Kim [16], inspired by simultaneous location and mapping (SLAM) challenges, derived an underwater-specific degradation model. Drews *et al.* [19] considered only two color channels when using the dehazing approach of He *et al.* [24], adapting this method to underwater scenes. Marques *et al.* [39] introduced a contrast-based approach inspired in the Dark Channel Prior [24] that significantly improved the quality of low-light underwater images.

Aerial images dehazing. Dehazing methods aim for the recovery of a scene's original radiance intensity and the correction of color shifts caused by scattering and absorption of light by fluctuating particles and water droplets. Initial approaches proposed to address this challenge [35, 42, 36, 30, 44] were limited by the need of multiple im-

ages (captured under different weather conditions) as input. He *et al.* [24] proposed a popular method for single-image dehazing that introduced the *Dark Channel* and the *Dark Channel Prior* (DCP), which allow for the estimation of a scene's atmospheric lighting and transmission map, important parameters of the dehazing process (as detailed in sub-section 2.2). Comparative results [3, 40, 2, 39] attest to this method's performance in scenarios where only a single hazy image is available. Numerous underwater-specific enhancement frameworks [39, 19, 13, 15, 21, 32] are based on variations of the DCP. Recently a number of data-driven methods [12, 40, 50, 31, 41, 48] focused on the use of Convolution Neural Networks (CNNs) to train systems capable of efficiently performing the dehazing/image recovering task. However, these systems typically require time-consuming processes of data gathering and preparation, hyper-parameter tuning, training, and evaluation.

Low-light image enhancement. Dong *et al.* [18] observed that dark regions in low-light images are visually similar to haze in their inverted versions. The authors proposed to use the DCP-based dehazing method to remove such haze. Ancuti *et al.* [4] proposed to use a non-uniform lighting distribution model and the DCP-based dehazing method to generate two inputs (an additional input is the Laplacian of the original image), followed by a multi-scale fusion process. Zhang *et al.* [49] introduced the maximum reflectance prior, which states that the maximum local intensity in low-light images depends solely on the scene illumination source. This prior is used to estimate the atmospheric lighting model and transmission map of a dehazing process (refer to sub-section 2.2 for details). Guo *et al.* [22] proposed LIME, where the atmospheric lighting model is also initially constructed by finding the maximum intensity throughout the color channels. The LIME framework refines this initial model by introducing a structure prior that guarantees structural fidelity to the output, as well as smooth textural details. Data-driven approaches were also proposed for the enhancement of low-light images [28, 46, 43, 29]. These methods employ CNNs to extract features from datasets composed of low-light/normal-light pairs of images ([29] requires only the degraded images), and then train single-image low-light enhancement frameworks.

2.2. Previous Works Supporting the Proposed Approach

Dark Channel Prior-based dehazing of single images. Equation 1 describes the formation of a hazy image I as the sum of two additive components, *direct attenuation* and *airlight*.

$$I(x) = J(x)t(x) + A_\infty(1 - t(x)) \quad (1)$$

where J represents the haze-less version of the image,

x is an spatial location, transmission map t indicates the amount of light that reaches the optical system and A_∞ is an estimation of the global atmospheric lighting. The first term, direct attenuation $D(x) = J(x)t(x)$, indicates the attenuation suffered by the scene radiance due to the properties of the medium. The second term, airlight $V(x) = A_\infty(1 - t(x))$, is due to previously scattered light and may result in color shifts on the hazy image. The goal of dehazing efforts is to find the haze-free version of the image (J) by determining A_∞ and t .

He *et al.* [24] introduced the *Dark Channel* and the *Dark Channel Prior* (DCP), which can be used to infer atmospheric lighting A_∞ and to derive a simplified formula for the calculation of t . The dark channel is described in Equation 2.

$$J^{dark}(x) = \min_{y \in \Omega(x)} (\min_{c \in \{r,g,b\}} I^c(y)) \quad (2)$$

where x and y represent two pairs of spatial coordinates in the dark channel J^{dark} and in the hazy image I (or any other arbitrary image), respectively. The intensity of each pixel in the single-channel image J^{dark} is the lowest value between the pixels inside patch Ω in I^c , where $c \in \{R, G, B\}$. The DCP is an empirical observation stating that pixels from non-sky regions have at least one significantly low intensity across the three color channels. Thus the dark channel is expected to be mostly dark (i.e., intensities closer to 0).

A single three-dimensional global atmospheric lighting vector A_∞^c ($c \in \{R, G, B\}$) can be inferred by looking at the 0.1% [24] or 0.2% [39] brightest pixels in the dark channel (which represent the most haze-opaque regions from input), then considering the brightest intensity pixels in these same spatial coordinates from the *input* image I . Ancuti *et al.* [4] observed that a single global value might not properly represent the illumination of low-light scenes, thus proposing the estimation of local atmospheric light intensities $A_{L\infty}^c$ inside patches Ψ following Equation 3.

$$A_{L\infty}^c(x) = \max_{y \in \Psi(x)} (\min_{z \in \Omega(y)} (I^c(z))) \quad (3)$$

where x , y and z represent, respectively, spatial coordinates in the local atmospheric image, “minimum” image $I_{\min}(z) = \min_{z \in \Omega(y)} (I(z))$, and hazy image I . For each color channel $c \in \{R, G, B\}$, the local atmospheric lighting $A_{L\infty}^c$ is calculated by first finding I_{\min}^c , which represents the minimum intensities inside patch Ω on I^c , and then calculating the maximum intensities inside a patch Ψ on I_{\min}^c . By arguing that the influence of lighting sources goes beyond patch Ω , Ancuti *et al.* [4] used patch Ψ with twice the size of Ω . The DCP is used in [24] to derive Equation 4 for the calculation of the transmission map t .

$$t(x) = 1 - \omega \min_{y \in \Omega(x)} (\min_{c \in \{r,g,b\}} \frac{I^c(y)}{A_\infty^c}) \quad (4)$$

where A_∞^c indicates the atmospheric lighting in the range [0, 255], resulting in a normalized image ($\frac{I^c(y)}{A_\infty^c}$) ranging from [0, 1]. The constant ω ($0 \leq \omega \leq 1$) preserves a portion of the haze, generating more realistic output. Note that for local estimation of atmospheric lighting, A_∞^c would be substituted by $A_{L\infty}^c$ in Equation 4. The haze-less version of the image, $J(x)$, is obtained using Equation 5 [24].

$$J^c(x) = \frac{I^c(x) - A_\infty^c}{\max(t(x), t_0)} + A_\infty^c \quad (5)$$

Since the direct attenuation $D(x) = J(x)t(x)$ can be close to zero when $t(x) \approx 0$, [24] adds the t_0 term as a lower bound to $t(x)$, effectively preserving small amounts of haze in haze-dense regions of I .

Contrast-guided approach for the enhancement of low-light images. Marques *et al.* [39] observed and addressed three problems that arrive from the use of single-sized patches Ω throughout the image dehazing process: 1) small patch sizes would result in oversaturation of the radiance from the recovered scene (non-natural colors); 2) large patch sizes better estimate and eliminate haze, but since they consider that the transmission profile (i.e., amount of light that reaches a camera) is constant inside patch Ω , undesired halos might appear around intensity discontinuities and 3) a single patch size is typically not optimal for images of varying scales.

Marques *et al.* [39] argues that homogeneous regions of an image possess lower contrast, and that the assumption that their transmission profile is approximately the same holds for patches Ω of varying sizes (in particular, from 3×3 up to 15×15). In these scenarios, the use of larger patch sizes generates darker dark channels (strengthening the DCP) and are, therefore, preferred. For regions with complex content (i.e., intensity changes), [39] uses smaller patch sizes to capture the transmission profile nuances. To determine the correct patch size for each pixel in image I , [39] introduced the *contrast code image* (*CCI*), whose calculation is summarized in Equation 6.

$$CCI(x) = \arg \min_i [\sigma(\Omega_i(x))] \quad (6)$$

where $\Omega_i(x) \in I$ represents a square patch of size $(2i + 1) \times (2i + 1)$ centered at the pair of spatial coordinates x , $i = \{1, 2, \dots, 7\}$, and σ represents the standard deviation between intensities inside of patch Ω_i (considering the three color channels). Therefore, the *CCI* is populated by i (referred to as code c in [39]), which indicates the size of the patch Ω that generated the smallest σ in each pixel location x . A variable named *tolerance* is also introduced

to incentivize the use of larger patch sizes by changing the measured values of σ for different i .

The CCI is then used to calculate the transmission map and dark channel. For a pixel location x , one would use patches of size $(2c + 1) \times (2c + 1)$ (where $c = CCI(x)$) instead of fixed-size patches. This contrast-guided approach significantly mitigates the three aforementioned problems.

Multi-scale fusion for image enhancement. The work of Ancuti *et al.* [7] proposes to perform dehazing by first calculating two inputs \mathcal{I}^k ($k = \{1, 2\}$) from the original image: a white-balanced and a contrast-enhanced version of it. Then the authors introduced the now-popular multiscale fusion process, where three *weight maps* containing important features of each input \mathcal{I}^k are calculated: 1) **Luminance** \mathcal{W}_L^k : responsible for assigning high values to pixels with good visibility, this weight map is computed by observing the deviation between the R, G and B color channels and luminance L (average of pixel intensities at a given location) of the input; 2) **Chromaticity** \mathcal{W}_C^k : controls the saturation gain on the output image, and can be calculated by measuring the standard deviation across each color channel from the input; 3) **Saliency** \mathcal{W}_S^k : in order to highlight regions with greater saliency, this weight map is obtained by subtracting a Gaussian-smoothed version of the input by its mean value (method initially proposed by Achanta *et al.* [1]). Aiming to minimize visual artifacts introduced by the simple combination of the weight maps, [7] uses a multiscale fusion process where a Gaussian pyramid is calculated for the normalized weight map $\bar{\mathcal{W}}^k$ (i.e., per-pixel division between the multiplication of all three weights and their sum) of each input, while the inputs \mathcal{I}^k themselves are decomposed into a Laplacian pyramid. Given that the number of levels from both pyramids is the same, they can be independently fused using Equation 7.

$$\mathcal{R}_l(x) = \sum_k G_l\{\bar{\mathcal{W}}^k(x)\} L_l\{\mathcal{I}_k(x)\} \quad (7)$$

where l indicates the number of levels (typically 5), $L\{\mathcal{I}\}$ represents the Lapacian of \mathcal{I} , and $G\{\bar{\mathcal{W}}\}$ denotes the Gaussian-smoothed version of $\bar{\mathcal{W}}$. The fused result is a sum of the contributions from different layers, after the application of an appropriate upsampling operator.

The authors applied the multi-scale fusion with different strategies in other works, for example by changing the inputs to gamma-corrected and sharpened versions of a white-balanced underwater image [8] or by calculating new weight maps (e.g., local contrast weight map [4]).

3. Proposed approach

Although providing good enhancement results for low-light underwater images, the contrast-guided approach of [39] oversimplifies the scene atmospheric lighting using a single 3-channel value A_∞ . This common practice, which

assumes a mostly white-colored lighting source, works well with aerial hazy images under sunlight [24], but fails to represent low-light scenarios properly [4], since those can present non-homogeneous and non-white illumination. This misrepresentation in [39] results in images with regions that suffer an excess of darkness removal, generating outputs that are overly bright and have a washed-out appearance, often belittling intensity discontinuities (e.g., edges, textures) that could represent important visual features in other computer vision-based applications. This undesirable phenomenon results in Marques *et al.* method’s low *e-score* [23] and count of SURF [9] features (results presented in [39]).

With L^2UWE we propose a better image enhancement mechanism by deriving more realistic models for underwater illumination. We consider local contrast information as a guiding parameter to derive lighting distribution models under two distinct “lenses”: one very focused (i.e., using smaller local regions) that captures the finer details of the original image, and a second, wider one, which considers larger local regions to create brighter models. Each model drives a different dehazing process, and the outputs are combined using a multi-scale fusion approach. This fusion strategy preserves both the details and darkness removal obtained with the two lighting models. As a result, the output of L^2UWE drastically reduces the amount of darkness from the original images while highlighting their intensity changes. Figure 1 details the computational pipeline of L^2UWE .

3.1. Contrast-aware local atmospheric lighting models for low-light scenes

Using parts of the dark channel to derive a single *global estimate* A_∞ for underwater images creates overly-bright and washed-out results. The underlying assumption in [39] that the lighting in inverted input images is mostly white is not reasonable for underwater scenes: see the “inversion” step of Figure 1, where the lighting presents non-white colors. Given that the images from the OceanDark dataset [39] are captured using artificial (and often multiple) lighting sources, a single global estimate A_∞ can not properly model their non-uniform nature. L^2UWE addresses these problems by calculating the atmospheric lighting for each color channel, and by considering the CCI code at each spatial position x when determining *local estimates* of lighting. This approach is similar to that described in Equation 3 [4], but instead of using a fixed-size patch Ψ , we introduce the contrast-guided patch Υ for lighting calculation.

We observed that regions with high contrast (i.e., lower code c in the CCI) should have their local lighting component modeled by considering a larger Υ , based on the heterogeneous influence that illumination sources place on them. On the other hand, since regions with lower con-

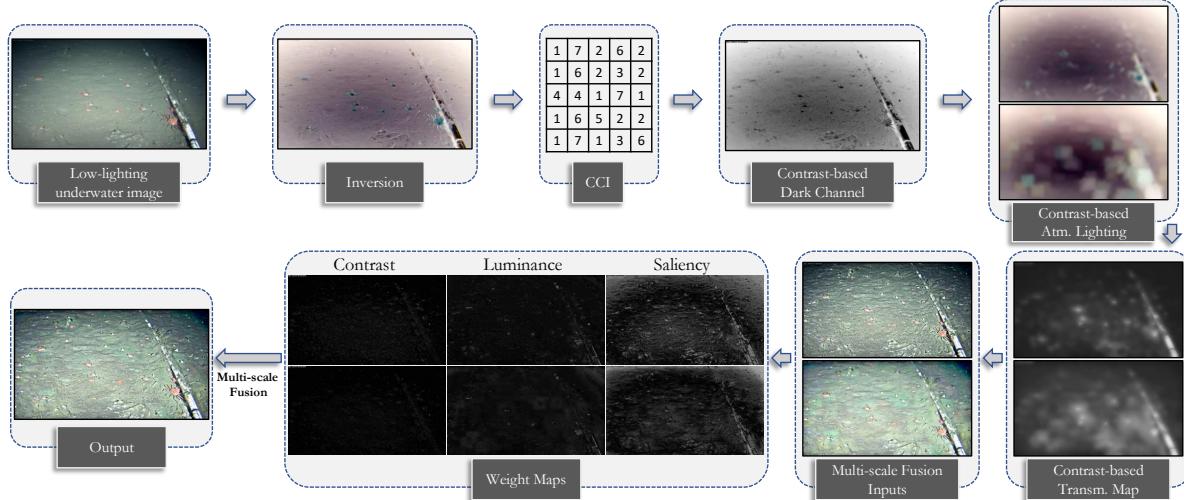


Figure 1. Pipeline of L²UWE. First, the *CCI* [39] is calculated in the inverted version of the image. The contrast-based Dark Channel is then used to derive two atmospheric lighting models that consider local illumination. Two transmission maps generate the inputs of a multi-scale fusion process. Three weight maps are calculated for each input, which are then combined to offer the framework’s output.

trast (i.e., higher codes c in the *CCI*) are illuminated in a roughly homogeneous manner, only smaller patches Υ need to be studied to properly model local illumination. We offer a formalization of this reasoning in Equation 8, which specifies the relationship between *CCI* code c and the size of the lighting square patch Υ used in our local atmospheric lighting model calculation.

$$S_{\Upsilon}(m, c) = 3m - \left[\frac{m}{3}(c-1)\right] \quad (8)$$

where m represents an arbitrary multiplication factor and $c = \{1, 2, \dots, 7\}$ represents code c in the *CCI* at a certain position. Parameter m has a multiplicative effect on the contrast-guided patch, but an offset is also added to progressively constrain it based on the patch size: smaller patches will be more influenced than larger patches. For example, for $m = 15$ and a position x where $CCI(x) = 1$, patch $\Upsilon(x, m)$ would be of dimensions $S_{\Upsilon}(15, 1) \times S_{\Upsilon}(15, 1)$, or 45×45 . Similarly, for $m = 15$ and $CCI(x) = 5$ (lower contrast), patch $\Upsilon(x, m)$ would be of dimensions 25×25 .

With the use of contrast-aware patch Υ and multiplication factor m , we define the local, contrast-guided atmospheric intensity $A_{LCG\infty}^c$ for a color channel c as:

$$A_{LCG\infty}^c(x, m) = \max_{y \in \Upsilon(x, m)} (\min_{z \in \Omega(y)} (I^c(z))) \quad (9)$$

The main distinction between $A_{LCG\infty}^c(x, m)$ (Equation 9) and $A_{L\infty}^c(x)$ (Equation 3) is that the former uses contrast-aware patches $\Upsilon(x, m)$, while the latter uses fixed-size patches $\Psi(x)$. Thus we refer the reader to Equation 3 and its discussion on sub-section 2.2. Since $A_{LCG\infty}^c$ is calculated for each color channel c , by maximizing contrast-aware patch $\Upsilon(x, m)$ one is actually determining a local

position y where the radiance for a certain color channel c is maximum in the I_{min}^c (i.e., a dark channel specific for color c). Figure 2 compares the different atmospheric lighting models obtained using $A_{LCG\infty}$ and A_{∞} . The $A_{LCG\infty}$ is filtered using a Gaussian kernel with $\sigma = 10$ to prevent the creation of abrupt, square-like intensity discontinuities (“halos”) when normalizing the input image by the atmospheric lighting (see Equation 4). Differently from A_{∞} , $A_{LCG\infty}$ captures the color characteristics of the illumination, as well as its local distribution throughout the image.



Figure 2. Different approaches for atmospheric lighting estimation on inverted OceanDark sample. **Left:** a single 3-channel value (A_{∞}), implying that the illumination is homogeneous and roughly white. **Center:** contrast-aware local lighting estimation $A_{LCG\infty}^c$. **Right:** Gaussian-filtered version of $A_{LCG\infty}^c$, preventing the creation of halos in the output.

3.2. Fusion process

The choice of parameter m in Equations 8 and 9 determines the size of local window Υ and therefore the radius of influence from each illumination source. In other words, higher m creates brighter lighting models. While this generates output with less darkness, it also might apply an excess of radiance correction (i.e., bringing intensities close to saturation) in regions of the image and hide important intensity changes (as previously discussed about [39]).

In order to harvest the advantages of both choices, we derive two $A_{LCG\infty}$: one with $m = 5$ and the other choosing $m = 30$. These lighting models are used with Equation 4 to determine two transmission maps. These maps are filtered using a Fast Guided filter [25], and finally used to recover two haze-less versions of the original image (Equation 5). The image enhancement results obtained using each $A_{LCG\infty}$ serve as inputs \mathcal{I}^k ($k = \{1, 2\}$) to a multi-scale fusion process. Figure 3 shows two inputs generated with $m = \{5, 30\}$ for an OceanDark sample.



Figure 3. Multi-scale process inputs. **Left:** original OceanDark sample. **Center:** first input, an image enhanced using the contrast-guided CCI and $A_{LCG\infty}$ with $m = 5$. **Right:** second input, an image enhanced using the contrast-guided CCI and $A_{LCG\infty}$ with $m = 30$.

By fusing the inputs generated from different $A_{LCG\infty}$, we preserve two important aspects of the enhanced images: the efficient darkness removal of the input obtained with $m = 30$ (Figure 3 right) and the edges, textures and overall intensity changes of the input generated with $m = 5$ (Figure 3 center). Experiments evaluating different pairs of m values yielded the best performance when using $m = 5$ and $m = 30$. We also performed tests using a higher number of inputs, but similar results were obtained as for fusing the inputs corresponding to $m = 5$ and $m = 30$, at the expense of a higher computational complexity.

In order to properly combine the two inputs, we calculate three weight maps from each of them: saliency, luminance and local contrast. These weight maps guarantee that regions in the inputs that have high saliency and contrast, or that possess edges and texture variations will be emphasized in the fused output [7, 4].

As discussed in sub-section 2.2, the **saliency weight map** is calculated by subtracting a Gaussian-smoothed version of input k , $\mathcal{I}_k^{G_s}$, by the mean intensity value of this same input, \mathcal{I}_k^μ (constant for each input), as detailed in Equation 10.

$$\mathcal{W}_S^k(x) = \| \mathcal{I}_k^{G_s}(x) - \mathcal{I}_k^\mu \| \quad (10)$$

where x represents a spatial location of input k and $\mathcal{I}_k^{G_s}$ is obtained with a 5×5 ($\frac{1}{16}[1, 4, 6, 4, 1]$) Gaussian kernel.

Considering that saturated colors present higher values in one or two of the R, G, B color channels [7], we use Equation 11 to calculate the **luminance weight map** \mathcal{W}_L^k .

$$\mathcal{W}_L^k = \sqrt{\frac{1}{3}[(R^k - L^k)^2 + (G^k - L^k)^2 + (B^k - L^k)^2]} \quad (11)$$

where L^k represents, at each spatial position, the mean of R, G, B intensities for input k . R^k, G^k and B^k are the three color channels of input \mathcal{I}^k .

The **local contrast weight map** \mathcal{W}_{LCOn}^k , also used in [4], is responsible for highlighting regions of input \mathcal{I}^k where there is more local intensity variation. We calculate it by applying a $\frac{1}{8}[-1 -1 -1; -1 8 -1; -1 -1 -1]$ Laplacian kernel on L^k .

The three weight maps ($\mathcal{W}_{LCOn}^k, \mathcal{W}_L^k, \mathcal{W}_S^k$) are combined into a normalized weight map \mathcal{W}^k , from which a 5-level Gaussian pyramid $G\{\mathcal{W}^k\}$ is derived. Our choice of using Gaussian pyramids is based on their efficacy in representing weight maps, as demonstrated by Aucutie et al. [7]. Figure 4 illustrates the three weight maps calculated for one input image, as well as their equivalent normalized weight map.

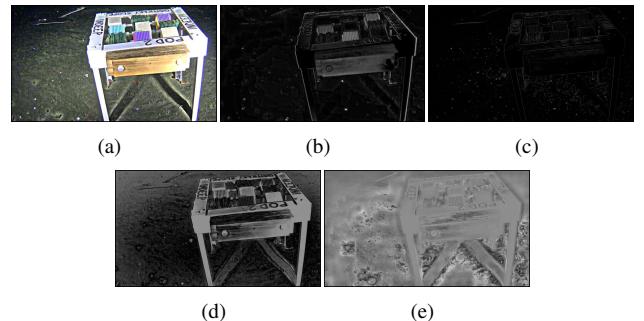


Figure 4. Weight maps of an input \mathcal{I}^k . **(a)** Input image. **(b)** Luminance weight map \mathcal{W}_L^k . **(c)** Local contrast weight map \mathcal{W}_{LCOn}^k . **(d)** Saliency weight map \mathcal{W}_S^k . **(e)** Normalized weight map \mathcal{W}^k .

Following the procedure of [4, 7, 8, 11], each input \mathcal{I}^k is decomposed into a 5-level Laplacian pyramid $L\{\mathcal{I}^k\}$. The multi-scale fusion process is then carried out using Equation 7. The output of L²UWE (Figure 5(d)) reduces the darkness from the original image, retains colors and enhances important visual features.

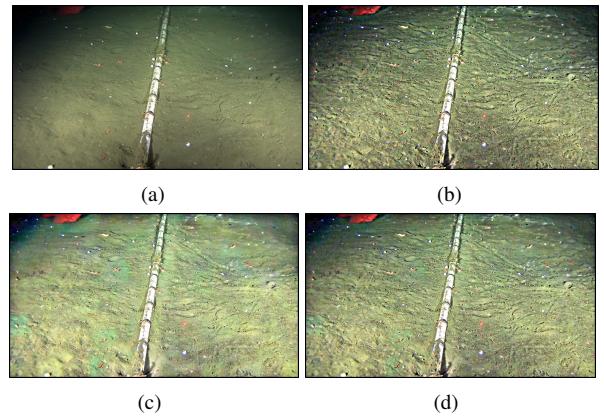


Figure 5. **(a)** Original sample from OceanDark. **(b)** Input calculated with $m = 5$. **(c)** Input calculated with $m = 30$. **(d)** Final output of L²UWE, a multi-scale fusion between the two inputs.

4. Experimental results

The performance of L²UWE is evaluated using the OceanDark dataset [39], composed by 183 samples of low-light underwater images captured by Ocean Networks Canada (ONC). Seven metrics are used in a comprehensive comparison between L²UWE and five underwater-specific image enhancers, as well as two low-light-specific image enhancers.

Metrics. Evaluating the results of image enhancers in datasets without ground truth (such as OceanDark [39]) is non-trivial. Individual metrics can not, alone, indicate the performance of the enhancement; e.g., the **e-score** [23] measures the number of new visible edges obtained after enhancement, which can represent noise, while the FADE score [17] can be used to measure darkness, but it does not account for visual features lost because of over-illumination. Thus we employed a group of seven metrics in the evaluation. **UIQM** [37]: inspired on the human visual system, this no-reference underwater image quality indicator combines measures of colorfulness, sharpness and contrast; **PCQI** [45]: a method developed to assess the quality of contrast-changed images, PCQI considers local contrast quality maps; **GCF** [34]: this metric reflects the level of contrast present in the whole image (dehazed images typically present higher contrast); **e- and r-scores** [23]: indicate, based on the original and enhanced images, the increase in number of visible edges, and the boost in gradient value for these edge’s pixels (“visibility”), respectively; **FADE** [17]: measures the amount of perceived fog (or darkness in inverted images), thus assigning, in our analysis, lower values to better-illuminated images; **SURF** [9]: a popular method that extracts useful features for image matching, reconstruction, stitching, object detection, among others.

Comparison with state-of-the-art methods. We evaluate L²UWE against five frameworks designed specifically for the enhancement of underwater images: Marques *et al.* [39], Berman *et al.* [10], Drews *et al.* [19], Fu *et al.* [20] and Cho *et al.* [16]. Since OceanDark [39] is composed of low-light images, we also considered two popular low-light-specific image enhancers: Zhang *et al.* [49] and Guo *et al.* [22]. All methods are discussed in sub-section 2.1. The implementations used are those made publicly available by the authors.

Qualitative analysis. Figure 6 shows that L²UWE is able to generate output images that highlights important visual features (e.g., fishes that were not visible, small rocks and overall geography of the sites) of the original images without excessively brightening the scenes. While the methods of Marques *et al.* [39] and Guo *et al.* [22] yielded similar results that greatly reduced low-light regions, they also concealed important visual features because of a lighting over-correction: close-to-saturation pixel in-

tensities might hide the image’s finer details, such as edges and textures. The methods of Drews *et al.* [19], Zhang *et al.* [49] and Cho *et al.* [16] actually darkened the images, contrary to the goal of the enhancement. The methods of Zhang *et al.* [49] and Fu *et al.* [20] highlighted edges and textures of the outputs, but were not able to properly illuminate dark regions. The unnatural colors generated by the method of Berman *et al.* [10] can be attributed to the automatic, but often sub-optimal, choice of water type done for each sample.

Quantitative analysis. L²UWE obtained the highest score on the UIQM [37] metric, confirming the human-visual-system-inspired perceived image quality of its outputs. The quality of the contrast modification measured by PCQI [45] also favors the results obtained with the proposed method. The framework of Zhang *et al.* [49] obtained the highest GCF [34], however a qualitative analysis of its results indicates that this is due to the strong saturation of the output, which creates a high *global* contrast with respect to the undesired dark regions still present in the enhanced images. The highest increase in visible edges (**e-score** [23]) and SURF features [9] is attributed to L²UWE (we reiterate that these two metrics should be interpreted cautiously, as they may reflect noise introduced to the image). Finally, the highest scores in the metrics related to visibility (**r-score** [23]) and FADE [17] indicate that L²UWE successfully achieved the goal of increasing the illumination of low-light underwater images while preserving their colors and highlighting their salient visual structures.

5. Conclusion

Our novel single-image framework, L²UWE, uses a contrast-guided approach for the efficient modelling of lighting distributions in low-light underwater scenes. It then generates two dehazed inputs that are combined employing a multi-scale fusion process, ultimately reducing dark regions and highlighting important visual features of the original image without changing its color distribution.

Experimental results show the proposed method’s capacity of drastically reducing low-light regions of inputs without creating washed-out outputs (see Figure 6). Although other methods can be used for the enhancement of similar scenarios with remarkable results, our experiments show that L²UWE outperforms the current state-of-the-art approaches in the task of enhancing low-light underwater images. Our proposed contrast-guided computational pipeline is also expected to work well in other mediums, such as night-time and low-light aerial scenes.

Future developments will focus on the adaptation of L²UWE for aerial low-light images, its evaluation in additional datasets, and the use of data-driven systems in the framework’s computational pipeline (e.g., a CNN-based network for the estimation of atmospheric lighting models).



Figure 6. Output of diverse image enhancement methods on samples from the OceanDark dataset [39].

Method	UIQM [37]	PCQI [45]	GCF [34]	e-score [23]	r-score [23]	FADE [17]	SURF [9]
Original	0.88 ± 0.13	1	3.28 ± 0.62	N/A	1	1.91 ± 0.79	340 ± 293
Marques [39]	0.99 ± 0.12	0.85 ± 0.03	3.41 ± 0.71	0.28 ± 0.32	2.75 ± 0.76	0.46 ± 0.18	705 ± 470
Berman [10]	1.00 ± 0.18	0.78 ± 0.15	3.84 ± 1.07	0.25 ± 0.50	2.91 ± 1.96	1.15 ± 0.40	425 ± 317
Fu [20]	0.93 ± 0.15	0.93 ± 0.09	3.28 ± 0.57	0.09 ± 0.39	1.72 ± 0.35	1.75 ± 0.25	865 ± 478
Cho [16]	1.24 ± 0.15	0.87 ± 0.04	4.11 ± 0.84	0.89 ± 0.54	1.72 ± 0.09	1.81 ± 0.52	751 ± 428
Drews [19]	1.38 ± 0.14	0.85 ± 0.05	4.70 ± 0.89	1.06 ± 0.83	1.29 ± 0.31	2.08 ± 0.90	589 ± 324
Zhang [49]	1.28 ± 0.08	1.03 ± 0.07	6.34 ± 0.74	1.48 ± 0.88	3.70 ± 1.00	0.72 ± 0.39	1719 ± 620
Guo [22]	0.93 ± 0.13	0.86 ± 0.03	3.50 ± 0.73	0.16 ± 0.14	2.21 ± 0.64	0.60 ± 0.24	607 ± 452
L ² UWE	1.38 ± 0.11	1.17 ± 0.07	4.89 ± 0.66	1.82 ± 0.83	4.61 ± 0.58	0.42 ± 0.20	1856 ± 655

Table 1. Mean and standard deviation of seven metrics computed on all samples of the OceanDark dataset [39]. The results compare the output of diverse image enhancing methods. Best results (i.e., higher, with the exception of darkness-indicating FADE [17]) are bolded.

References

- [1] Radhakrishna Achanta, Sheila Hemami, Francisco Estrada, and Sabine Susstrunk. Frequency-tuned salient region detection. In *2009 IEEE conference on computer vision and pattern recognition*, pages 1597–1604. IEEE, 2009. [4](#)
- [2] E. M. Alharbi, P. Ge, and H. Wang. A research on single image dehazing algorithms based on dark channel prior. *Journal of Computer and Communications*, 4(02):47, 2016. [2](#)
- [3] C. Ancuti, C. O. Ancuti, and C. De Vleeschouwer. D-hazy: a dataset to evaluate quantitatively dehazing algorithms. In *IEEE International Conference on Image Processing (ICIP)*, pages 2226–2230. IEEE, 2016. [2](#)
- [4] Cosmin Ancuti, Codruta O Ancuti, Christophe De Vleeschouwer, and Alan C Bovik. Night-time dehazing by fusion. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 2256–2260. IEEE, 2016. [2, 3, 4, 6](#)
- [5] Cosmin Ancuti, Codruta O Ancuti, Christophe De Vleeschouwer, Rafael Garcia, and Alan C Bovik. Multi-scale underwater descattering. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 4202–4207. IEEE, 2016. [2](#)
- [6] Cosmin Ancuti, Codruta Orniana Ancuti, Tom Haber, and Philippe Bekaert. Enhancing underwater images and videos by fusion. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 81–88. IEEE, 2012. [2](#)
- [7] Codruta Orniana Ancuti and Cosmin Ancuti. Single image dehazing by multi-scale fusion. *IEEE Transactions on Image Processing*, 22(8):3271–3282, 2013. [1, 2, 4, 6](#)
- [8] Codruta O Ancuti, Cosmin Ancuti, Christophe De Vleeschouwer, and Philippe Bekaert. Color balance and fusion for underwater image enhancement. *IEEE Transactions on image processing*, 27(1):379–393, 2017. [2, 4, 6](#)
- [9] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3):346–359, 2008. [1, 4, 7, 8](#)
- [10] D. Berman, T. Treibitz, and S. Avidan. Diving into hazelines: Color restoration of underwater images. In *Proceedings of the British Machine Vision Conference (BMVC)*, volume 1, 2017. [1, 2, 7, 8](#)
- [11] Peter Burt and Edward Adelson. The laplacian pyramid as a compact image code. *IEEE Transactions on communications*, 31(4):532–540, 1983. [6](#)
- [12] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11):5187–5198, 2016. [2](#)
- [13] Nicholas Carlevaris-Bianco, Anush Mohan, and Ryan M Eu-stice. Initial results in underwater single image dehazing. In *Oceans 2010 Mts/IEEE Seattle*, pages 1–8. IEEE, 2010. [2](#)
- [14] Majed Chambah, Dahbia Semani, Arnaud Renouf, Pierre Courtellemont, and Alessandro Rizzi. Underwater color constancy: enhancement of automatic live fish recognition. In *Color Imaging IX: Processing, Hardcopy, and Applications*, volume 5293, pages 157–169. International Society for Optics and Photonics, 2003. [2](#)
- [15] John Y Chiang and Ying-Ching Chen. Underwater image enhancement by wavelength compensation and dehazing. *IEEE transactions on image processing*, 21(4):1756–1769, 2011. [2](#)
- [16] Younggun Cho and Ayoung Kim. Visibility enhancement for underwater visual slam based on underwater light scattering model. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 710–717. IEEE, 2017. [1, 2, 7, 8](#)
- [17] Lark Kwon Choi, Jaehee You, and Alan Conrad Bovik. Referenceless prediction of perceptual fog density and perceptual image defogging. 24(11):3888–3901, 2015. [1, 7, 8](#)
- [18] X. Dong, G. Wang, Y. Pang, W. Li, J. Wen, W. Meng, and Y. Lu. Fast efficient algorithm for enhancement of low lighting video. In *2011 IEEE International Conference on Multimedia and Expo*. IEEE, 2011. [2](#)
- [19] Paul Drews, Erickson Nascimento, Filipe Moraes, Silvia Botelho, and Mario Campos. Transmission estimation in underwater single images. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 825–830, 2013. [1, 2, 7, 8](#)
- [20] Xueyang Fu, Peixian Zhuang, Yue Huang, Yinghao Liao, Xiao-Ping Zhang, and Xinghao Ding. A retinex-based enhancing approach for single underwater image. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 4572–4576. IEEE, 2014. [1, 2, 7, 8](#)
- [21] Adrian Galdran, David Pardo, Artzai Picón, and Aitor Alvarez-Gila. Automatic red-channel underwater image restoration. *Journal of Visual Communication and Image Representation*, 26:132–145, 2015. [2](#)
- [22] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on Image Processing*, 26(2):982–993, 2016. [1, 2, 7, 8](#)
- [23] Nicolas Hautiere, Jean-Philippe Tarel, Didier Aubert, and Eric Dumont. Blind contrast enhancement assessment by gradient ratioing at visible edges. *Image Analysis & Stereology*, 27(2):87–95, 2011. [1, 4, 7, 8](#)
- [24] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12):2341–2353, 2011. [2, 3, 4](#)
- [25] K. He, J. Sun, and X. Tang. Guided image filtering. *IEEE transactions on pattern analysis & machine intelligence*, 35(6):1397–1409, 2013. [6](#)
- [26] Muhammad Suzuri Hitam, Ezmahamrud Afreen Awalludin, Wan Nural Jawahir Hj Wan Yussof, and Zainuddin Bachok. Mixture contrast limited adaptive histogram equalization for underwater image enhancement. In *2013 International conference on computer applications technology (ICCAT)*, pages 1–5. IEEE, 2013. [2](#)
- [27] Kashif Iqbal, Rosalina Abdul Salam, Azam Osman, and Abdullah Zawawi Talib. Underwater image enhancement using an integrated colour model. *IAENG International Journal of Computer Science*, 34(2), 2007. [2](#)
- [28] Lincheng Jiang, Yumei Jing, Shengze Hu, Bin Ge, and Weidong Xiao. Deep refinement network for natural low-light image enhancement in symmetric pathways. *Symmetry*, 10(10):491, Oct 2018. [2](#)

- [29] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. *arXiv preprint arXiv:1906.06972*, 2019. 2
- [30] Johannes Kopf, Boris Neubert, Billy Chen, Michael Cohen, Daniel Cohen-Or, Oliver Deussen, Matt Uyttendaele, and Dani Lischinski. Deep photo: Model-based photograph enhancement and viewing. *ACM transactions on graphics (TOG)*, 27(5):1–10, 2008. 2
- [31] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4770–4778, 2017. 2
- [32] Huimin Lu, Yujie Li, Lifeng Zhang, and Seiichi Serikawa. Contrast enhancement for images in turbid water. *JOSA A*, 32(5):886–893, 2015. 2
- [33] Delphine Mallet and Dominique Pelletier. Underwater video techniques for observing coastal marine biodiversity: a review of sixty years of publications (1952–2012). *Fisheries Research*, 154:44–62, 2014. 1
- [34] Kresimir Matkovic, László Neumann, Attila Neumann, Thomas Psik, and Werner Purgathofer. Global contrast factor-a new approach to image contrast. *Computational Aesthetics*, 2005:159–168, 2005. 1, 7, 8
- [35] S. G. Narasimhan and S. K. Nayar. Chromatic framework for vision in bad weather. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 598–605. IEEE, 2000. 2
- [36] S. G. Narasimhan and S. K. Nayar. Contrast restoration of weather degraded images. *IEEE transactions on pattern analysis and machine intelligence*, 25(6):713–724, 2003. 2
- [37] Karen Panetta, Chen Gao, and Sos Agaian. Human-visual-system-inspired underwater image quality measures. *IEEE Journal of Oceanic Engineering*, 41(3):541–551, 2015. 1, 7, 8
- [38] Yan-Tsung Peng, Xiangyun Zhao, and Pamela C Cosman. Single underwater image enhancement using depth estimation based on blurriness. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 4952–4956. IEEE, 2015. 2
- [39] Tunai Porto Marques, Alexandra Branzan Albu, and Maia Hoeberichts. A contrast-guided approach for the enhancement of low-lighting underwater images. *Journal of Imaging*, 5(10):79, 2019. 1, 2, 3, 4, 5, 7, 8
- [40] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang. Single image dehazing via multi-scale convolutional neural networks. In *European conference on computer vision*, pages 154–169. Springer, 2016. 2
- [41] Wenqi Ren, Lin Ma, Jiawei Zhang, Jinshan Pan, Xiaochun Cao, Wei Liu, and Ming-Hsuan Yang. Gated fusion network for single image dehazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3253–3261, 2018. 2
- [42] Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar. Instant dehazing of images using polarization. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, page 325. IEEE, 2001. 2
- [43] Liang Shen, Zihan Yue, Fan Feng, Quan Chen, Shihao Liu, and Jie Ma. Msr-net: Low-light image enhancement using deep convolutional network. *arXiv preprint arXiv:1711.02488*, 2017. 2
- [44] Tali Treibitz and Yoav Y Schechner. Polarization: Beneficial for visibility enhancement? In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 525–532. IEEE, 2009. 2
- [45] Shiqi Wang, Kede Ma, Hojatollah Yeganeh, Zhou Wang, and Weisi Lin. A patch-structure representation method for quality assessment of contrast changed images. *IEEE Signal Processing Letters*, 22(12):2387–2390, 2015. 1, 7, 8
- [46] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018. 2
- [47] H.-Y. Yang, P.-Y. Chen, C.-C. Huang, Y.-Z. Zhuang, and Y.-H. Shiao. Low complexity underwater image enhancement based on dark channel prior. In *Second International Conference on Innovations in Bio-inspired Computing and Applications (IBICA)*, pages 17–20. IEEE, 2011. 2
- [48] He Zhang and Vishal M Patel. Density-aware single image de-raining using a multi-stream dense network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 695–704, 2018. 2
- [49] Jing Zhang, Yang Cao, Shuai Fang, Yu Kang, and Chang Wen Chen. Fast haze removal for nighttime image using maximum reflectance prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7418–7426, 2017. 1, 2, 7, 8
- [50] Jiawei Zhang, Jinshan Pan, Wei-Sheng Lai, Rynson WH Lau, and Ming-Hsuan Yang. Learning fully convolutional networks for iterative non-blind deconvolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3817–3825, 2017. 2