Analysis of Underwater Image Processing Methods for Annotation in Deep Learning Based Fish Detection

JOSE-LUIS LISANI¹, (Member, IEEE), ANA-BELÉN PETRO¹, CATALINA SBERT¹, AMAYA ÁLVAREZ-ELLACURÍA², IGNACIO A. CATALÁN², AND MIQUEL PALMER²

¹Ciències Matemàtiques i Informàtica, Universitat de les Illes Balears, Palma de Mallorca, 07122 Illes Balears, Spain
²IMEDEA, UIB-CSIC, Esporles, 07190 Illes Balears, Spain

Corresponding author: Ana-Belén Petro (anabelen.petro@uib.es)

This work was supported by the Comunitat Autonoma de les Illes Balears through the Direcció General de Política Universitària i Recerca with funds from the Tourist Stay Tax Law ITS 2017-006 under Grant PRD2018/26.

ABSTRACT With the advent of deep-learning (DL) techniques, image annotation has become a fundamental part of the research process. In the case of underwater image annotation, the human in charge of the task is faced with the inherent quality problems of this kind of images. A large number of underwater image enhancement (UIE) methods have been developed aimed at improving the colors and contrast of these images. However, no attention has been paid to the specific problem of image annotation. In this case the global image quality of the processed image is less important than the fact that the objects to be annotated stand out and that their contours are easy to delineate. In this paper we evaluate seven state-of-the-art UIE techniques and rank them, through a subjective approach, according to their utility for the annotation process. The conclusion of our study is that, in general, the model-free Multiscale Retinex algorithm is preferred over more complex techniques that try to model the formation of the underwater images.

INDEX TERMS Deep learning, fish detection, pre-processing, underwater image enhancement.

I. INTRODUCTION

Deep-learning (DL) methods have become immensely popular in the field of computer vision since they can tackle successfully complex tasks such as classification [53], object detection [48], [49] and instance segmentation [18]. They have also been used for low-level tasks such as image restoration [58] or super-resolution [16]. In recent years these techniques have been applied in marine science [11], [36], [47], [50]. It is of particular interest the use of detection algorithms for monitoring fish populations.

One specific requirement of DL techniques is the need of a vast amount of labeled data to train and test the models that capture the hierarchical relations between different image features. Correct models can only be learned if the ground-truth annotations are abundant and of good quality, therefore, obtaining an initial good data annotation is basic to get a successful solution. Even though different annotation tools have been developed [3], the annotation process is a tedious and laborious task, which must be carried out manually, or in a semi-supervised way, and involves hours and even weeks of work.

In the case of underwater images, the depth-dependent attenuation of the light wavelengths and the scattering effect produce color cast, low contrast, noise and haze [31]. Under these circumstances it is particularly challenging for the human annotators to delineate the contours of the objects of interest, e.g. fish. Furthermore, the variety of underwater backgrounds due to, for example, bottom types, adds complexity to fish detection and delineation.

In order to improve these images, a large number of underwater image enhancement (UIE) methods have been developed. Most recent reviews (for example, [55] and [61]) classify UIE methods into three types: restoration (model-based) approaches, enhancement (model-free) approaches and DL (data-driven) approaches. Restoration methods [9],...
[15], [32], and [44] use a model of the degradation process that leads to the formation of the observed image and aim to recover a clean version of the image by inverting this process. Image enhancement techniques [4], [23] waive away the image formation model and use qualitative and quantitative criteria to improve the contrast and color of the images. Finally, DL models are trained using pairs of degraded and undegraded images that provide clues on how the degradation process can be reverted [6], [14].

In general, model-based methods obtain good results when the scene fits the model, and its parameters are accurately estimated. The use of incorrect assumptions about the model leads to visually unpleasing results. Model-free methods permit to deal with a wider type of scenes but in some cases accentuate noise and produce color distortions. Finally, DL approaches suffer from a lack of enough training data. In most cases synthetic images are used for training, which limits the generalization capability of the models and make these approaches fall behind conventional methods [5], [31].

All these methods have been used, in general, for improving the visual quality of the images, aimed at improving their colors and contrast. However, no attention has been paid to the specific problem of image annotation. As Liu et al. [35] mention, “improving the accuracy for the subsequent higher-level detection/classification tasks is one additional objective of enhancement when the UIE algorithms serve as preprocessing step” (p. 4861). In this case the global image quality of the processed image is less important than the fact that the objects to be annotated stand out and that their contours are easy to delineate.

Our goal in this paper is to assess the most common underwater image processing methods under the prism of helping in the annotation task. To our knowledge, this is the first study of this kind, since most previous studies have focused on the visual quality of the results. We shall evaluate seven techniques representative of the state-of-the-art. This number is not arbitrary but determined by the design of the experimental setup (see Section III for details).

Since image annotation is performed by humans, we have conducted a subjective experiment in which 30 participants have ranked the performance of these seven algorithms in terms of the improvement of the features that help in the annotation process. We have focused on the problem of fish detection, therefore the annotation task consists in outlining the shape of the fish. The obtained results have undergone a statistical analysis and some quantitative measures have also been computed to check whether they are correlated with the subjective results.

The paper is organized as follows: in Section II the seven UIE methods selected in our study are described; in Section III the setting of the subjective experiment is presented; the results of the experiment are shown and analyzed in Section IV; finally, the conclusions of our study are exposed in Section V.

II. SELECTED METHODS

The literature on underwater color enhancement methods is very abundant and several surveys and reviews have been published in recent years. We have analyzed thirteen of these surveys ([1], [5], [17], [25], [31], [33], [35], [37], [51], [55], [56], [59], [61]) and selected the most prominent methods.

The taxonomy criteria used by the different surveys are similar and we have collected the salient methods in each category. Some reviews, as [31] and [35], conduct qualitative and quantitative comparisons between the methods, others rank them based on some metrics [17], while others simply describe the methods without comparing them [25]. As some of these articles indicate, the algorithms having the best quantitative results do not always produce visually pleasant images, from the human point of view. Moreover, there is no algorithm that always wins when faced with a large and varied dataset. Taking these considerations into account, we have selected the most cited and better-ranked methods in the indicated surveys, among the two main categories of algorithms: model-free and physical model based. DL methods are excluded since, in most cases, they “fall behind state-of-the-art conventional methods” [5].

A. MODEL-FREE METHODS

These methods aim to improve the contrast and color of images for a better visual quality, modifying image pixel intensity values without any prior knowledge about the underwater conditions or scene structure.

In this group we have considered three image processing techniques which include a color correction method [23], the Retinex algorithm [26], [45] and a fusion based method [4].

1) UNSUPERVISED COLOR CORRECTION (UCM)

We have considered the unsupervised color correction method (UCM) proposed by Iqbal et al. [23]. This method deals with the problem of low contrast and color cast due to underwater scattering and absorption. Color cast is reduced by equalizing the color channels to the blue one. Contrast correction is performed in RGB and HSI color spaces. First, the red color is increased by stretching the red histogram to the maximum and, similarly, the blue color is reduced by stretching the blue histogram to the minimum. Finally, the method stretches saturation and intensity in HSI color space to enhance the color authenticity and to increase the global brightness.

2) MULTISCALE RETINEX (MSR)

Retinex theory models the mechanism of the human vision system allowing humans to perceive the world under different lighting conditions. It attempts to achieve color constancy when the scene is dominated by a certain illumination, which is a common situation in the underwater environment.

Retinex theory was first proposed by Land and McCann [30]. They established that the visual system does
not perceive absolute lightness but rather variations of it in local regions. Based on this idea, Jobson et al. [26] proposed the well-known multiscale Retinex (MSR) defined by

$$\text{MSR}_c(x) = \sum_{n=1}^{N} w_n [\log(I_c(x)) - \log((G_n \ast I_c)(x))],$$

where $I_c$ and $\text{MSR}_c$ are, respectively, the input and output images at channel $c$, $N$ is the number of scales, $w_n$ the weight of each scale and $G_n(x) = C_n \exp \left(-|x|^2/2\sigma_n^2\right)$.

In images which violate the gray-world assumption, that is, images where a certain color may dominate, multiscale Retinex produces grayish images. To solve this problem Jobson et al. [26] proposed a color restoration step which consists on multiplying MSR by a function of the chromaticity.

$$\text{MSRCR}_c(x) = C_c(x)\text{MSR}_c(x), \quad c \in \{R, G, B\}$$

where

$$C_c(x) = \beta \log \left(\frac{\alpha I_c(x)}{\sum_{i \in \{R,G,B\}} I_i(x)}\right)$$

with $\beta$ the gain constant and $\alpha$ which controls the non-linearity. Finally a channel by channel linear stretching is performed so that the minimum and maximum values of each channel are mapped to 0 and 255, respectively.

3) FUSION-BASED ENHANCEMENT (FUSION)

Fusion-based approaches adopt the fusion strategy to merge images with different image characteristics. In [4], the authors combine, using a multiscale strategy, a white-balanced and a sharpened version of the input image.

The white balance compensates the color cast caused by the depth-dependent absorption of colors in underwater images. Under the assumption that the red channel attenuation is the fastest and that the green channel is well preserved, the white balance consists of the following steps (we assume that the range of values for each channel is [0, 1]):

- The red channel $R$ is compensated, at each pixel, as follows:

$$R_c(x) = R(x) + \alpha(\bar{G} - \bar{R})(1 - R(x))G(x)$$

where $\bar{G}$ and $\bar{R}$ represent the mean value of $R$ and $G$ and $\alpha$ is a constant parameter. Moreover in the cases where the blue channel is strongly attenuated, it is also compensated in the same manner

$$B_c(x) = B(x) + \alpha(\bar{G} - \bar{B})(1 - B(x))G(x)$$

- After channel compensation, the image is assumed to follow the gray-world assumption and a classical white-balance algorithm can be used to process the image. The sharpened version of the image $I$ is defined by $S = I + \beta(I - G_{\sigma} \ast I)$, where $G_{\sigma}$ is a Gaussian kernel with standard deviation $\sigma$ and $\beta$ is a parameter.

The final result is obtained by fusing the white-balanced and sharpened images using the multiscale fusion strategy described in [40]. The value at each pixel in the merged result is obtained as a weighted sum of the values of the component images, the weights depending on the local contrast, the saliency level and the color saturation at each pixel.

B. PHYSICAL MODEL-BASED METHODS

These methods set the enhancement of an underwater image as an inverse problem, and estimate the parameters of the imaging formation model from the observation and some prior assumptions.

The most used image formation model is due to Jaffe-McGlamery [24], [39]

$$I_c(x) = J_c(x)\tau_c(x) + A_c(1 - \tau_c(x)), \quad c \in \{R, G, B\} \quad (1)$$

where $I_c$ and $J_c$ are, respectively, the degraded and restored image to be recovered, at channel $c$. In this model there are two parameters that must be estimated, $A$ the global background light and $\tau$ the transmission map defined as

$$\tau_c(x) = e^{-A(x)/d(x)}, \quad c \in \{R, G, B\} \quad (2)$$

where $d(x)$ represents the scene depth and $\beta_c$ is the scattering of the transmission medium.

Model (1) is similar to the one used in haze removal problems, for this reason most underwater imaging methods adopt the same prior assumptions than dehazing methods, with some modifications due to the particularity of underwater images.

The dark channel prior (DCP), proposed by He et al [19], is frequently used to estimate the transmission map. It is based on the observation that most local patches in haze-free outdoor images contain some pixels which have very low intensities in at least one color channel. Mathematically this can be formalized as

$$J_{\text{dark}}(x) = \min_{y \in N(x)} (\min_{c \in \{R,G,B\}} J_c(y)) \approx 0$$

where $J$ is a haze-free image and $N(x)$ is a neighborhood of $x$. According to [19], considering the formation model (1), and assuming the global background light is known, an estimation $\bar{\tau}$ of the transmission map based on the image $I$ is computed by

$$\bar{\tau}(x) = 1 - \min_{y \in N(x)} (\min_{c \in \{R,G,B\}} \frac{I_c(y)}{A_c}),$$

where the authors assume that scattering is wavelength-independent.

Different methods have been proposed to estimate the background light $A$, for instance, He et al. [19] propose to compute it using the 0.1% brightest pixels in the dark channel. Once $A$ and $\tau(x)$ have been estimated, Equation (1) allows to compute the haze-free image $J$.

In the current paper we have selected four models which propose different modifications of DCP and global background light estimation, adapted to underwater images.
1) UNDERWATER DARK CHANNEL PRIOR (UDCP)
As the red light attenuates much faster than the green and blue lights when it propagates in water, Drew et al. in [9] observed that the red channel in an underwater image is unreliable to estimate the dark channel. Under this assumption they proposed a version of the DCP that estimates the dark channel using only the green and blue channels. Using the same method proposed in [19] to estimate the global background light, they proposed to estimate the transmission map as

\[ \hat{t}(x) = 1 - \min_{y \in N(x)} \left( \frac{\min_{c \in \{G,B\}} I_c(y)}{A_c} \right). \]

2) DEHAZING WITH MINIMUM INFORMATION LOSS AND HISTOGRAM DISTRIBUTION PRIOR (InfoLoss)
Li et al. [32] propose an enhancement method in two steps: first, an image dehazing step, followed by a contrast enhancement of the result.

For the dehazing part they based their method on [19], but with some modifications for both the estimation of the background light and the transmission map.

To estimate the global background light the image is divided into four rectangular regions, and the region with the maximum difference between its mean value and its standard deviation is selected. From among the brightest pixels within the selected region, the one with the maximum blue-red difference is chosen and its value is used as an estimation of the global background light.

The transmission map \( \hat{t} \) is estimated in a way that reduces the information loss, that is, the number of pixels mapped outside the range \([0, 255]\) in the output image. First, the transmission map of the most degraded channel (i.e., the red one) is estimated. Then, using this initial estimation and taking advantage of optical properties of underwater imaging [60], the transmission maps of the green and blue channels are also estimated. Finally an adaptive exposure map is employed to adjust the results for better visual quality, preventing that some regions become too dark or too bright.

For the contrast enhancement step the authors propose to adjust the histogram of the haze-free image to the average histogram distribution of natural-scene images.

3) AUTOMATIC RED-CHANNEL UNDERWATER IMAGE RESTORATION (ARC)
In [15] the authors proposed a variant of the dark channel prior [19]. Taking into account that the red channel intensity decays faster as distance increases, they proposed a Red Channel Prior as:

\[ J^{\text{RED}}(x) = \min_{y \in N(x)} (1 - J_r(y)) \]

\[ \min_{y \in N(x)} J_g(x), \min_{y \in N(x)} J_b(x) \approx 0 \]

To estimate the global background light the authors take the value of the pixel in the input image that corresponds to the brightest pixel in \( J^{\text{RED}} \).

Moreover, in order to take into account the presence of artificial illumination in the image, the transmission map is adjusted to effectively enhance the artificial light region and improve the overall color fidelity of the images. Defining saturation as

\[ S(x) = \frac{\max_{c \in \{R,G,B\}} I_c(x) - \min_{c \in \{R,G,B\}} I_c(x)}{\max_{c \in \{R,G,B\}} I_c(x)} \]

the authors propose to estimate the transmission map as

\[ \tilde{t}(x) = 1 - \min_{y \in N(x)} \left( \frac{1 - I_r(y)}{1 - A_r}, \frac{\min_{y \in N(x)} I_p(y)}{A_p}, \frac{\min_{y \in N(x)} I_b(y)}{A_b}, \lambda \min_{y \in N(x)} S(y) \right) \]

for some \( \lambda \in [0, 1] \).

4) UNDERWATER IMAGE RESTORATION BASED ON IMAGE BLURRNESINESS AND LIGHT ABSORPTION (BAL)
Based on the observation that the blur of the objects in the underwater image increases with the depth, the authors in [44] proposed to use this prior and the light absorption property to estimate background light and underwater scene depth.

The image blurriness is computed as

\[ BP(x) = GF(\max_{y \in N(x)} \frac{1}{n} \sum_{i=1}^{n} |I_g(x) - I_g * G_{r_i}(x)|) \]

where \( I_g \) is the grayscale intensity image, \( G_{r_i} \) is a Gaussian filter with standard deviation \( r_i = 2^n + 1 \) and \( n = 4 \). A guided filtering is applied to obtain a smooth blurriness map.

The background light is estimated as a weighted average between the region with lowest variance and the one with the largest blurriness. Regions are selected using a quadtree decomposition.

The scene depth is estimated by sigmoidally combining three depth estimations. The first one is obtained from the red channel map. The second one is based on the assumption that for points closer to the camera the difference between the values of the red channel and the values of the blue and green channels is large. The third one is obtained from the blurriness map. From the estimated depth, the transmission map for the red channel is obtained from (2) with \( \beta \in \left[ \frac{1}{2}, \frac{3}{2} \right] \). Using optical properties of underwater imaging [60], it is estimated the transmission map for the green and blue channels from the red one.

III. EXPERIMENTAL SETTING
As we have mentioned, the main aim of this work is to select the best preprocessing method to improve the underwater image for a correct fish detection. We are aware that in the literature we can find a wide range of metrics to measure the quality of these methods, but we also know, and we can read in other articles (for example, [56] and [35]) that, in general, there exist discrepancies between the qualitative evaluation of the methods and the scores of the different metrics. “The quantitative evaluation system of underwater image quality needs to be improved, (…) and the establishment of a standard quantitative evaluation system is the focus.
of future research” ([59], p. 182276). For this reason and due to the fact that the initial annotation task is carried out manually, we have decided to prioritize the subjective evaluation. As [38] mentions, “Given that the ultimate receivers of images are human eyes, the human subjective opinion is the most reliable value for indicating the image perceptual quality” (p. 626).

The design of a subjective test for underwater image is a challenging task for which no established procedure exists, in contrast with standard quality evaluation [8]. Several issues must be addressed: selection of participants, selection of scenes, selection of processing methods, evaluation protocol, evaluation criteria, and time allotted to each participant to perform the test. A review of several of these tests can be found in [33] and a summary is displayed in Table 1. The most commonly used protocols are: Single Stimulus (SS), Simultaneous Double Stimulus (SDS), and Ranking (R). SDS is a pairwise comparison scheme in which each image is compared to a reference. With the SS method one image is displayed at a time and the participant must assign it a score. Finally, with the R-n protocol, n images are displayed simultaneously and the participant must rank them. The later method is less time consuming since the n images are evaluated simultaneously, while in SDS and SS the images are shown sequentially.

The time constraints for carrying out the test are an important factor in the design, since they influence the selection of the evaluation protocol, the number of chosen scenes and the number of compared methods. In our case, we decided that each participant should be able to perform the test in 1 hour, approximately, since a larger period could lead to fatigue and attention drift. For this reason we chose the ranking evaluation protocol, which is faster than SDS and SS. In this case, for each scene, all the images under comparison must be presented to the participant simultaneously in the same display, with enough resolution so that their quality can be assessed easily. This limits the number of images that can be used in the test and therefore the number of enhancement methods to be compared. We decided to display 8 images per scene, one of them the original (degraded) image, which implies that only 7 processed results could be used in the comparison. The selected methods have been described in Section II. The initial tests showed that the participants needed, on average, about 4 minutes to rank the 8 images, which implied that no more than 15 scenes could be used in the experiment to fulfill the 1 hour time constraint. Finally, a group of 30 participants were chosen to perform the test, which is within the usual range used in other studies (see Table 1) and is a number large enough so the results are statistically meaningful.

A. DESCRIPTION OF THE EXPERIMENT

A custom GUI application (Figure 1) was specifically built for the experiment. It presents eight images corresponding to the same scene: seven results from different UIE algorithms, and the original (un-processed) image. The images are shuffled and un-labeled, so that the user ignores how they have been generated. To help in the visualization of the images, they can be enlarged by clicking on them.

Fifteen varied underwater scenes were used in the experiment, selected by marine sciences experts, which represent different habitats and some typical drawbacks of underwater images (i.e. blur, haze, low saturation, low visibility, noise, and deep water scenarios). One or more fish appear in each scene. This point is important, since the participants are required to decide in which of the displayed images these fish can be better observed. The selected scenes are shown in Figure 2.

Regarding the enhancement algorithms, we used our own versions of UCM, Fusion, InfoLoss and UDCP, which were implemented following the description in the original papers. For MSR, we used the implementation in [45], while for ARC and BAL we obtained the processed results through the Platform for Underwater Image Quality Evaluation (PUIQE) [33]. For each method, the default parameters recommended by the authors were used.

From the thirty participants, 19 were experts in fish ecology and the rest in image enhancement methods. They were asked whether they have a normal vision (or corrected to normal vision) during the recruiting process. A confidentiality sheet was signed to ensure that no personal information would be released.

The participants were introduced to the goal of the experiment before the start, and a brief training session followed, using a scene which was subsequently discarded from the results. The text of the guidelines given to the participants was: “The user will be presented with several images corresponding to different underwater scenes. Then, he/she must assign a rank to the images, from 1 to 8, where higher ranks are given to images in which is easier to see more fish and discern their contours.” Participants were asked to enlarge each image before deciding the final ranking. The ranked images were shown on the right panel of the display and the user could modify the ranking by dragging them to its corresponding position (labeled from 1 to 8, with 1 being the best and 8 being the worst). After sorting all the images for a given scene the user clicked the submit button and a new scene was presented. The scenes were displayed in a random order.

The experiments were conducted in a dark room with minimal ambient lighting (below 25 lux) which is within the recommended luminance levels according to ITU-R recommendations [8]. The distance between the display and the participant was set to approximately 80 cm.

IV. ANALYSIS OF THE RESULTS

This section summarizes the results of the subjective experiment and performs a statistical analysis of them.

Table 2 shows, for each scene, the average rank assigned to each method, computed over all the participants in the experiment. The methods are sorted according to their average rank, from the highest rank (low value, i.e. best) to the lowest (high value, i.e. worst). We observe that, in some cases, several
TABLE 1. Subjective tests with general images (I), underwater images (UI) and videos (V). For each test we list the number of images, the number of methods under evaluation, and the number of participants. The used methods have been classified in Single Stimulus (SS), Simultaneous double stimulus (SDS), or Ranking (R), where the number is the amount of the images in the ranking. Finally, the proposed task or question is indicated.

| Reference | Images | Number of Methods | Part. | Method | Task / Question |
|-----------|--------|-------------------|-------|--------|-----------------|
| [7] Aydin et al. | 20        | 1                  | 21    | SS     | Rating a single aesthetic attribute |
| [29] Koho et al. | -         | 1                  | -     | SS     | Rating from 1-5 the microscopy image |
| [38] Ma et al. | 69-102    | 10                 | 30-34 | SDS    | Rating from 1-5 the quality of target image |
| [52] Shi et al. | 176       | 8                  | 33    | SDS    | Rating from 1-5 the quality of segmented image |
| [10] Duan et al. | 15        | 5                  | 24    | R-5    | Which image is the most natural looking? |
| [31] Li et al. | 950       | 12                 | 50    | SDS    | Which one is better than the other? |
| [46] Protasiuk et al. | 50        | 9                  | 10    | R-4    | Ranking of least to most realistic (1-4) |
| [57] Yang and Sowmia | 44        | 1                  | 12    | SS     | Rating from 1-5 the quality |
| [22] Hou et al. | 30        | 10                 | 15    | SS     | Rating from 1-10 the quality |
| [12] Emberton et al. | 10        | 4                  | 10    | SDS    | Which one you prefer? (Two images) |
| [20] Hou et al. | 50        | 7                  | 20    | R-7    | Ranking 7 images |
| [13] Emberton et al. | 100       | 5                  | 260   | SDS    | Which one you prefer? (Two images) |
| [42] Nahtinen et al. | 40        | 1                  | 30    | SS     | Artifacts detection |
| [41] Mukherjee et al. | 6         | 6                  | 32    | R-7    | Ranking by resemblance to reference |

FIGURE 1. Image of the web-based GUI application developed for the subjective experiment.

methods get similar average ranks (for example, Infoloss = 3.03, MSR = 3.06, and BAL = 3.12, in scene 8), and the question arises whether these small differences are really meaningful or all these methods should be considered to have the same rank. To address this issue we perform a statistical analysis of the results using Kendall’s coefficient of concordance.

We follow [41] and [10] to analyze the level of agreement between the rankings assigned by the participants in the subjective test. Kendall [27] proposed to compute the agreement between different rankings of a set of objects using the coefficient of concordance \( W \)

\[
W = \frac{12S}{m^2(n^3 - n)}
\]

where \( m \) is the number of participants (judges or raters) and \( n \) is the number of enhancement methods (objects) to be evaluated. \( S \) measures the differences between the rankings assigned by each participant. Let \( r_{ij} \) be the rank given by participant \( j \) to algorithm \( i \), then

\[
S = \sum_{i=1}^{n} (R_i - \bar{R})^2,
\]

where \( R_i \) is the sum of the ranks assigned to method \( i \) (\( R_i = \sum_{j=1}^{n} r_{ij} \)), and \( \bar{R} \) is the average of all \( R_i \) (\( \bar{R} = \frac{1}{n} \sum_{i=1}^{n} R_i \)).

\( W \) takes values in [0, 1]. Total agreement between participants implies \( W = 1 \), while complete disagreement (random rankings) entails \( W = 0 \). A significant test for \( W \) was proposed in [28], in which the null hypothesis \( H_0 \) was the lack of agreement. The test statistic is computed as

\[
\chi^2 = m(n - 1)W,
\]

and it follows a chi-squared distribution with \( n - 1 \) degrees of freedom. The null hypothesis is rejected when the p-value of the test is below 0.05. Remark that acceptance of the null
hypothesis implies that all the methods under test receive random rankings, and therefore none of them can be considered better than the rest.

We want to decide whether there are groups of algorithms for which the ranks are not significantly different (that is, groups for which $H_0$ is accepted). For this reason we perform a multiple test that involves all possible groupings of algorithms. The significance of the tests is adjusted using the Bonferroni correction.

Table 2 shows the results of the tests for each one of the scenes of the experiment. The methods are sorted according to their average rank $\bar{R}_i = \frac{R_i}{m}$. For each group of 8, 7, 6, \ldots, 2 consecutive methods the significance of their coefficient of concordance $W$ is computed, and those groups for which $H_0$ is accepted are encircled with the same color in the table.

A global analysis of the enhancement methods is performed by computing, for each method $i$, the average of its $\bar{R}_i$ over the 15 scenes. The methods are then sorted according
Table 2. Enhancement methods sorted by their average ranks, for each scene. Groups of similarly ranked methods are encircled with the same color.

| Scene | Enhancement Methods with Average Rankings |
|-------|------------------------------------------|
| 1     | Fusion (2.81) | BAL (3.44) | ARC (3.50) | Original (3.56) | UDCP (5.03) | InfoLoss (5.28) | UCM (5.56) | MSR (6.81) |
| 2     | MSR (1.78) | ARC (3.00) | Fusion (4.16) | UDCP (5.09) | BAL (5.12) | Original (5.47) | UCM (5.69) | InfoLoss (5.69) |
| 3     | Fusion (1.81) | BAL (3.84) | MSR (4.50) | ARC (4.56) | InfoLoss (4.81) | Original (4.84) | UCM (4.84) | UDCP (6.44) |
| 4     | MSR (1.94) | InfoLoss (2.81) | Fusion (2.62) | UCM (3.62) | BAL (5.34) | ARC (6.12) | Original (6.78) | UCM (6.28) |
| 5     | MSR (1.94) | InfoLoss (2.81) | Fusion (2.62) | UCM (3.62) | BAL (5.34) | ARC (6.12) | Original (6.78) | UCM (6.28) |
| 6     | MSR (1.94) | UDCP (3.25) | ARC (3.44) | ARC (3.43) | UCM (5.06) | Fusion (5.25) | MSR (5.28) | UCM (6.90) |
| 7     | MSR (1.75) | Fusion (3.12) | UDCP (3.72) | ARC (3.72) | Original (4.62) | Fusion (5.34) | MSR (5.72) | UCM (6.37) |
| 8     | MSR (1.53) | Fusion (3.06) | UCM (3.12) | ARC (3.25) | Original (5.87) | Fusion (5.97) | MSR (6.25) | UCM (6.25) |
| 9     | UDCP (2.53) | InfoLoss (2.91) | Fusion (4.44) | Original (3.53) | ARC (4.41) | Fusion (4.68) | MSR (5.19) | BAL (5.22) |
| 10    | MSR (1.69) | UDCP (2.91) | InfoLoss (3.73) | Fusion (4.56) | Original (3.53) | ARC (4.41) | Fusion (4.68) | MSR (5.19) | BAL (5.22) |
| 11    | UDCP (2.53) | InfoLoss (2.91) | Fusion (4.44) | Original (3.53) | ARC (4.41) | Fusion (4.68) | MSR (5.19) | BAL (5.22) |
| 12    | MSR (2.16) | Fusion (2.37) | UDCP (3.72) | fusion (3.72) | Original (3.00) | UCM (4.51) | Fusion (5.47) | UCM (5.47) | UDCP (6.94) |
| 13    | InfoLoss (2.37) | MSR (3.16) | Fusion (3.14) | UDCP (3.56) | UCM (4.69) | BAL (5.75) | BAL (5.75) | UCM (5.75) | UDCP (7.94) |
| 14    | MSR (2.25) | InfoLoss (3.28) | Fusion (3.62) | UDCP (3.78) | UCM (3.91) | ARC (3.31) | UCM (6.46) | UCM (6.46) | UCM (6.46) |
| 15    | MSR (1.03) | InfoLoss (3.12) | Fusion (3.97) | UDCP (4.56) | UCM (3.34) | Fusion (6.22) | UCM (6.22) | Original (6.07) |

Table 3. Enhancement methods sorted by their global average ranks. Groups of similarly ranked methods are encircled with the same color.

| Enhancement Methods with Global Average Rankings |
|-----------------------------------------------|
| MSR (2.73) | Fusion (3.51) | InfoLoss (4.09) | BAL (4.44) | ARC (4.97) | UDCP (5.05) | UCM (5.13) | Original (6.07) |

The results in Table 4 permit to conclude that, in general, MSR is the method with the highest ranking. In fact, it gets the highest rank in most of the cases, except for scenes 1, 3, 10 and 13. Fusion is the second method in the global ranking, but only ranks first in 4 out of the 15 scenes. The rest of methods are rarely chosen as the first option by the observers and get worse global rankings than MSR or Fusion. Finally, the original image is never preferred to some of the processed results, which endorses the need of the enhancement.

Figures 3 and 4 display the results of the different methods when applied to two of the chosen scenes. MSR removes the color cast and strongly increases the contrast of the images. This process doesn’t always lead to the best visual quality (e.g. Fusion, BAL or InfoLoss in Figure 3 produce better looking results), but it permits to discern better the contours of the objects, which facilitates the annotation task. ARC produces reddish colors in the brightest parts of the images and UDCP creates strong halo artifacts, which explains why these methods use to get low ranks in the experiment. In Figure 4 we observe that Fusion, BAL and UCM are unable to remove the haze present in the original image. In general, these methods get low rankings when the input images are hazy.

Figure 5 displays some excerpts of scenes 1, 3 and 10, for which MSR gets very low ranks. We can see that scene 10...
TABLE 4. Effective rankings of each method per scene, according to the groupings shown in Table 2. The last column corresponds to the global groupings in Table 3.

| Method  | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | 11  | 12  | 13  | 14  | 15  | Global |
|---------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--------|
| MSR     | 3   | 1   | 2.5 | 1   | 1   | 1   | 1   | 1   | 1   | 3   | 1   | 1   | 1.5 | 1   | 1.5   | 1.5   |
| Fusion  | 1   | 2.5 | 1   | 3.5 | 3.5 | 2   | 1   | 1.5 | 2   | 3   | 1   | 1.5 | 1.5 | 1.5 | 1   | 1.5   |
| InfoLoss| 2.5 | 3   | 2.5 | 1   | 2   | 3   | 1   | 3   | 1.5 | 2   | 1   | 2   | 2.5 | 1   | 2.5   | 2.5   |
| BAL     | 1.5 | 2.5 | 3   | 5   | 3.5 | 2   | 1   | 2   | 3   | 5   | 1   | 3   | 1.5 | 3   | 3     | 3     |
| ARC     | 1.5 | 2   | 2.5 | 4   | 4   | 1   | 2   | 3   | 5   | 1   | 3   | 1.5 | 3   | 4   | 3.5   | 4     |
| UDCP    | 2   | 3   | 3   | 5   | 3   | 4   | 1   | 4   | 1   | 3   | 1.5 | 4.5 | 4   | 4   | 4     | 4     |
| UCM     | 2.5 | 3   | 3   | 4   | 4.5 | 2   | 2   | 4   | 4   | 2   | 2.5 | 2.5 | 4   | 4   | 4     | 4     |
| Original| 1.5 | 3   | 2.5 | 4   | 5   | 5   | 3   | 3   | 2   | 3.5 | 2   | 4   | 5   | 5   | 5     | 5     |

FIGURE 3. Scene 8 with the results of the different enhancement methods. The ranks and UCIQE values associated to each image are also displayed. The two best values of UCIQE are highlighted in red and green, respectively. See text for details.

is very noisy and MSR excessively enhances the noise. The same is true for all the other methods. Even for the best ranked method (UDCP) it is hard to distinguish the contours of the fish. In scenes 1 and 3 MSR produces color artifacts, in the form of red dots near the contours of the objects. For these scenes, the red component of the light is very low w.r.t. the blue and green components (R/G and R/B ratios are very small), and the algorithm tends to overenhance this component, giving raise to the observed artifacts. In these cases the Fusion method produces better results than MSR (see second and third rows in Figure 5).

A. QUANTITATIVE EVALUATION

To complete the assessment of the UIE methods, we compute in this section three popular underwater-specific no-reference image quality metrics, namely, UCIQE [57], UIQM [43]
and CCF [54]. UCIQE is a linear combination of chroma, saturation, and contrast, measured in the CIELab color space. UIQM is a combination of three attributes: colorfulness, sharpness, and contrast. Finally, CFM is a combination of three attributes: colorfulness, contrast, and fog density. For all the metrics, the higher the value the better the image quality, although it is well known [21] that the subjective perception of quality doesn’t always agree with the obtained values.

The three metrics have been computed for all the images used in our experiments. Our goal is to check whether the values of these metrics are correlated with the results of our subjective experiment. Table 5 shows the average values, over the 15 scenes, for the different enhancement methods. The last row displays the Spearman’s rank correlation between the effective global ranks of each method (last column in Table 4) and the metrics’ values. We observe a relatively high correlation between UCIQE values and effective ranks. In Figures 3 and 4 the ranks and UCIQE values for the original and processed results computed for two scenes are displayed. It must be remarked that MSR, which ranks first in both scenes doesn’t get the highest UCIQE value. This is an expected result, since MSR was not preferred by the observers due the visual quality of its results but for the enhancement of the image contours. On the other hand, high values of UCIQE do not always imply that the images are good for the annotation task (see Figure 4).

**B. UNDERWATER IMAGE ENHANCEMENT AND DEEP LEARNING**

We introduce in this section a new element for the analysis of the different image processing methods described in the paper. So far, we have concluded that MSR is the method that better helps humans to detect and label fish. A natural question arises: is the same statement true for a machine? In other words, would the object detection task performed by a deep learning algorithm be improved by the use of these processing techniques?

In order to answer this question we have trained a well-known deep learning model for object detection, Mask RCNN [18], using as input the images processed with the 7 methods described in Section II, and also the original, un-processed versions of the images. The network has been trained to detect one single category (fish), using 400 images (with 4252 annotated fish). For testing, a set of 100 images (1004 fish) has been used. Figure 6 displays one of the images used for training, with superimposed annotations in color.

All the images (train+test) have been processed using the methods described in the paper. Each pair of train+test sets have been used to train and evaluate the performance of the fish detector. The Matterport implementation of Mask RCNN [2] has been used in our tests. The upper layers of the network (‘heads’) have been trained for 30 epochs and
a learning rate of 0.001; then the intermediate and higher layers (‘fc3’) have been trained for 30 additional epochs with a learning rate of 0.0005; and finally, all the layers have been trained for other 30 epochs with the same learning rate.

The detection accuracy of the trained network has been evaluated in terms of the mean Average Precision (mAP). In order to reduce the effect of the random nature of the network optimization process, the above strategy has been
repeated five times, for each pair of train+test processed images. Table 6 displays the average mAP values obtained on the test sets, for each enhancement method. The last row displays the Spearman’s rank correlation between the effective global ranks of each method (last column in Table 4) and the mAP values. Although there exists a positive correlation between rank and mAP values, it is not very strong. However, it can be remarked that when using MSR the detection results are significantly better than with the other methods. Previous studies [34] have already suggested that the use of image enhancement techniques may improve the performance of deep learning models, and, in the future, we plan to dive further in the subject.

### V. CONCLUSION

In this paper we have analyzed seven underwater image enhancement algorithms representative of the state of the art. We have studied them in terms of how well they serve the task of image annotation for deep learning purposes. Since this task is performed by humans we have carried out a subjective experimental test in which 30 observers have ranked the results of these algorithms on 15 underwater scenes. We have made a statistical analysis of the results and checked whether they correlate with some popular quantitative metrics. Our conclusion is that the Multiscale Retinex algorithm is, in general, preferred over the rest of methods, since it enhances significantly the contours of the objects in the scene. This enhancement may be excessive when the image is noisy and may produce color artifacts when the red component of the light is very low w.r.t. the blue and green components, but overall, the results are good. Moreover, some tests also suggest that this method can improve the performance of a CNN object detector when used for pre-processing the images. This will be further investigated in our future research and, besides Mask RCNN, other popular architectures (such as YOLOv6 or RetinaNet) will be examined.

### REFERENCES

[1] J. Raihan, A. P. E. Abas, and L. C. D. Silva, “Review of underwater image restoration algorithms,” *IEEE Image Process.*, vol. 13, no. 10, pp. 1587–1596, Aug. 2019.

[2] W. Abdulla. (2017). *Mask R-CNN for Object Detection and Instance Segmentation on Keras and Tensorflow*. [Online]. Available: https://github.com/matterport/Mask_RCNN

[3] M. Aljabri, M. Alamir, M. AlGhandi, M. Abdel-Mottaleb, and F. Collado-Mesa, “Towards a better understanding of annotation tools for medical imaging: A survey,” *Multimedia Tools Appl.*, vol. 81, no. 18, pp. 25877–25911, Jul. 2022, doi: 10.1007/s11042-022-12100-1.

[4] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and P. Bekaert, “Color balance and fusion for underwater image enhancement,” *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 379–393, Jan. 2018.

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

[6] S. Anwar, C. Li, and F. Porikli, “Deep underwater image enhancement,” Tech. Rep., 2018.

[7] T. O. Aydin, A. Smolic, and M. Gross, “Automated aesthetic analysis of photographic images,” *IEEE Trans. Vis. Comput. Graphics*, vol. 21, no. 1, pp. 31–42, Jan. 2015.

[8] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, BT, R.I.R, International Telecommunication Union, Geneva, Switzerland, 2002.

[9] P. Drews, Jr., E. do Nascimento, F. Moraes, S. Botelho, and M. Campos, “Transmission estimation in underwater single images,” in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Dec. 2013, pp. 825–830, doi: 10.1109/ICCVW.2013.113.

[10] L. Duan, K. Debattista, Z. Lei, and A. Chalmers, “Subjective and objective evaluation of local dimming algorithms for HDR images,” *IEEE Access*, vol. 8, pp. 51692–51702, 2020.

[11] A. Álvarez-Ellacuria, M. Palmer, I. A. Catalán, and J.-L. Lisani, “Image-based, unsupervised estimation of fish size from commercial landings using deep learning,” *ICES J. Mar. Sci.*, vol. 77, no. 4, pp. 1330–1339, Jul. 2020.

[12] S. Emberton, L. Chittka, and A. Cavallaro, “Hierarchical rank-based visual light estimation for underwater dehazing,” in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, X. Xie, M. W. Jones, and G. K. L. Tam, Eds. BMVA Press, Sep. 2015, pp. 125–125.12, Art. no. 125, doi: 10.5244/C.29.125.

[13] S. Emberton, L. Chittka, and A. Cavallaro, “Underwater image and video dehazing with pure hae region segmentation,” *Comput. Vis. Image Understanding*, vol. 168, pp. 145–156, Mar. 2019.

[14] C. Fabbi, M. I. Islam, and I. Sattar, “Enhancing underwater imagery using generative adversarial networks,” in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 7159–7165, doi: 10.1109/ICRA.2018.8460552.

[15] A. Galdran, D. Pardo, A. Picón, and A. Alvarez-Gilá, “Automatic red-channel underwater image restoration,” *J. Vis. Commun. Image Represent.*, vol. 26, pp. 132–145, Jan. 2015.

[16] C. Guo, C. Li, J. Guo, R. Cong, H. Fu, and P. Han, “Hierarchical features driven residual learning for depth map super-resolution,” *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2545–2557, May 2019, doi: 10.1109/TIP.2018.2887029.

[17] M. Han, Z. Lyu, T. Qiu, and M. Xu, “A review on intelligence dehazing and color restoration for underwater images,” *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 50, no. 5, pp. 1820–1832, May 2020, doi: 10.1109/TSMC.2017.2789902.

[18] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2980–2988.

[19] K. He, J. Sun, and X. Tang, “Single image haze removal using dark channel prior,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2414–2435, Dec. 2011, doi: 10.1109/TPAMI.2010.168.

[20] G. Hou, J. Li, G. Wang, H. Yang, B. Huang, and Z. Pan, “A novel dark channel prior guided variational framework for underwater image restoration,” *J. Vis. Commun. Image Represent.*, vol. 66, Jan. 2020, Art. no. 102732.

[21] G. Hou, Y. Li, H. Yang, K. Li, and Z. Pan, “UID2021: An underwater image dataset for evaluation of no-reference quality assessment metrics,” arXiv:2204.08813.

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

[23] K. Iqbal, M. Odetayo, A. James, R. A. Salam, and A. Z. H. Talib, “Enhancing the low quality images using unsupervised colour correction method,” in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, Oct. 2017, pp. 1703–1709.

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

[25] R. Liu, X. Fan, M. Zhu, M. Hou, and Z. Luo, “Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 12, pp. 4861–4875, Dec. 2020, doi: 10.1109/TCVT.2019.2963772.

### TABLE 6. Mean average Precision (mAP) values for a Mask RCNN fish detector trained and tested using images processed with the different methods described in the paper.

| Method      | mAP   |
|-------------|-------|
| MSR         | 0.6642|
| Fusion      | 0.6338|
| InfoLoss    | 0.6366|
| BAL         | 0.5744|
| ARC         | 0.6146|
| UDCP        | 0.6192|
| UCM         | 0.5966|
| Original    | 0.6144|
| Spearman’s rank correlation | 0.6587 |
H. Lu, Y. Li, Y. Zhang, M. Chen, S. Serikawa, and H. Kim, “Underwater R. Protasiuk, A. Bibi, and B. Ghanem, “Local color mapping combined A. B. Petro, C. Sbert, and J.-M. Morel, “Multiscale retinex,” Image K. Panetta, C. Gao, and S. Agaian, “Human-visual-system-inspired under- R. Mukherjee, K. Debattista, T. Bashford-Rogers, P. Vangorp, R. Mantiuk, T. Mertens, J. Kautz, and F. Van Reeth, “Exposure fusion: A simple and B. L. McGlamery, “A computer model for underwater camera systems,” T. M. Wa, L. Deng, C. Dong, and K. N. Ngan, “Image retargeting quality J. Lisani et al.: Analysis of Underwater Image Processing Methods for Annotation in DL Based Fish Detection

H. Land and J. J. McCann, “Lightness and retinex theory,” J. Opt. Soc. D. Jobson, Z. Rahman, and G. A. Woodell, “A multiscale retinex
J.-L. Lisani et al., “Classical and state-of-the-art approaches for underwater image defogging: A comprehensive survey,” Frontiers Inf. Technol. Electron. Eng., vol. 21, no. 12, pp. 1745–1769, Dec. 2020, doi: 10.1631/FITEE.2000190.

Y. Zhang, W. Zuo, S. Gu, and L. Zhang, “Learning deep CNN denoiser prior for image restoration,” in Proc. IEEE Conf. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 3929–3938, doi: 10.1109/CVPR.2017.300.

X. Zhao, T. Jin, and S. Qu, “Deriving inherent optical properties from background color and underwater image enhancement,” Ocean Eng., vol. 94, pp. 163–172, Jan. 2015, doi: 10.1016/j.oceaneng.2014.11.036.

J.-C. Zhou, D.-H. Zhang, and W.-S. Zhang, “Classic and state-of-the-art approaches for underwater image defogging: A comprehensive survey,” Mobile Netw. Appl., vol. 20, no. 3, pp. 726, Jan. 2020.

A. Petro, C. Sbert, and J.-M. Morel, “Multiscale retinex,” Image Processing On Line, vol. 4, no. 1, pp. 71–88, Apr. 2014.

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

Y. Wang, W. Song, G. Fortino, L.-Z. Qi, W. Zhang, and A. Liotta, “An experimental-based review of image enhancement and image restoration methods for underwater imaging,” IEEE Access, vol. 7, pp. 140233–140251, 2019, doi: 10.1109/ACCESS.2019.2932130.

M. Yang and A. Sowmya, “An underwater color image quality evaluation metric,” IEEE Trans. Image Process., vol. 24, no. 12, pp. 6062–6071, Dec. 2015.

K. Zhang, W. Zuo, S. Gu, and L. Zhang, “Learning deep CNN denoiser prior for image restoration,” in Proc. IEEE Conf. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 3929–3938, doi: 10.1109/CVPR.2017.300.

W. Zhang, L. Dong, X. Pan, P. Zou, L. Qin, and W. Xu, “A survey of restoration and enhancement for underwater images,” IEEE Access, vol. 7, pp. 182259–182279, 2019, doi: 10.1109/ACCESS.2019.2959560.

X. Zhao, T. Jin, and S. Qu, “Deriving inherent optical properties from background color and underwater image enhancement,” Ocean Eng., vol. 94, pp. 163–172, Jan. 2015, doi: 10.1016/j.oceaneng.2014.11.036.

J.-C. Zhou, D.-H. Zhang, and W.-S. Zhang, “Classic and state-of-the-art approaches for underwater image defogging: A comprehensive survey,” Frontiers Inf. Technol. Electron. Eng., vol. 21, no. 12, pp. 1745–1769, Dec. 2020, doi: 10.1631/FITEE.2000190.

J.-C. Zhou, D.-H. Zhang, and W.-S. Zhang, “Classic and state-of-the-art approaches for underwater image defogging: A comprehensive survey,” Frontiers Inf. Technol. Electron. Eng., vol. 21, no. 12, pp. 1745–1769, Dec. 2020, doi: 10.1631/FITEE.2000190.

J. Lisani, A. Petro, A. Álvarez-Ellacuría, I. Catalán, and M. Palmer, “On the importance of color pre-processing for object detection in submarine images,” in Proc. Int. Conf. Catalan Assoc. Artif. Intell., 2022, pp. 239–242.

R. Liu, X. Fan, M. Zhu, M. Hou, and Z. Luo, “Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light,” IEEE Trans. Circuits Syst. Video Technol., vol. 30, no. 12, pp. 4861–4875, Dec. 2020, doi: 10.1109/TCSVT.2019.2963772.

V. Lopez-Vazquez, J. M. Lopez-Guade, S. Marin, E. Fanelli, E. Johnsen, and J. Aguzzi, “Video image enhancement and machine learning pipeline for underwater animal detection and classification at cabled observatories,” Sensors, vol. 20, no. 3, p. 726, Jan. 2020.

H. Lu, Y. Li, Y. Zhang, M. Chen, S. Serikawa, and H. Kim, “Underwater optical image processing: A comprehensive review,” Mobile Netw. Appl., vol. 22, no. 6, pp. 1204–1211, Dec. 2017.

L. Ma, W. Lin, C. Dong, and K. N. Ngan, “Image retargeting quality assessment: A study of subjective scores and objective metrics,” IEEE J. Sel. Topics Signal Process., vol. 6, no. 6, pp. 626–639, Oct. 2012.

B. L. McClamery, “A computer model for underwater camera systems,” Proc. SPIE, vol. 208, pp. 221–231, Mar. 1980.

T. Mertens, J. Kautz, and F. Van Reeth, “Exposure fusion: A simple and practical alternative to high dynamic range photography,” Comput. Graph. Forum, vol. 28, no. 1, pp. 161–171, Mar. 2009.

R. Mukherjee, K. Debattista, T. Bashford-Rogers, P. Vangorp, R. Mantuik, M. Bessa, B. Waterfield, and A. Chalmers, “Objective and subjective evaluation of high dynamic range video compression,” Signal Process., Image Commun., vol. 47, pp. 426–437, Sep. 2016.

M. Nuantine, T. Viritanen, M. Vahteranoksa, T. Vuori, P. Oittinen, and J. Hakkinen, “CV2D—A database for evaluating no-reference video quality assessment algorithms,” IEEE Trans. Image Process., vol. 25, no. 7, pp. 3073–3086, Jul. 2016.

K. Panetta, C. Gao, and S. Agaian, “Human–visual–system–inspired underwater image quality measures,” IEEE J. Ocean. Eng., vol. 41, no. 3, pp. 541–551, Jul. 2016.

Y.-T. Peng and P. C. Cosman, “Underwater image restoration based on image bluriness and light absorption,” IEEE Trans. Image Process., vol. 26, no. 4, pp. 1579–1594, Apr. 2017.

A. B. Petro, C. Sbert, and J.-M. Morel, “Multiscale retinex,” Image Processing Line, vol. 4, pp. 71–88, Apr. 2014.

M. Protasiev, A. Bibi, and G. Ghanem, “Local color mapping combined with color transfer for underwater image enhancement,” in Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV), Jan. 2019, pp. 1433–1439.

H. Qin, X. Li, J. Liang, Y. Peng, and C. Zhang, “DeepFish: Accurate underwater live fish recognition with a deep architecture,” Neurocomputing, vol. 187, pp. 49–58, Apr. 2016.

J. Komron, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 779–788.

S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” in Advances in Neural Information Processing Systems, vol. 28, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, Eds., Curran Associates, 2015, [Online Available: https://proceedings.neurips.cc/paper/2015/file/14a6aabbb1485c45bb82298be2de38046-Paper.pdf].
CATALINA SBERT received the degree in mathematics, in 1987, and the Ph.D. degree in computer science, in 1995. She is currently an Associate Professor with the University of the Balearic Islands. She has coauthored over 20 papers indexed in JCR and over 25 conference papers. Her research interests include variational models and partial differential equations applied to image restoration and image enhancement.

AMAYA ÁLVAREZ-ELLACURÍA received the Ph.D. degree in marine science from the University of the Balearic Islands, in 2010. From 2010 to 2017, she was hired as a Technician at the Balearic Islands Coastal Observing and Forecasting System, SOCIB, Spain. Since 2018, she has been working as a Technician with the Mediterranean Institute for Advanced Studies (IMEDEA), a joint center between the Spanish National Research Council (CSIC) and the University of the Balearic Islands (UIB). Formerly, she was the Vice-Director of IMDEA, where she is currently the Head of the Department of Marine Ecology. He has authored over 75 SCI papers on marine ecology, particularly fisheries oceanography, and has led 15 (three EU) research projects. He has supervised five Ph.D. and several postdoctoral students (including two Marie Curie grantees).

IGNACIO A. CATALÁN received the M.Sc. degree in fisheries and shellfish culture from the School of Ocean Sciences, Bangor, U.K., in 1998, and the Ph.D. degree in biology from the University of Barcelona, in 2003. He conducted several postdoctoral contracts and stages at UiB Bergen (Norway) and CSIC (Spain), and since 2009, he has been working as a Tenured Scientist with the Mediterranean Institute for Advanced Studies (IMEDEA), a joint center between the Spanish National Research Council (CSIC) and the University of the Balearic Islands (UIB). Formerly, he was the Vice-Director of IMDEA, where he is currently the Head of the Department of Marine Ecology. He has authored over 75 SCI papers on marine ecology, particularly fisheries oceanography, and has led 15 (three EU) research projects. He has supervised five Ph.D. and several postdoctoral students (including two Marie Curie grantees).

MIQUEL PALMER received the Ph.D. degree in biology from the University of Illes Balears (UIB), in 1994. Since 2006, he has been working as a Tenured Scientist with the Mediterranean Institute for Advanced Studies (IMEDEA), a joint center between the Spanish National Research Council (CSIC) and UIB. He has authored over 120 SCI papers. He has supervised seven (plus three ongoing) Ph.D. projects. His current research interests include quantitative ecology of fish, especially on individual variation. He is investing a growing effort to develop new technological and analytical tools based on underwater cameras, since he believes that a paradigm shift in the conventional methods of marine organisms observation is needed for affording the current challenges.