An Underwater Image Quality Assessment Metric

Pengfei Guo, Hantao Liu, Delu Zeng, Tao Xiang, *Senior Member, IEEE*, Leida Li, *Member, IEEE*, and Ke Gu, *Member, IEEE*

***Abstract*—Various image enhancement algorithms are adopted to improve underwater images that often suffer from visual distortions. It is critical to assess the output quality of underwater images undergoing enhancement algorithms, and use the results to optimise underwater imaging systems. In our previous study, we created a benchmark for quality assessment of underwater image enhancement via subjective experiments. Building on the benchmark, this paper proposes a new objective metric that can automatically assess the output quality of image enhancement, namely UWEQM. By characterising specific underwater physics and relevant properties of the human visual system, image quality attributes are computed and combined to yield an overall metric. Experimental results show that the proposed UWEQM metric yields good performance in predicting image quality as perceived by human subjects.**

***Index Terms*—Underwater, image quality, enhancement, human visual system, objective metric.**

I. Introduction

THE acquisition of high quality images plays an important role in marine industry [1]–[5] and other maritime activities [6]–[11]. However, the complexity of underwater environment poses many challenges to image acquisition. First, water absorbs different wavelengths of light at different degrees [12]; and underwater images often suffer from greenish or bluish color cast [13]. Second, object edges and details are often blurred due to the scattering of light [14]. Third, the backward scattering of light by particles can cause the effect of low contrast [15]. Therefore, many algorithms have been developed to improve the quality of underwater images [14], [16]–[22].

Manuscript received 7 July 2021; revised 30 March 2022; accepted 20 June 2022. Date of publication 30 June 2022; date of current version 30 October 2023. This work was supported in part by Guangdong basic and applied basic research foundation under Grant 2020A1515110958 and the Fundamental Research Program 401 of Guangdong, China under Grant 2020B1515310023. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Joao M Ascenso. *(Corresponding author: Hantao Liu.)*

Pengfei Guo is with the School of Computational Science, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China, and also with the School of Mathematics, South China University of Technology, Guangzhou 510641, China (e-mail: guopfzhku@163.com).

Hantao Liu is with the School of Computer Science and Informatics, Cardiff University, CF24 3AA Cardiff, United Kingdom (e-mail: liuh35@cardiff.ac.uk).

Delu Zeng is with the School of Mathematics, South China University of Technology, Guangzhou 510641, China (e-mail: dltsang@xmu.edu.cn).

Tao Xiang is with the College of Computer Science, Chongqing University, Chongqing 400044, China (e-mail: txiang@cqu.edu.cn).

Leida Li is with the School of Artificial Intelligence, Xidian University, Xi'an 710071, China (e-mail: ldli@xidian.edu.cn).

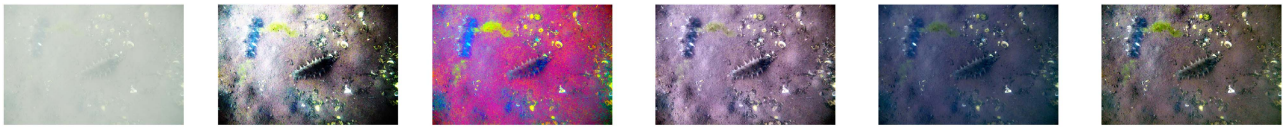
Ke Gu is with the Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China (e-mail: guke.doctor@gmail.com).

Digital Object Identifier 10.1109/TMM.2022.3187212

The underwater image enhancement algorithms can be categorized into image formation model-based (IFM-based) and image formation model-free (IFM-free) approaches [23]. The IFM-based methods improve the quality of underwater images by estimating the optical properties of underwater imaging and recovering color, sharpness and contrast [14], [17], [18], [22], [24]; and the IFM-free methods attempt to correct the color and details of underwater images by re-distributing their intensity values [16], [19], [20]. Assessing the output image quality of these enhancement algorithms remains an academic challenge [25], [26]. The first challenge lies in the lack of subjective study on the perceived quality of underwater images undergoing enhancement algorithms or processes. The second challenge is to build a reliable algorithm to predict image quality in close agreement with human judgements. In our previous study [27], a benchmark was created for quality assessment of underwater images undergoing various enhancement algorithms. The ground truth of perceived image quality revealed the way human viewers judge the output image quality of these enhancement algorithms. When properly conducted, subjective evaluation is the most reliable means of measuring perceived image quality; however, it is cumbersome and impractical in many circumstances. A more realistic way to measure image quality is to develop an objective metric, which can automatically assess perceived quality.

Image quality assessment (IQA) algorithms can be divided into full-reference (FR), reduced-reference (RR), and no-reference (NR) approaches [28]. The FR IQA algorithms use the whole information of the reference image to evaluate the quality of the test image [29], [30]. The RR IQA methods use partial information of the reference image to assess the quality of the test image [31], [32]. NR IQA methods rely on the test image only and extract the IQA features for the measure of image quality [33]–[38].

It should be noted that in the context of image enhancement, a pristine reference image of "perfect" quality does not exist, therefore, the objective IQA metrics must be available in the no-reference (NR) framework. In the literature, developing an NR metric has been the most challenging task in the image quality research community, mainly due to our lack of knowledge on the human visual system (HVS) [39]–[41]. Some successful NR metrics have been developed and proven effective for specific applications with specific visual distortions [33]–[35]. These metrics, however, are generally challenged when dealing with the complexity and diversity in natural content. For underwater images, it is known that the image statistics, due to e.g., light attenuation, scattering and reflection, are different from the ordinary natural scene statistics [27]. This makes the problem



(a) (b) (c) (d) (e) (f)

Fig. 1. Illustration of an original underwater image (a) and results (b)-(f) of five image enhancement algorithms, HE, DS, IF, UDCP and NLD in the UEIQA database [27]. (a) represents the original image with MOS (mean opinion score) = 1.6161, (b) represents the result of HE with MOS = 4.3007, (c) represents the result of DS with MOS = 2.2353, (d) represents the result of IF with MOS = 5.8368, (e) represents the result of UDCP with MOS = 4.0820, and (f) represents the result of NLD with MOS = 6.2122.

of developing an NR IQA metric even more challenging, which is to be investigated in this paper.

The contributions of this paper are:

- First, we exploit the image formation model [42]–[44], which describes the processes of underwater optical imaging. We propose to utilize the restoration model parameters that represent the physical optics information of underwater imaging to build relevant features for image quality assessment.
- Second, we investigate the properties of the human visual system (HVS), and build sophisticated image quality descriptors for underwater images. More specifically, we construct a new feature representation (i.e., Michaelson-like) to assess the contrast of underwater images, which can adequately reveal the changes of contrast under the low light environment; a new feature representation combining saliency and local binary patterns for the assessment of texture quality of underwater images; and a new feature representation based on color correlograms to capture the color properties of underwater images.
- Third, the above underwater-specific and HVS-based IQA features are integrated into a single metric. We thoroughly analyse the selection and individual contribution of IQA features, and compare our proposed metric to existing state-of-the-art IQA metrics, including general-purpose, underwater-specific and deep learning-based metrics. Our proposed metric demonstrates superior performance in assessing output quality of underwater image enhancement.

II. Problem Definitions and Challenges

We briefly describe the UEIQA database as the benchmark for measuring the output quality of underwater image enhancement. Also, we analyse the limitations and challenges of objective quality measurement for this particular application.

A. The Benchmark: Subjective Measure of Output Quality of Underwater Image Enhancement

We built a benchmark database for Underwater Enhancement Image Quality Assessment (UEIQA) [27]. This database contained 40 different underwater scenes with a resolution of 1280×720 pixels; and each scene yielded the results of five enhancement algorithms, including Contrast Limited Adaptive Histogram Equalization (HE) [16], Decorrelation Stretch (DS) [19], Image Fusion (IF) [20], Underwater Dark Channel Prior (UDCP) [18], and Non-local Dehazing (NLD) [45]. Hence, the UEIQA benchmark consists of 40 originally acquired images and 200 algorithmically enhanced images. A fully controlled perception experiment was conducted, where each image was evaluated by 18 subjects using a single-stimulus method as prescribed in [46]. The mean opinion score (MOS) of each stimulus was computed as the most reliable measure of perceived image quality [29]. Fig. 1 illustrates an original underwater image and results of five image enhancement algorithms in the UEIQA database; and the perceived quality (i.e., MOS) resulted from the perception study.

B. Limitations and Challenges: Objective Measure of Output Quality of Underwater Image Enhancement

As mentioned in Section I, the IQA metrics used for the application of underwater image enhancement must be in a no-reference (NR) framework. Now, we investigate the existing NR metrics, including general-purpose and underwater-specific metrics, and analyse the limitations and challenges.

1) General-Purpose Image Quality Metrics: There are several no-reference (NR) IQA metrics that have proven effective for general-purpose applications, including BRISQUE [35], NIQE [34], BLIINDS II [47], SSEQ [48], LPC [49]. These general-purpose NR metrics are based on measuring structural changes when images undergoing certain distortion types (e.g., BRISQUE, SSEQ, and LPC) or capturing universal statistical information inspired by the human visual system (e.g., NIQE and BLIINDS II). However, it is unknown whether these general-purpose metrics are helpful for measuring the output quality of underwater image enhancement. To this end, we evaluate these metrics on the UEIQA benchmark. The performance (i.e., Spearman rank order correlation coefficient (SROCC)) of BRISQUE, NIQE, BLIINDS II, SSEQ, and LPC metrics on the UEIQA benchmark is 0.40, 0.24, 0.27, 0.18, and 0.75, respectively (note the complete results of performance comparison can be found in Table IV). All metrics show poor or unsatisfactory correlation (e.g., SROCC < 0.8) with subjective quality, suggesting that these general-purpose NR metrics are useless for our application.

There are two possible reasons for the observed poor performance: (1) The above-mentioned IQA metrics have been designed to handle natural scenes, which exhibit different image characteristics than underwater images. Fig. 2 illustrates the histograms of an underwater image and an image of natural scene. It can be seen that the intensities of former concentrate on a narrow range of the histogram, whereas the histogram of latter

TABLE I
ESTIMATION OF BL AND TM BY DIFFERENT OPTICAL PRIOR BASED UNDERWATER IMAGE RESTORATION MODELS

Model index	Optical prior	BL Estimation	TM Estimation
U-Model 1	DCP [17]	$\mathbf{B_c}(\mathbf{I}) = \mathbf{I_c}(\arg\max_{\mathbf{x}\in\mathbf{I}}(\min_{\mathbf{c}}(\min_{\mathbf{y}\in\Omega(\mathbf{x})}\mathbf{I_c}(\mathbf{y}))))$	$\mathbf{t}(\mathbf{x}) = \mathbf{1} - \min_{\mathbf{c}}(\min_{\mathbf{y}\in\Omega(\mathbf{x})}(\mathbf{I_c}(\mathbf{y})/\mathbf{B_c}))$
U-Model 2	UDCP [18]	$\mathbf{B_c}(\mathbf{I}) = \mathbf{I_c}(\arg\max_{\mathbf{x}\in\mathbf{I}}(\min_{\mathbf{c}}(\min_{\mathbf{y}\in\Omega(\mathbf{x})}\mathbf{I_c}(\mathbf{y}))))$	$\mathbf{t}(\mathbf{x}) = \mathbf{1} - \min_{\mathbf{c}^*}(\min_{\mathbf{y}\in\Omega(\mathbf{x})}(\mathbf{I_{c^*}}(\mathbf{y})/\mathbf{B_{c^*}}))$
U-Model 3	MIP [51]	$\mathbf{MIP}(\mathbf{x}) = \max_{\mathbf{y}\in\Omega(\mathbf{x})}\mathbf{I_r}(\mathbf{y}) - \max_{\mathbf{c}^*}(\max_{\mathbf{y}\in\Omega(\mathbf{x})}\mathbf{I_{c^*}}(\mathbf{y}))$ $\mathbf{B_c}(\mathbf{I}) = \mathbf{I_c}(\arg\max_{\mathbf{x}\in\mathbf{I}}(\mathbf{MIP}(\mathbf{x})))$	$\mathbf{t}(\mathbf{x}) = \mathbf{MIP}(\mathbf{x}) + (\mathbf{1} - \max_{\mathbf{x}\in\mathbf{I}}(\mathbf{MIP}(\mathbf{x})))$
U-Model 4	MIP/UDCP [24]	$\mathbf{I_c^{dark}}(\mathbf{x}) = \min_{\mathbf{y}\in\Omega(\mathbf{x})}\mathbf{I_c}(\mathbf{y})$ $\mathbf{B_c}(\mathbf{I}) = \mathbf{I_c}(\arg\max_{\mathbf{x}\in\mathbf{I}}(\mathbf{I_c^{dark}}(\mathbf{x}) - (\max_{\mathbf{c}^*}\mathbf{I_{c^*}^{dark}}(\mathbf{x}))))$	$\mathbf{t_{c^*}}(\mathbf{x}) = \mathbf{1} - \min_{\mathbf{c}^*}(\min_{\mathbf{y}\in\Omega(\mathbf{x})}(\mathbf{I_{c^*}}(\mathbf{y})/\mathbf{B_{c^*}}))$ $\mathbf{t_r}(\mathbf{x}) = \alpha\max_{\mathbf{y}\in\Omega(\mathbf{x})}\mathbf{I_r}(\mathbf{y})$ $\alpha = \frac{\mathrm{Avg}(\mathbf{t_{c^*}})}{\mathrm{Avg}(\max_{\mathbf{y}\in\Omega(\mathbf{x})}\mathbf{I_r}(\mathbf{y}))}$
U-Model 5	IBLAP [14]	$\mathbf{B_c^{sel}} = \{\mathrm{Avg}(\mathbf{LV}(\mathbf{I_c})), \mathrm{Avg}(\mathbf{LB}(\mathbf{I_c}, \mathbf{P_{blr}})), \mathbf{DCP_c}\}$ $\mathbf{B_{min,c}} = \min\{\mathbf{B_c^{sel}}\}, \mathbf{B_{max,c}} = \max\{\mathbf{B_c^{sel}}\}$ $\mathbf{B_c}(\mathbf{I}) = \beta\mathbf{B_{max,c}} + (\mathbf{1} - \beta)\mathbf{B_{min,c}}$, β is sigmoid function based weight.	$\mathbf{d_D} = \mathbf{1} - \mathrm{Strech}(\mathbf{MIP}(\mathbf{x}))$ $\mathbf{d_R} = \mathbf{1} - \mathrm{Strech}(\max_{\mathbf{y}\in\Omega(\mathbf{x})}\mathbf{I_r}(\mathbf{y}))$ $\mathbf{d_B} = \mathbf{1} - \mathrm{Strech}(\mathrm{Reconstruct}(\mathbf{P_r}(\mathbf{x})))$, $\mathbf{P_r}(\mathbf{x}) = \max_{\mathbf{y}\in\Omega(\mathbf{x})}\frac{1}{\mathbf{n}}\sum_{\mathbf{i}=1}^{\mathbf{n}}\lvert\mathbf{I_{gray}}(\mathbf{x}) - \mathbf{G}^{\hat{\mathbf{i}},\hat{\mathbf{i}}}((\mathbf{x}))\rvert$ $\mathbf{t}(\mathbf{x}) = \gamma_2[\gamma_1\mathbf{d_D} + (\mathbf{1} - \gamma_1)\mathbf{d_R}] + (\mathbf{1} - \gamma_2)\mathbf{d_B}$, γ_1, γ_2 are the sigmoid function based weights.
U-Model 6	ULAP [52]	$\mathbf{d}(\mathbf{x}) = \mu_1 + \mu_2\max_{\mathbf{y}\in\Omega(\mathbf{x})}\mathbf{I_r}(\mathbf{y}) + \mu_3\max_{\mathbf{c}^*}(\max_{\mathbf{y}\in\Omega(\mathbf{x})}\mathbf{I_{c_*}}(\mathbf{y}))$, $\mu_1 = 0.5321, \mu_2 = -0.9106, \mu_3 = 0.5130$ $\mathbf{B_c}(\mathbf{I}) = \mathbf{I_c}(\arg\max_{\mathbf{x}\in\mathbf{P_{0.1\%}}}(\mathbf{d}(\mathbf{x})))$	$\mathbf{t_c}(\mathbf{x}) = \mathbf{Nrer_c^{d_a(x)}}$, $\mathbf{d_a}(\mathbf{x}) = \mathbf{10}\times\mathbf{d}(\mathbf{x})$, $\mathbf{Nrer_r} = \mathbf{0.83}, \mathbf{Nrrr_g} = \mathbf{0.95}, \mathbf{Nrer_b} = \mathbf{0.97}$

TABLE II
PERFORMANCE OF SIX VARIANTS OF UWEQM USING DIFFERENT TM FEATURE VECTORS $\mathbf{H_{TM}}$ DERIVED FROM DIFFERENT OPTICAL PRIOR BASED UNDERWATER IMAGE RESTORATION MODELS. RESULTS ARE BASED ON THE UEIQA DATABASE

Criteria \ Model	$\mathrm{UWEQM_{DCP}}$	$\mathrm{UWEQM_{UDCP}}$	$\mathrm{UWEQM_{MIP}}$	$\mathrm{UWEQM_{MIP/UDCP}}$	$\mathrm{UWEQM_{IBLAP}}$	$\mathrm{UWEQM_{ULAP}}$
PLCC	0.8774	0.8764	**0.8922**	0.8590	0.8794	0.8759
SROCC	0.8698	0.8602	**0.8816**	0.8520	0.8740	0.8705
KROCC	0.6998	0.6834	**0.7137**	0.6759	6976	0.6948
RMSE	1.1262	1.1250	**1.0534**	1.1851	1.1075	1.1095

distributes in a wider range of intensity. (2) These general-purpose IQA metrics have been designed to measure conventional signal distortions caused by compression and transmission. They may not necessarily capture the properties of the unique distortions caused by underwater environment, such as color cast caused by light attenuation in water [27]. More importantly, distortions that are inevitably introduced by an underwater image enhancement process/algorithm could complicate the inherent distortions [27]. Therefore, a dedicated IQA metric for underwater images undergoing enhancement is desirable.

2) Underwater-Specific Image Quality Metrics: Research has been undertaken to develop dedicated IQA metrics for underwater images. In 2014, Yang *et al.* first reported an underwater image/video quality evaluator which included a discriminator of underwater environment based on the log-contrast power spectrum. This block-based metric calculated a weighted-average of gray-scale contrast, the global brightness and block average gradient [39]. Then, in 2015, Yang *et al.* conducted a series of experiments to measure underwater image quality subjectively and objectively, and proposed an Underwater Color Image Quality Evaluator (UCIQE) [40] by transferring an RGB image to CIELab color space and linearly combining the chroma, contrast, and saturation. This metric aims to evaluate the color and sharpness factors of underwater images. Panetta *et al.* presented

TABLE III
PERFORMANCE OF UWEQM VARIANT MODELS BASED ON SINGLE FEATURE OR A COMBINED VECTOR OF MULTIPLE SUB-FEATURES. RESULTS ARE BASED ON THE UEIQA DATABASE

Model \ Criteria	PLCC	SROCC	KROCC	RMSE
$\mathbf{H_{TM}}$	0.7950	0.7791	0.6034	1.2925
$\mathbf{H_{MLC}}$	0.8183	0.8092	0.6208	1.2761
$\mathbf{H_{SLBP}}$	0.6636	0.6463	0.4738	1.6926
$\mathbf{H}_{\mathrm{RGB-CAC_s}}$	0.7040	0.6746	0.5086	1.5792
$\mathbf{H}_{\mathrm{HSV-CAC_s}}$	0.5662	0.5443	0.3912	1.8739
$\mathbf{H}_{\mathrm{RGB-CAC_s}}$, $\mathbf{H}_{\mathrm{HSV-CAC_s}}$	0.7319	0.7067	0.5226	1.5370
$\mathbf{H_{TM}}$, $\mathbf{H_{MLC}}$	0.8219	0.8072	0.6431	1.2561
$\mathbf{H_{TM}}$, $\mathbf{H_{SLBP}}$	0.8190	0.8062	0.6431	1.3010
$\mathbf{H_{MLC}}$, $\mathbf{H_{SLBP}}$	0.8344	0.8266	0.6466	1.2351
$\mathbf{H_{TM}}$, $\mathbf{H}_{\mathrm{RGB-CAC_s}}$, $\mathbf{H}_{\mathrm{HSV-CAC_s}}$	0.8134	0.7932	0.6106	1.3262
$\mathbf{H_{MLC}}$, $\mathbf{H}_{\mathrm{RGB-CAC_s}}$, $\mathbf{H}_{\mathrm{HSV-CAC_s}}$	0.8040	0.7880	0.6222	1.3283
$\mathbf{H_{SLBP}}$, $\mathbf{H}_{\mathrm{RGB-CAC_s}}$, $\mathbf{H}_{\mathrm{HSV-CAC_s}}$	0.7451	0.7189	0.5333	1.5205
$\mathbf{H_{TM}}$, $\mathbf{H_{MLC}}$, $\mathbf{H}_{\mathrm{RGB-CAC_s}}$, $\mathbf{H}_{\mathrm{HSV-CAC_s}}$	0.8570	0.8451	0.6647	1.1832
$\mathbf{H_{TM}}$,$\mathbf{H_{SLBP}}$, $\mathbf{H}_{\mathrm{RGB-CAC_s}}$, $\mathbf{H}_{\mathrm{HSV-CAC_s}}$	0.8154	0.7978	0.6113	1.3275
$\mathbf{H_{TM}}$, $\mathbf{H_{MLC}}$, $\mathbf{H_{SLBP}}$	0.8645	0.8586	0.6801	1.1591
$\mathbf{H_{MLC}}$, $\mathbf{H_{SLBP}}$, $\mathbf{H}_{\mathrm{RGB-CAC_s}}$, $\mathbf{H}_{\mathrm{HSV-CAC_s}}$	0.8228	0.8178	0.6302	1.3041
Proposed UWEQM	**0.8922**	**0.8816**	**0.7137**	**1.0534**

a new underwater IQA framework [50] inspired by the human visual system. The factors that cause low quality underwater images such as light attenuation and scattering effects were explicitly considered; and the color-cast, blurring and stereoscopic distortions were described by three individual metrics i.e., the Underwater Image Color Measure (UICM), Underwater Image Sharpness Measure (UISM), and Underwater Image Contrast Measure (UIConM), respectively. An overall metric, Underwater Image Quality Metric (UIQM) is a linear combination of UICM, UISM, and UIConM.

Now, we evaluate the underwater-specific IQA metrics on the UEIQA benchmark. The performance (i.e., Spearman rank order correlation coefficient (SROCC)) of UICM, UISM, UIConM, UIQM, and UCIQE metrics is 0.69, 0.49, 0.77, 0.29, and 0.57, respectively (note the complete results of performance comparison can be found in Table IV). It can be seen that these metrics that include underwater-specific features still show limited capabilities (e.g., SROCC<0.8) in assessing the output quality of underwater enhancement. The possible reasons are (1) Some methods are based on the histogram of features, which intuitively represents the statistical characteristics of underwater images. However, local information is not sufficiently reflected by these histogram-based methods. Some methods include the human visual system (HVS) properties, e.g., by measuring the perceptual preference of color, sharpness and contrast. These methods adopt simple HVS features and combine these features using a simple linear model, which however, neglect the way (non-linear behaviour) humans perceive the overall quality of images. (2) Moreover, image enhancement algorithms applied to underwater images often adopt physical optics prior information for signal restoration. The output quality of underwater image enhancement is not only related to the inherent distortions caused by underwater environment but also related to the signal distortions induced by the enhancement algorithms. Therefore, for the problem of assessing the output quality of image enhancement, the unique challenges regarding what makes an enhanced image of "better" quality and how to measure the output quality are largely unsolved and are to be investigated in this paper.

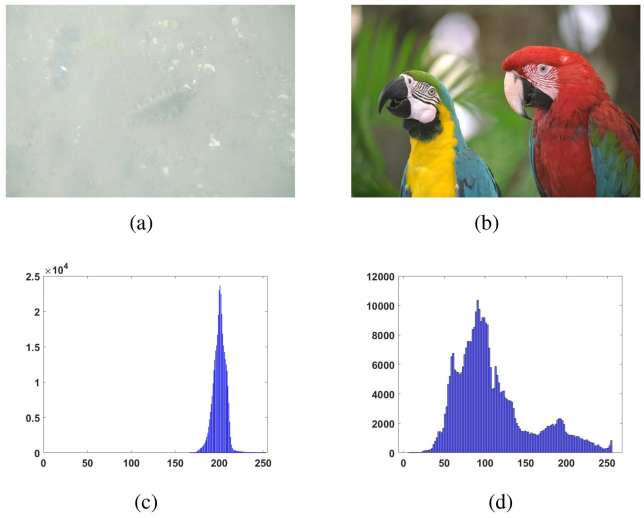


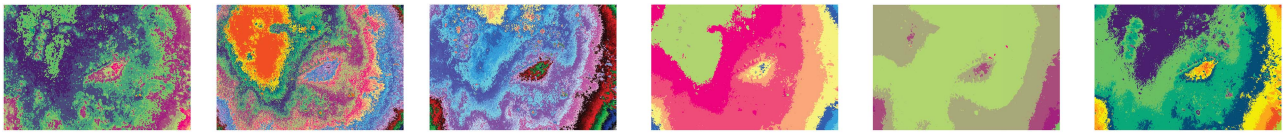
Fig. 2. Illustration of the histograms of an underwater image (a) and an image of natural scene (b). The intensities of former concentrate on a narrow range of the histogram (c), whereas the histogram, and (d) of latter distributes in a wider range of intensity.

III. PROPOSED METHOD

Since our goal is to evaluate the output quality of underwater image enhancement and there is no reference image of "ideal" quality, we consider image attributes that explicitly reflect underwater optical imaging and human perceptions of image quality aspects. To make a realistic metric, we construct two pivotal types of attributes: the first type defines the underlying physics

TABLE IV
PERFORMANCE COMPARISON OF THE PROPOSED UWEQM METRIC TO THE STATE-OF-THE-ART NO-REFERENCE IQA METRICS ON UEIQA DATABASE

Criteria	BRISQUE	NIQE	BLIINDS II	SSEQ	LPC	UICM	UISM	UIConM	UIQM	UCIQE	UEIQM	UWEQM
PLCC	0.4148	0.3999	0.1718	0.1789	0.7591	0.7082	0.5204	0.7813	0.5230	0.6159	0.8055	**0.8922**
SROCC	0.4024	0.2366	0.2708	0.1764	0.7583	0.6947	0.4930	0.7720	0.2886	0.5691	0.7657	**0.8816**
KROCC	0.2737	0.1662	0.1948	0.1051	0.5659	0.4969	0.3429	0.5819	0.1832	0.4101	0.5806	**0.7137**
RMSE	1.9200	2.1998	2.0779	2.2058	1.4595	1.5829	1.9145	1.3994	1.7977	1.6617	1.2508	**1.0534**



(a) DCP-TM map (b) UDCP-TM map (c) MIP-TM map (d) MIP/UDCP-TM map (e) IBLAP-TM map (f) ULAP-TM map

Fig. 3. Illustration of TM maps (visualized by MATLAB's colormap function) of an original underwater image (i.e., Fig. 1(a)) estimated by U-Model 1, U-Model 2, U-Model 3, U-Model 4, U-Model 5, and U-Model 6, respectively.

of underwater optical imaging; the second type addresses the human visual system (HVS) properties, including contrast, texture, color and attention. IQA features are formulated and spatial feature representations are calculated. These extracted features are combined to yield a final metric, using the random forest regression algorithm.

A. Underwater Physics Attribute

Based on the image acquisition system and environment, underwater optical imaging can be formulated as an image formation model (IFM) of Jaffe-McGlamery [53]. An underwater image can be represented as the linear superposition of three components:

$$E_{\mathrm{T}}(\mathbf{x}) = E_{\mathrm{DT}}(\mathbf{x}) + E_{\mathrm{FS}}(\mathbf{x}) + E_{\mathrm{BS}}(\mathbf{x}), \tag{1}$$

where $\mathbf{x}$ denotes the pixel position of the image, $E_{\mathrm{T}}(\mathbf{x})$ is the total energy received by the underwater camera, $E_{\mathrm{DT}}(\mathbf{x})$ represents the energy component of direct transmission, $E_{\mathrm{FS}}(\mathbf{x})$ and $E_{\mathrm{BS}}(\mathbf{x})$ are the energy components of forward light scattering and backward light scattering.

To approach underwater imaging as a restoration process, the Jaffe-McGlamery model of equation (1) could be interpreted as an inverse engineering problem. To this end, a simplified formation model [42]–[44] is proposed and widely used in the literature for dealing with underwater images. The simplified image formation model (or underwater image restoration model) describes the relationship between the observed image and its restored image under the conditions of underwater environment, and can be described as:

$$\mathbf{I}(\mathbf{x}) = \mathbf{t}(\mathbf{x})\mathbf{J}(\mathbf{x}) + (1 - \mathbf{t}(\mathbf{x}))\mathbf{B}, \tag{2}$$

where $\mathbf{I}$ is the raw captured image by the camera, $\mathbf{J}$ is the restored image, $\mathbf{t}(\mathbf{x})$ is the transmission medium (TM) map, and $\mathbf{B}$ is the background light (BL). As presented in the IFM model of equation (1), underwater image degradation is caused by light attenuation and light scattering. Therefore, a prior that represents the optical properties is often constructed to form the restored image $\mathbf{J}$ in the IFM model of equation (2), such as Dark Channel Prior (DCP), Underwater Dark Channel Prior (UDCP), Maximum Intensity Prior (MIP), Image Blurriness and Light Absorption Prior (IBLAP), Underwater Light Attenuation Prior (ULAP), as shown in Table I. Based on these optical priors, the BL and TM maps are estimated from the camera captured image $\mathbf{I}$ by solving the inverse problem of equation (2). These models (i.e., U-Model 1 to U-Model 6 in table I) represent state-of-the-art optical prior based underwater image restoration methods. Fig. 3 illustrates the TM maps (visualized by MATLAB's colormap function) of an original underwater image (i.e., Fig. 1(a)) estimated by U-Model 1 [17], U-Model 2 [18], U-Model 3 [51], U-Model 4 [24], U-Model 5 [14], and U-Model 6 [52], respectively. Fig. 4 illustrates the TM maps estimated by the one of these models (i.e., U-Model 1) for original and enhanced images (i.e., Fig. 1(a)-(f)).

For our problem of assessing output quality of image enhancement, there is no "ideal" reference image of perfect quality. We could formulate this problem using the above image formation model of equation (2) by assuming the current/test image (i.e., $\mathbf{I}$ in equation (2)) remains sub-optimal and its quality could be further improved towards a restored image of optimal quality (i.e., $\mathbf{J}$ in equation (2)). Since we do not need to produce the final restored image, we only use the optical prior based model to approximate the TM as the underwater physics-related attribute to form part of our IQA framework (note it can be seen from Table I that TM is highly related to BL hence only TM is used for the sake of simplicity). The formulas for TM estimation of DCP-based, UDCP-based, MIP-based, MIP/UDCP-based, IBLAP-based, and ULAP-based models are listed in Table I, where $\mathbf{c} \in \{r, g, b\}$, $\mathbf{c}^* \in \{g, b\}$, and $\mathbf{r}$, $\mathbf{g}$, $\mathbf{b}$ represent the R, G, B channels in the RGB colorspace.

B. HVS-Based Attributes

Now, we present the HVS-based descriptors for the evaluation of output quality of underwater image enhancement, including Michaelson-like contrast map, salient local binary patterns, and color autocorrelograms.

1) Michaelson-Like Contrast Map (MLC): Due to backward scattering of light, underwater images suffer serious contrast

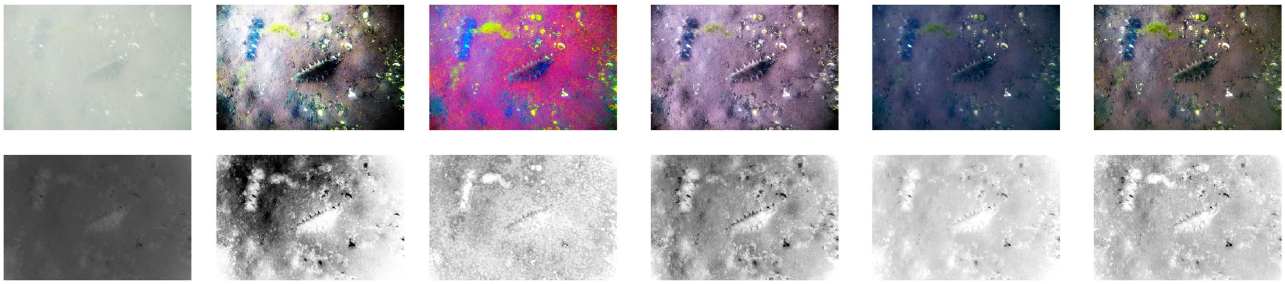


Fig. 4. Illustration of underwater physics attribute. Top row from left to right: original underwater image and results of five image enhancement algorithms used in the UEIQA database [27]. Second row shows the transmission medium (TM) maps estimated by one of the optical prior based underwater image restoration models (i.e., U-Model 1 [17] in Table I).

degradation. However, normal contrast measures are not able to quantify the contrast of underwater images under the low luminance conditions. Based on the properties of the human visual system (HVS), Panneta *et al.* introduced the Parameterized Logarithmic Image Processing (PLIP) operations that could overcome the problem of low luminance conditions.

Let an RGB image denoted as $\mathbf{I}$, and $\mathbf{I}$ be gray-scaled as $I_{gray} = (I_r + I_g + I_b)/3$. For any pixel $\mathbf{x} \in I_{gray}$, $\Omega(\mathbf{x})$ is the 3×3 window by both-side symmetric padding, and $I_{max}(\mathbf{x}) = \max_{\mathbf{y}\in\Omega(\mathbf{x})} I_{gray}(\mathbf{y})$, $I_{min} = \min_{\mathbf{y}\in\Omega(\mathbf{x})} I_{gray}(\mathbf{y})$. We define the Michaelson-like contrast map (MLC) at each pixel position $\mathbf{x}$ as follows:

$$\mathrm{MLC}(\mathbf{x}) = \frac{I_{max}(\mathbf{x})\Theta I_{min}(\mathbf{x})}{I_{max}(\mathbf{x}) \oplus I_{min}(\mathbf{x})} * \log\left(\frac{I_{max}(\mathbf{x})\Theta I_{min}(\mathbf{x})}{I_{max}(\mathbf{x}) \oplus I_{min}(\mathbf{x})}\right), \tag{3}$$

where the symbols Θ, $\oplus$ and $*$ represent the PLIP operations that are detailed below. First, the following gray-scale tone function is used to process the original image as absorption filters:

$$g(\mathbf{x}) = M - I_{gray}(\mathbf{x}), \tag{4}$$

where M represents the maximum of absorption of the human eye and is set to be 1026 according to [54]. Then, the operations Θ, $\oplus$ and $*$ are defined as follows:

$$p \oplus q = p + q - \frac{pq}{\gamma(M)}, \tag{5}$$

$$p\Theta q = \frac{p - q}{k(M) - q}, \tag{6}$$

$$p * q = \phi^{-1}(\phi(p)\phi(q)), \tag{7}$$

$$\phi(p) = -\lambda(M)\ln^{\beta}\left(1 - \frac{f}{\lambda(M)}\right), \tag{8}$$

$$\phi^{-1}(p) = -\lambda(M)\left\{1 - [\exp(-\frac{f}{\lambda(M)})]^{\frac{1}{\beta}}\right\}, \tag{9}$$

where $\oplus$ is denoted as PLIP subtraction, Θ as PLIP addition, $*$ as PLIP multiplication, p, q are the gray-scale tone pixel intensity values, f is the corresponding original gray-scale image intensity value, β is a constant, and $\gamma(M)$, $k(M)$, $\lambda(M)$ are arbitrary functions of M. According to the analysis of Panneta *et al.* [54], The parameter and functions are set as $\gamma(M) = k(M) = \lambda(M) = 1026$ and $\beta = 2$.

2) Salient Local Binary Patterns (SLBP): The local binary patterns (LBP) operator is adopted to describe the texture information of images. The LBP method was first proposed by Ojala et.al. [55]. Let I be a gray-scale image, $\Omega_R^P(\mathbf{x})$ is the P-neighbourhood centered at pixel $\mathbf{x} = (i_x, j_x)$ with a radius R. For arbitrary point $\mathbf{p} = (i_p, j_p)$ can be expressed as:

$$i_p = \left\lfloor i_x + R\cos\left(2\pi\frac{P - p + 2}{P}\right)\right\rceil, \tag{10}$$

$$j_p = \left\lfloor j_x - R\sin\left(2\pi\frac{P - p + 2}{P}\right)\right\rceil, \tag{11}$$

where $p \in \{0, 1, \ldots, P-1\}$, and $\lfloor m \rceil$ means the nearest integer of m. Define a detection function by the following formula:

$$\sigma(u, v) = \begin{cases} 1, & \text{if } u - v \geq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{12}$$

The LBP is computed in the P-neighbourhood with the center $\mathbf{x} = (i_x, j_x)$ and radius R by the following equation:

$$L_R^P(\mathbf{x}) = \sum_{p=0}^{P-1} \sigma(I(i_p, j_p), I(i_x, j_x))2^p. \tag{13}$$

From this definition, a feature vector (histogram) is yielded. For example, when $R = 1, P = 8$, and $L_R^P(\mathbf{x}) \in [0, 255]$, a 256-dimensional feature vector is formed. To reduce the dimension of the output feature vector, Ojala et.al. [55] improved LBP into the "uniform LBP" method that defined as

$$\tilde{L}_R^P(\mathbf{x}) = \begin{cases} \sum_{p=0}^{P-1} \sigma(I(i_p, j_p), I(i_x, j_x)), & \text{if } \Upsilon \leq 2, \\ P + 1, & \text{otherwise,} \end{cases} \tag{14}$$

where

$$\Upsilon = \sum_{p=1}^{P-1} |\sigma(I(i_p, j_p), I(i_x, j_x)) - \sigma(I(i_{p-1}, j_{p-1}), I(i_x, j_x))| + |\sigma(I(i_p, j_p), I(i_x, j_x)) - \sigma(I(i_0, j_0), I(i_x, j_x))|. \tag{15}$$

For an input image, the LBP feature well captures local texture information. However, the local spatial information is not equally visible to the human visual system (HVS) [56]. It is well known that visual attention plays a significant role in image quality assessment [57]–[60]. To incorporate visual attention, we calculate a saliency map (using the model proposed by Vikram *et al.* [56]) of the input image and use it to weight the

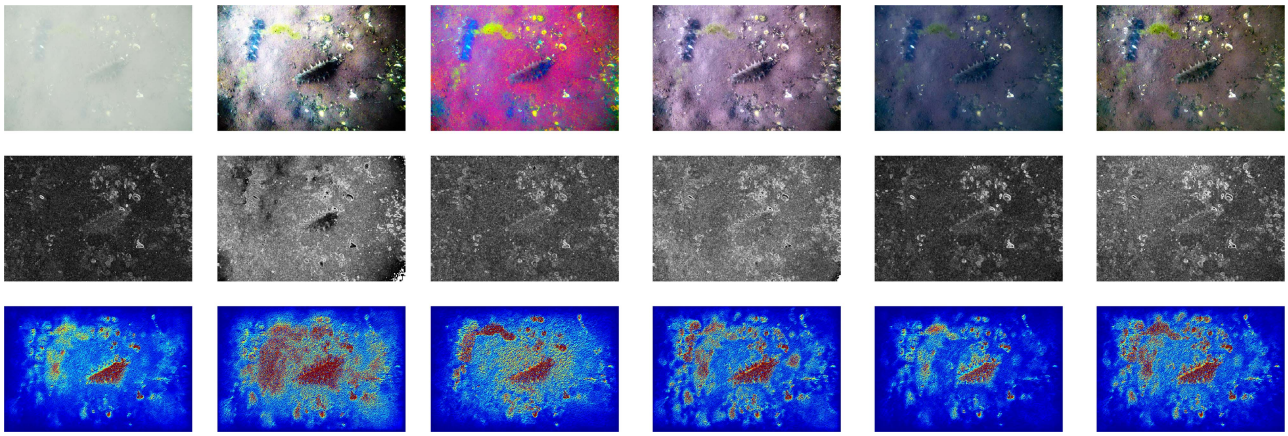


Fig. 5. Illustration of the HVS-based attributes. Top row from left to right: original underwater image and results of five image enhancement algorithms used in the UEIQA dataset [27]. Second row shows the Michaelson-like contrast (MLC) maps. Third row shows the salient local binary patterns (SLBP) maps.

corresponding LBP map. The saliency weighted LBP is called salient local binary patterns map (SLBP) [56]. Let SM denote the saliency map, $\overline{SM}$ is the [0,1]-normalized saliency map, and $LBP = (L_R^P(\mathbf{x}))_{\mathbf{x}\in I}$ is the LBP map of the input image I. The salient local binary patterns (SLBP) is defined pixel-wise as

$$SLBP(\mathbf{x}) = \overline{SM}(\mathbf{x}) L_R^P(\mathbf{x}). \tag{16}$$

Fig. 5 illustrates the feature maps, including Michaelson-like contrast (MLC) map and salient local binary patterns (SLBP) map.

3) Color Autocorrelogram (CAC): Color correction/ equalization plays an important role in underwater image enhancement algorithms. For our goal of assessing output quality of image enhancement, it would be beneficial for the IQA metric to include colour perception-related attributes that account for the behaviour of the human visual system. We adopt the color autocorrelogram index [61], which has proven superior in capturing the spatial correlation of colors in an image.

Let $\mathbf{I}_{in}$ denote an index image of $\mathbf{I}$ with size $m \times n$, we quantify the colors of $\mathbf{I}_{in}$ into m colors $c_1, c_2, \ldots, c_m$. Let $\mathbf{I}_{in}(\mathbf{x})$ denote the color of pixel $\mathbf{x}$ and define the color level set $\mathbf{I}_{in}^{c_i} \triangleq \{\mathbf{x} | \mathbf{I}_{in}(\mathbf{x}) = c_i\}$ for $i = 1, 2, \ldots, m$. Besides, the distance of different pixels $\mathbf{x}_1$ and $\mathbf{x}_2$ is measured by using the L_∞-norm, i.e., $\text{dist}(\mathbf{x}_1, \mathbf{x}_2) = ||\mathbf{x}_1 - \mathbf{x}_2||_\infty \triangleq \max\{|x_1 - x_2|, |y_1 - y_2|\}$ for pixels $\mathbf{x}_1(x_1, y_1)$ and $\mathbf{x}_2(x_2, y_2)$. Then, the color correlogram of image $\mathbf{I}_{in}$ at color piont (c_i, c_j) is defined as the following formula:

$$\mathbf{M}_{c_i,c_j}^k = \Pr(\{\mathbf{x}_1 \in \mathbf{I}_{in}^{c_i}, \mathbf{x}_2 \in \mathbf{I}_{in}^{c_j} | \text{dist}(\mathbf{x}_1, \mathbf{x}_2) = k\}), \tag{17}$$

where $i, j \in \{1, 2, \ldots, m\}$, $k \in \{1, 2, \ldots, d\}$ with a fixed priori distance d, and $\Pr(R_{i,j})$ denotes the probability of the random event $R_{i,j} = \{\mathbf{x}_1 \in \mathbf{I}_{in}^{c_i}, \mathbf{x}_2 \in \mathbf{I}_{in}^{c_j} | \text{dist}(\mathbf{x}_1, \mathbf{x}_2) = k\}$. The overall color correlogram $\mathbf{M}$ of image $\mathbf{I}_{in}$ is defined as a $m \times md$ matrix, i.e., $\mathbf{M} = [\{\mathbf{M}_{c_i,c_j}^1\}_{m\times m}, \{\mathbf{M}_{c_i,c_j}^2\}_{m\times m}, \ldots, \{\mathbf{M}_{c_i,c_j}^d\}_{m\times m}]$. Specially, the color autocorrelogram of image $\mathbf{I}_{in}$ that only captures spatial correlation of identical color c_i is formulated as follows:

$$\mathbf{M}_{c_i,c_i}^k = \Pr(\{\mathbf{x}_1, \mathbf{x}_2 \in \mathbf{I}_{in}^{c_i} | \text{dist}(\mathbf{x}_1, \mathbf{x}_2) = k\}), \tag{18}$$

where $i \in \{1, 2, \ldots, m\}$. Then, the overall color autocorrelogram CAC of image $\mathbf{I}_{in}$ is defined as a md vector, i.e.,

$$\text{CAC} = \left(\mathbf{M}_{c_1,c_1}^1, \ldots, \mathbf{M}_{c_m,c_m}^1, \ldots, \mathbf{M}_{c_1,c_1}^d, \ldots, \mathbf{M}_{c_m,c_m}^d\right). \tag{19}$$

In order to improve the efficiency of CAC algorithm, the probability definition in equation (18) can be transformed into the following counting problem:

$$\mathbf{S}_{c_i,c_i}^k = |\{\mathbf{x}_1, \mathbf{x}_2 \in \mathbf{I}_{in}^{c_i} | \text{dist}(\mathbf{x}_1, \mathbf{x}_2) = k\}|, \tag{20}$$

where $|\cdot|$ denotes the cardinality of a set. The overall simplified color autocorrelogram CAC_s of image $\mathbf{I}_{in}$ is defined as:

$$\text{CAC}_\text{s} = (\mathbf{S}_{c_1,c_1}^1, \ldots, \mathbf{S}_{c_m,c_m}^1, \ldots, \mathbf{S}_{c_1,c_1}^d, \ldots, \mathbf{S}_{c_m,c_m}^d). \tag{21}$$

To enhance the feature robustness [62], we compute the autocorrelogram on both RGB colorspace (conventionally used, i.e., RGB-CAC_s) and HSV colorspace (perceptually uniform, i.e., HSV-CAC_s).

C. Overall Metric

To generate an overall IQA metric, individual features are formulated to vector representations, and combined to yield a predictor using the random forest regression algorithm.

1) Feature Vectors: We have generated features representing underwater physics attribute, i.e., transmission medium (TM) map; and HVS-based attributes i.e., Michaelson-like contrast (MLC), Salient local binary patterns (SLBP), and simplified color autocorrelograms (RGB-CAC_s, HSV-CAC_s). Now, we transform these features into feature vectors for subsequent processes. Since the local binary patterns (LBP) operator is a powerful visual descriptor, we use this method to construct feature vectors for certain feature maps as extracted in this paper. For the TM and MLC feature maps, the transformation process is performed using the equations (12)-(15), where an LBP map

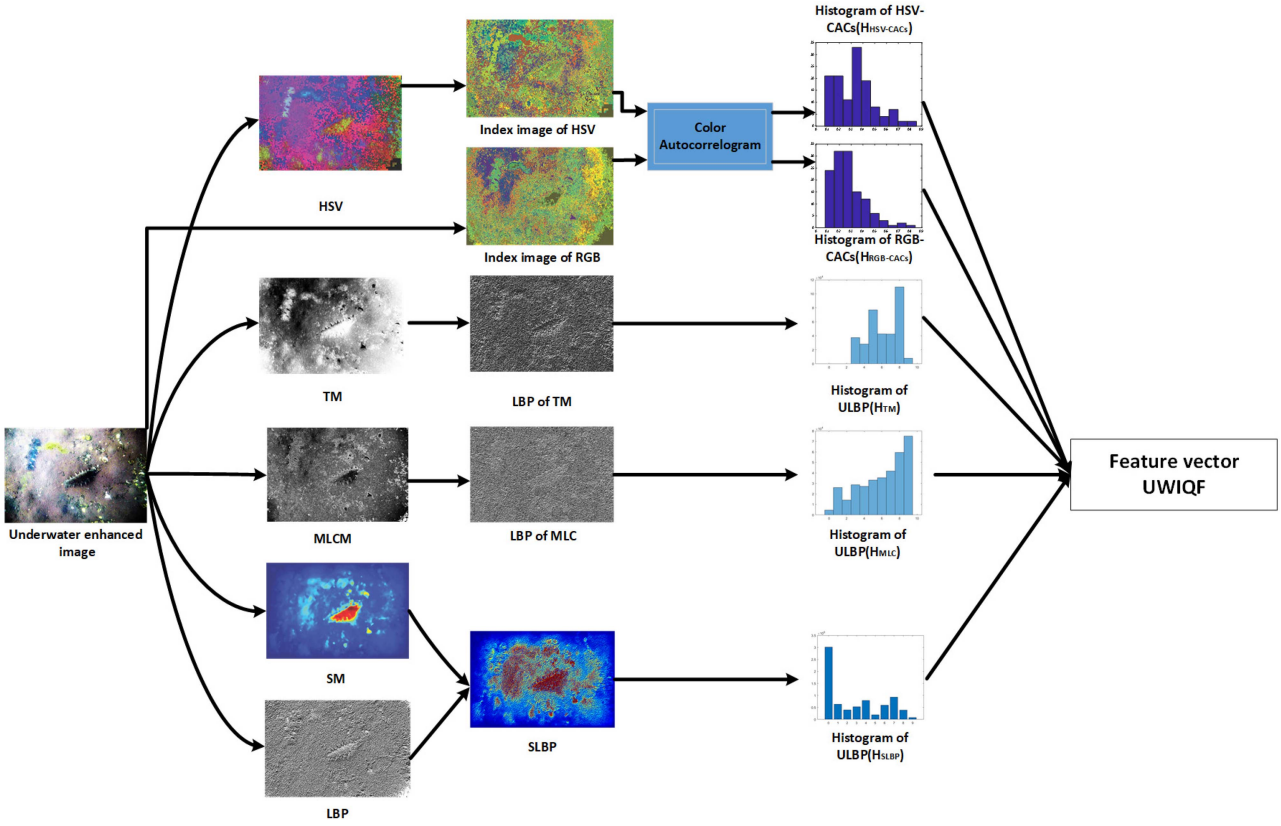


Fig. 6. The schematic overview of the proposed underwater image quality-aware feature (UWIQF). TM, MLCM, SM, LBP indicate transmission medium, Michaelson-like contrast map, saliency map and local binary patterns. HSV and RGB indicate HSV colorspace and RGB colorspace.

is first calculated and then converted to a 10-dimensional feature vector. This process yields two feature vectors, $\mathbf{H_{TM}}$ and $\mathbf{H_{MLC}}$. For the SLBP feature map, we follow the approach taken in [56], which improves the robustness of a feature vector when saliency weighting is applied for LBP. The implementation is detailed as below.

The histogram/feature vector of SLBP (i. e., $\mathbf{H_{SLBP}}$) is defined as

$$H^P_{SLBP,R} = \{f^P_R(0), f^P_R(1), \ldots, f^P_R(P+1)\}, \tag{22}$$

where the frequency of $H^P_{SLBP,R}$ at each bin is defined as:

$$f^P_R(p) = \sum_{\mathbf{x}\in I} SM(\mathbf{x})\delta(\tilde{L}^P_R(\mathbf{x}), p), \tag{23}$$

where $p \in \{0, \ldots, P+1\}$, and the δ-function is defined as

$$\delta(u,v) = \begin{cases} 1, & \text{if } u = v, \\ 0, & \text{otherwise.} \end{cases} \tag{24}$$

Moreover, in order to keep the uniform property of extracted features, the simplified color autocorrelograms of RGB-$\mathrm{CAC_s}$, HSV-$\mathrm{CAC_s}$ are transformed into two histograms of 10 bins each. This forms two 10-dimensional feature vectors $\mathbf{H}_{\mathrm{RGB-CAC_s}}$ and $\mathbf{H}_{\mathrm{HSV-CAC_s}}$. After individual feature vectors are yielded, they are each normalised to represent the probability density function of the feature. These resulting vectors are combined to form an underwater image quality-aware feature (UWIQF) vector:

$$[\mathbf{H_{TM}}|\mathbf{H_{MLC}}|\mathbf{H_{SLBP}}|\mathbf{H}_{\mathrm{RGB-CAC_s}}|\mathbf{H}_{\mathrm{HSV-CAC_s}}]. \tag{25}$$

Fig. 6 illustrates the schematic overview of the proposed underwater image quality-aware feature (UWIQF).

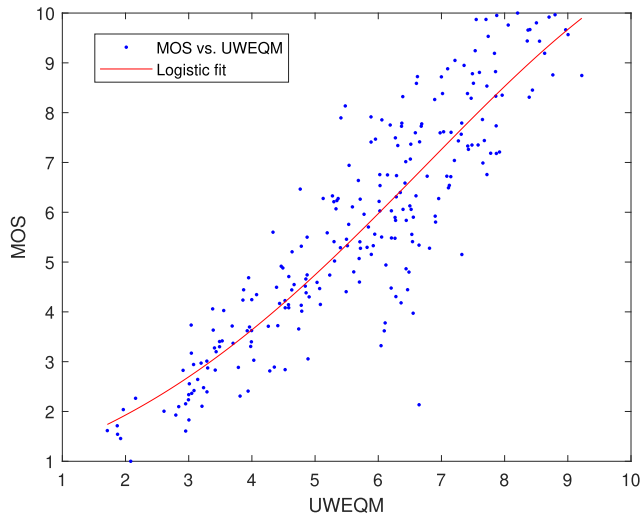


Fig. 7. Scatter plot of MOS versus our proposed UWEQM metric on the UEIQA database. The curved line shows the nonlinear logistic fit.

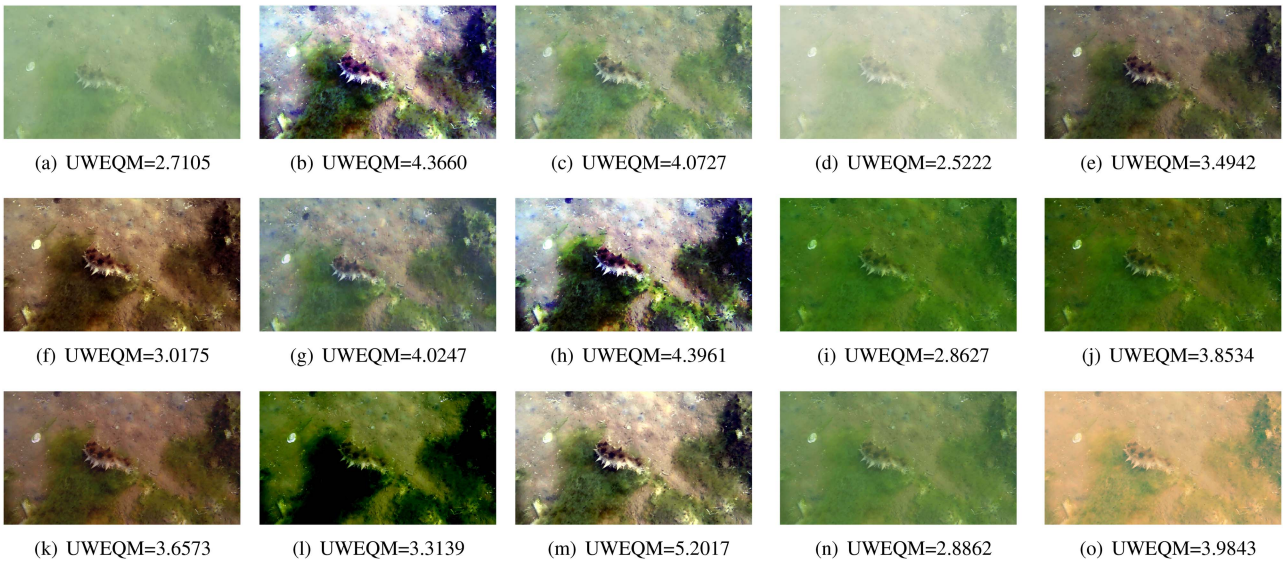


Fig. 8. An example to demonstrate the generalization ability of our proposed UWEQM metric. (a) an original underwater image; (b)-(o) results of underwater enhancement algorithms that are not used in the UEIQA database, including HE, CLAHE, GC, ICM, UCM, RD, HVSMIF, DCP, UDCP, GB-UDCP, IBLA, RGHS, RoWS, ULAP, respectively. In terms of the output quality, subjective quality could be visually assessed by the readers, and objective quality is predicted by our UWEQM metric.

2) Image Quality Metric: The UWIQF is used to build a metric for the assessment of quality of underwater image enhancement. Due to the high-dimensional and multi-modal feature space, we use the learning-based random forest regression method to establish the relationship between the UWIQF feature and the perceived image quality (i.e., MOS). Let $\mathcal{L} = \{(\text{UWIQF}_n, \text{MOS}_n)|i = 1, \ldots, N\}$ denote the training set, Θ is a probability space, and $\Theta_1, \ldots, \Theta_M \in \Theta$ are random variables in the same probability space. Assume that a set of M random regression tree models $\{\psi_{\mathcal{L},\Theta_m}|m = 1, \ldots, M\}$ trained on the same dataset $\mathcal{L}$ with different random seed Θ_m. A random forest regression model $\psi_{\mathcal{L},\Theta_1,\ldots,\Theta_M}$ integrates the set of M random regression tree models by averaging, yielding a ensemble model [63]:

$$\psi_{\mathcal{L},\Theta_1,\ldots,\Theta_M}(\text{UWIQF}) = \frac{1}{M}\sum_{m=1}^{M}\psi_{\mathcal{L},\Theta_m}(\text{UWIQF}), \quad (26)$$

The training error of the random forest regression model $\psi_{\mathcal{L},\Theta_1,\ldots,\Theta_M}$ can be computed as follows:

$$\begin{aligned}\text{Err}_{train} &= \frac{1}{N}\sum_{n=1}^{N}(\psi_{\mathcal{L},\Theta_1,\ldots,\Theta_M}(\text{UWIQF}_n) - \text{MOS}_n)^2 \\ &= \frac{1}{N}\sum_{n=1}^{N}\left(\frac{1}{M}\sum_{m=1}^{M}\psi_{\mathcal{L},\Theta_m}(\text{UWIQF}_n) - \text{MOS}_n\right)^2\end{aligned} \quad (27)$$

The hyperparameters $(\Theta_1, \ldots, \Theta_M)$ are trained by minimizing the above error of the ensemble model $\psi_{\mathcal{L},\Theta_1,\ldots,\Theta_M}$. Moreover, the ambiguity decomposition of Err_{train} guarantees that the performance of random forest model is better than a single regression tree [64]. Assume $(\tilde{\Theta}_1, \ldots, \tilde{\Theta}_M)$ are the trained model parameters, in the prediction stage, for any test quality-aware feature UWIQF, the predicted image quality is given by the following formula:

$$\psi_{\mathcal{L},\tilde{\Theta}_1,\ldots,\tilde{\Theta}_M}(\text{UWIQF}) = \frac{1}{M}\sum_{m=1}^{M}\psi_{\mathcal{L},\tilde{\Theta}_m}(\text{UWIQF}). \quad (28)$$

This gives an UnderWater image Enhancement Quality Metric (UWEQM).

IV. Experimental Results and Analysis

In this section, we detail the implementation of the proposed UWEQM algorithm. We also analyse the selection of optical prior based models for the underwater physics attribute, and contribution of individual IQA features. Moreover, the performance of the proposed UWEQM metric is compared to state-of-the-art metrics on the UEIQA database to show its superiority.

A. Experimental Setup

In this subsection, we present the experimental setups of our proposed UWEQM algorithm that include evaluation criteria, alternative IQA metrics for comparative analysis and model implementation.

1) Evaluation Criteria: To evaluate the performance of an IQA metric, the correlation and/or difference between the metric outputs and subjective mean opinion scores (MOSs) must be calculated. Four measures are commonly used including the Pearson linear correlation coefficient (PLCC), root mean square error (RMSE), Spearman rank order correlation coefficient (SROCC), and Kendall's rank order correlation coefficient (KROCC) [29]:

$$\text{PLCC} = \frac{\sum_{i=1}^{K}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{K}[(x_i - \bar{x})(y_i - \bar{y})]^2}}, \quad (29)$$

$$\text{RMSE} = \sqrt{\frac{1}{K}\sum_{i=1}^{K}(x_i - y_i)^2}, \tag{30}$$

$$\text{SROCC} = \frac{\sum_{i=1}^{K}(R(x_i) - \overline{R(x)})(R(y_i) - \overline{R(y)})}{\sqrt{\sum_{i=1}^{K}(R(x_i) - \overline{R(x)})^2}\sqrt{\sum_{i=1}^{K}(R(y_i) - \overline{R(y)})^2}}, \tag{31}$$

$$\text{KROCC} = \frac{2[N(\text{concordant pairs}) - N(\text{discordant pairs})]}{K(K-1)}, \tag{32}$$

where K is the number of test images, x_i and y_i indicate the metric score and subjective MOS of the i-th test image, respectively, $\bar{x}$ and $\bar{y}$ are the mean values, $R(\cdot)$ denotes the rank, and $N(\cdot)$ represents the amount of variable values. To account for the non-linear behaviour of subjective scoring, a logistic non-linear regression is commonly used to fit the metric scores to subjective MOSs [46]:

$$f(x_c) = \beta_1 \left[\frac{1}{2} - \frac{1}{1 + \exp[\beta_2(x_c - \beta_3)]}\right] + \beta_4 x_c + \beta_5, \tag{33}$$

where x_c represents the metric score, and the parameters are estimated using the MATLAB's $nlinfit$ function. PLCC and RMSE are computed between $f(x_c)$ and MOSs, whereas SROCC and KROCC are computed between x_c and MOSs. The closer the values of PLCC, SROCC and KROCC to 1 and the value of RMSE to 0, the better the performance of an IQA metric.

2) Alternative IQA Metrics for Comparative Analysis: We compare our proposed UWEQM metric with 11 state-of-the-art IQA metrics including BRISQUE [35], NIQE [34], BLIINDS II [47], SSEQ [48], LPC [49], UICM [50], UISM [50], UIConM [50], UIQM [50], UCIQE [40] and UEIQM [27]. UEIQM is a simple proof-of-concept IQA metric proposed in our recent work [27]. The UEIQM metric is built by combining three simple off-the-shelf features representing color, sharpness, and contrast. Note all metrics are no-reference (NR) models, so they are suitable for our target application of assessing output quality of underwater image enhancement.

3) Model Implementation - K-Fold Cross-Validation: To fairly evaluate the proposed UWEQM metric and its generalizability, we adopt a standard k-fold cross-validation. To this end, the benchmark UEIQA database is randomly partitioned into k (k=5 in our experiment) equal sized non-overlapped subsets; and one subset is used for testing and the other subsets are used for training. The cross-validation process is iterated 5 times, with each of the 5 subsets used exactly once as the testing set. The 5-time test results are averaged to yield a single estimation. Also, to reduce variability, multiple (i.e., 100 rounds in our experiment) rounds of cross-validation are performed using different random partitions, and the results are averaged over all rounds to give a final estimation of the IQA metric's predictive performance.

B. Selection of Underwater Physics Attribute

As mentioned in Section III.A, the transmission medium (TM) represents the underwater physics attribute that is derived from a optical prior based underwater image restoration model. Based on the six state-of-the-art models in Table I, six different TM maps can be estimated, resulting in six alternatives of feature vector $\mathbf{H_{TM}}$. Now, we investigate which TM estimation gives the best performance to our proposed UWEQM metric on the UEIQA database. To this end, six variants of UWEQM are created using different TM feature vectors, including UWEQM $_{\text{DCP}}$, UWEQM$_{\text{UDCP}}$, UWEQM$_{\text{MIP}}$, UWEQM$_{\text{MIP/UDCP}}$, UWEQM$_{\text{IBLAP}}$, and UWEQM$_{\text{ULAP}}$. In Table II, we list the performance of UWEQM variants on the UEIQA database. It can be seen that the UWEQM$_{\text{MIP}}$ yields the best performance; however, other UWEQM variants also produce comparable results. This suggests that MIP could be the most suitable optical prior for the images of the UEIQA database, where the depth of objects is about $3-4$ m in water. Hence the UWEQM$_{\text{MIP}}$ is directly denoted as UWEQM in the following of this paper. It should be noted that the selection of underwater physics attribute can adapt to specific imaging conditions, and the procedure can be used to find the feature that better suits the specific IQA problem or application.

C. Analysis of Individual Features

Since five different IQA features (i.e., $\mathbf{H_{TM}}$, $\mathbf{H_{MLC}}$, $\mathbf{H_{SLBP}}$, $\mathbf{H}_{\text{RGB}-\text{CAC}_\text{s}}$, and $\mathbf{H}_{\text{HSV}-\text{CAC}_\text{s}}$) are extracted to form the overall UWIQF vector of the proposed UWEQM metric, it is worth investigating how well individual sub-features or different combinations of sub-features can predict the ground truth image quality. To this end, we simply replace UWEQM metric's UWIQF vector by a single sub-feature or a combined vector of multiple sub-features; and re-run the experiments as described above. Table III illustrates the performance of these UWEQM variants. It can be seen that in general the variant models based on multiple sub-features tend to give better performance than the variant models based on single sub-feature. Our final metric (i.e., proposed UWEQM) based on all sub-features yields the best performance.

D. Performance Comparison to State-of-The-Art IQA Metrics

We compare the proposed UWEQM metric to the state-of-the-art no-reference (NR) IQA metrics, including general-propose metrics (i.e., BRISQUE, NIQE, BLIINDS II, SSEQ and LPC) and underwater-specific metrics (i.e., UICM, UISM, UIConM, UIQM, UCIQE and UEIQM). For metrics that are not machine learning-based, the results are produced by calculating a metric on the entire UEIQA database. The learning-based metrics (i.e., BRISQUE, BLIINDS II, SSEQ, UEIQM) are evaluated using the 5-fold cross-validation method as described in Section IV.A. Fig. 7 shows the scatter plot of MOS versus the predictions of our proposed UWEQM metric.

The results of metric performance in terms of PLCC, SROCC, KROCC and RMSE are listed in Table IV, with the best performance for each evaluation criterion highlighted in boldface.

It can be seen that most general-purpose metrics, BRISQUE, NIQE, BLIINDS II and SSEQ fail in predicting the output quality of underwater image enhancement. These metrics are based on natural scene statistics and do not contain visual descriptors for the characterises of underwater environment, such as light attenuation, light scattering, and specific artifacts. LPC metric addresses the sharpness measure using multi-scale local phase coherence in the wavelet transform domain. Since sharpness/contrast is an important image quality aspect of underwater images, the LPC metric produces relatively better performance than other general-purpose metrics. For the underwater-specific metrics, UISM, UIQM, UCIQE metrics are unable to accurately predict the quality of underwater images undergoing enhancement algorithms. These models contain underwater-specific visual features; however, they simply combine these features and, hence cannot deal with the complex interactive relationships between these features for the determination of the overall image quality. The other metrics, UICM, UIConM and UEIQM adopt a more sophisticated way for integrating IQA-aware features, and therefore significantly outperform the UISM, UIQM, UCIQE metrics. In particular, it should be noted that the UEIQM and UIConM metrics give the best and second-best performance among the exiting metrics. This might be due to the fact that both metrics contain a good measure of contrast under low-luminance conditions, which is one of the most relevant features for underwater image quality assessment. This tends to suggest that the development of reliable underwater IQA metrics should focus on identifying and quantifying relevant underwater-specific image features as well as finding ways to express the complex relationships between these features in determining the perceived quality. Our proposed metric follows such concept and hence outperforms all existing metrics in predicting output quality of underwater image enhancement.

The F-test (as used in [65]) is adopted to test the statistical significance between each existing metric and our proposed UWEQM metric. The value of the F-test is the ratio of the standard deviations of normalized results (normalized by the non-linear logistic regression model) of the selected metric and our proposed metric. The F-test critical region is defined by the F-distribution with the (N-1,N-1) freedom degree and the significance level of 0.05, where N is the number of the test images. The two metrics in question are statistically distinct if the F-test value is in the critical region. The pairwise (i.e., a selected metric versus our proposed metric) significance testing results show that the proposed UWEQM is significantly superior to any other metric in comparison.

V. Discussion

A current trend in IQA research is to develop deep learning-based IQA metrics. However, the challenge lies in the fact that deep learning-based models heavily rely on large-scale annotated data, and that creating "large" IQA databases is nontrivial as reliable image quality scores must be derived from fully-controlled psychophysical experiments [66]. Some approaches have been attempted to exploit deep learning techniques in IQA metrics [67]–[72], e.g., data augmentation methods are applied to improve sample efficiency, and transfer learning and domain adaptation are adopted to boost learning ability. We compare the performance of our proposed UWEQM metric to the state-of-the-art deep learning-based IQA metrics (note only NR IQA metrics that have their open source code made publicly available are included to ensure a fair comparison), including VGG [67], CNN [70], BIECON [68], DIQaM [69], RankIQA [71], GraphIQA [72]. It can be seen from Table V that our proposed UWEQM gives the best performance, which demonstrates the effectiveness of the proposed approach that takes into account both underwater physics and human visual system attributes. This, however, does not mean deep learning is not a good alternative approach. To facilitate the development of deep learning-based IQA, more psychophysical experiments should be conducted to provide more ground truth IQA data, and advanced methods should be developed to enhance learning capabilities of the model. Our future work will focus on developing a deep learning-based model for assessing the output quality of underwater image enhancement.

The generalization ability is critical for an IQA metric. First, in the development of our proposed UWEQM metric, we consider the attributes that are general representations for the output quality of underwater image enhancement. These attributes include underwater-specific physics and human visual system properties for contrast, texture, color and attention. Second, in training our metric on the UEIQA database, the k-fold cross-validation is used to tune model parameters and ensure there is no data leakage. The training and test subsets per run (i.e., one 5-fold trial) do not overlap, and the data splitting is randomly iterated 100 times (i.e., 100 5-fold trials) to eliminate the performance bias. Third, it would be beneficial to conduct cross-database evaluation to measure the generalization capability of our metric. However, to the best of our knowledge, the UEIQA is so far the only database that is made publicly available for the assessment of output quality of underwater image enhancement. It should be also noted that the UEIQA database contains results of five image enhancement algorithms. Due to the innate limitation of psychophysical experimentation, increasing the number of enhancement algorithms would make an IQA database too large hence compromise the reliability of the subjective data [73]. Therefore, to facilitate cross-database evaluation, we encourage researchers to create new IQA databases with different underwater image enhancement algorithms following the same protocols for the subjective testing of the UEIQA database. Nonetheless, we expect our proposed model to generalise and adapt to new unseen data. To demonstrate this in a nutshell, we applied fourteen image enhancement algorithms that are not used in the UEIQA database to an original underwater image, including Bi-HE (note a different algorithm to the HE used in the UEIQA database) [74], CLAHE [75],, Gamma Correction (GC) [76], Integrated Colour Model (ICM) [77], Unsupervised Colour Correction Method (UCM) [78], Rayleigh Distribution (RD) [79], HVS-based multi-scale underwater image fusion (HVSMIF) [80], DCP [17], UDCP [18], Green-Blue channels Underwater Dark Channel Prior (GB-UDCP) [24], IBLA [14], Relative Global Histogram Stretching (RGHS) [81], Removal of Water Scattering (RoWS) [82], and ULAP [52]. Fig. 8 illustrates

TABLE V
PERFORMANCE COMPARISON OF THE PROPOSED UWEQM METRIC TO THE STATE-OF-THE-ART DEEP LEARNING-BASED NR IQA METRICS. RESULTS ARE BASED ON THE UEIQA DATABASE

Criteria \ Model	VGG	CNN	BIECON	DIQaM	RankIQA	GraphIQA	UWEQM
PLCC	0.7721	0.1693	0.0628	0.3761	0.8040	0.7842	**0.8922**
SROCC	0.7763	0.1411	0.0362	0.2415	0.7976	0.7876	**0.8816**
KROCC	0.6001	0.0963	0.0282	0.1673	0.6184	0.5993	**0.7137**
RMSE	1.3683	2.1234	2.2121	1.9963	1.2803	1.3352	**1.0534**

the results of these image enhancement algorithms. In terms of the output quality, subjective quality could be visually assessed by the readers (note a proper subjective IQA experiment could be treated in a separate contribution in the future), and objective quality is predicted by our UWEQM metric as shown in Fig 8. It can be seen that our proposed IQA metric shows a good generalization ability in assessing the output quality of underwater image enhancement.

VI. CONCLUSION

In this paper, we have presented an objective metric UWEQM for the assessment of the output quality of underwater images undergoing enhancement algorithms. Since there is no reference image of "perfect" quality in the context of image enhancement, we propose to consider the physics prior information of underwater optical imaging and relevant characteristics of the human visual system including contrast, texture, attention and color. The proposed metric UWEQM is validated against the ground truth image quality scores and it outperforms other alternative metrics in the literature. Moreover, our proposed UWEQM metric shows a good generalization ability for image quality assessment.

REFERENCES

[1] Y. Y. Schechner and N. Karpel, "Recovery of underwater visibility and structure by polarization analysis," *IEEE J. Ocean. Eng.*, vol. 30, no. 3, pp. 570–587, Jul. 2005.

[2] Z. Li and J. Zheng, "Edge-preserving decomposition-based single image haze removal," *IEEE Trans. image Process.*, vol. 24, no. 12, pp. 5432–5441, Dec. 2015.

[3] J. Han et al., "Resolution enhancement in active underwater polarization imaging with modulation transfer function analysis," *Appl. Opt.*, vol. 54, no. 11, pp. 3294–3302, 2015.

[4] 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, 2015.

[5] H. Lu et al., "Underwater image enhancement method using weighted guided trigonometric filtering and artificial light correction," *J. Vis. Commun. Image Representation*, vol. 38, pp. 504–516, 2016.

[6] G. N. Bailey and N. C. Flemming, "Archaeology of the continental shelf: Marine resources, submerged landscapes and underwater archaeology," *Quaternary Sci. Rev.*, vol. 27, no. 23–24, pp. 2153–2165, 2008.

[7] C. Alippi, R. Camplani, C. Galperti, and M. Roveri, "A robust, adaptive, solar-powered wsn framework for aquatic environmental monitoring," *IEEE Sensors J.*, vol. 11, no. 1, pp. 45–55, Jan. 2011.

[8] J.-O. Gaudron, F. Surre, T. Sun, and K. Grattan, "Long period grating-based optical fibre sensor for the underwater detection of acoustic waves," *Sensors Actuators A: Phys.*, vol. 201, pp. 289–293, 2013.

[9] J. Ling et al., "On bayesian channel estimation and fft-based symbol detection in mimo underwater acoustic communications," *IEEE J. Ocean. Eng.*, vol. 39, no. 1, pp. 59–73, Jan. 2014.

[10] Z. Yan et al., "A gravity gradient differential ratio method for underwater object detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 4, pp. 833–837, Apr. 2014.

[11] C. P. Chen, H. Li, Y. Wei, T. Xia, and Y. Y. Tang, "A local contrast method for small infrared target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 574–581, Jan. 2014.

[12] S. Q. Duntley, "Light in the sea," *J. Opt. Soc. Amer. A*, vol. 53, no. 2, pp. 214–233, 1963.

[13] S. Raimondo and C. Silvia, "Underwater image processing: State of the art of restoration and image enhancement methods," *EURASIP J. Adv. Signal Process.*, vol. 2010, no. 1, 2010, Art. no. 746052.

[14] Y.-T. Peng and P. C. Cosman, "Underwater image restoration based on image blurriness and light absorption," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1579–1594, Apr. 2017.

[15] S. Luria and J. A. S. Kinney, "Underwater vision," *Science*, vol. 167, pp. 1454–1461, 1970.

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

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

[18] H. Yang, P. Chen, C. Huang, Y. Zhuang, and Y. Shiau, "Low complexity underwater image enhancement based on dark channel prior," in *Proc. 2nd Int. Conf. Innovations Bio-inspired Comput. Appl.*, 2011, pp. 17–20.

[19] M. Zhao, C. Zhang, W. Zhang, W. Li, and J. Zhang, "Decorrelation-stretch based cloud detection for total sky images," in *Proc. Vis. Commun. Image Process.*, 2015, pp. 1–4.

[20] C. O. Ancuti and C. Ancuti, "Single image dehazing by multi-scale fusion," *IEEE Trans. Image Process.*, vol. 22, no. 8, pp. 3271–3282, Aug. 2013.

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

[22] Y. Peng, K. Cao, and P. C. Cosman, "Generalization of the dark channel prior for single image restoration," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2856–2868, Jun. 2018.

[23] Y. Wang et al., "An experimental-based review of image enhancement and image restoration methods for underwater imaging," *IEEE Access*, vol. 7, pp. 140233–140251, 2019.

[24] C. Li, J. Quo, Y. Pang, S. Chen, and J. Wang, "Single underwater image restoration by blue-green channels dehazing and red channel correction," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2016, pp. 1731–1735.

[25] Z. Chen, T. Jiang, and Y. Tian, "Quality assessment for comparing image enhancement algorithms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 3003–3010.

[26] J. Yan, K. Zhu, W. Zhang, J. Wang, and Y. Xiao, "Enhanced image quality assessment based on the joint similarity feature," *J. Imag. Sci. Technol.*, vol. 61, no. 5, pp. 50501-1–50501-9, 2017.

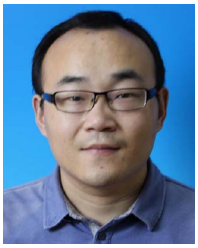
[27] P. Guo, L. He, S. Liu, D. Zeng, and H. Liu, "Underwater image quality assessment: Subjective and objective methods," *IEEE Trans. Multimedia*, vol. 24, pp. 1980–1989, 2022, doi: 10.1109/TMM.2021.3074825.

[28] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, Jan. 2009.

[29] Z. Wang et al., "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[30] H. Hadizadeh and I. V. Bajić, "Full-reference objective quality assessment of tone-mapped images," *IEEE Trans. Multimedia*, vol. 20, no. 2, pp. 392–404, Feb. 2018.
[31] Z. Wan, K. Gu, and D. Zhao, "Reduced reference stereoscopic image quality assessment using sparse representation and natural scene statistics," *IEEE Trans. Multimedia*, vol. 22, no. 8, pp. 2024–2037, Aug. 2020.
[32] W. Chen, K. Gu, T. Zhao, G. Jiang, and P. Le Callet, "Semi-reference sonar image quality assessment based on task and visual perception," *IEEE Trans. Multimedia*, vol. 23, pp. 1008–1020, Apr. 2021, doi: 10.1109/TMM.2020.2991546.
[33] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.
[34] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.
[35] A. Mittal, A. K. Moorthy, and A. C. Bovik, "Blind/referenceless image spatial quality evaluator," in *Proc. IEEE Conf. Rec. forty 5th Asilomar Conf. Signals, Syst. Comput.*, 2011, pp. 723–727.
[36] K. Gu et al., "Blind quality assessment of tone-mapped images via analysis of information, naturalness, and structure," *IEEE Trans. Multimedia*, vol. 18, no. 3, pp. 432–443, Mar. 2016.
[37] L. Li, W. Xia, W. Lin, Y. Fang, and S. Wang, "No-reference and robust image sharpness evaluation based on multiscale spatial and spectral features," *IEEE Trans. Multimedia*, vol. 19, no. 5, pp. 1030–1040, May 2017.
[38] Y. Liu, K. Gu, S. Wang, D. Zhao, and W. Gao, "Blind quality assessment of camera images based on low-level and high-level statistical features," *IEEE Trans. Multimedia*, vol. 21, no. 1, pp. 135–146, Jan. 2019.
[39] M. Yang and A. Sowmya, "New image quality evaluation metric for underwater video," *IEEE Signal Process. Lett.*, vol. 21, no. 10, pp. 1215–1219, Oct. 2014.
[40] 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.
[41] Z. Miao, F. Yuan, C. GAO, and E. Cheng, "New non-reference image quality evaluation method for underwater turbulence blurred images," in *Proc. 13th ACM Int. Conf. Underwater Netw. Syst.*, 2018, Art. no 19.
[42] S. G. Narasimhan and S. K. Nayar, "Chromatic framework for vision in bad weather," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2000, vol. 1, pp. 598–605.
[43] S. G. Narasimhan and S. K. Nayar, "Vision and the atmosphere," *Int. J. Comput. Vis.*, vol. 48, no. 3, pp. 233–254, 2002.
[44] R. Fattal, "Single image dehazing," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 1–9, 2008.
[45] D. Berman et al., "Non-local image dehazing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1674–1682.
[46] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.
[47] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the dct domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.
[48] L. Liu, B. Liu, H. Huang, and A. C. Bovik, "No-reference image quality assessment based on spatial and spectral entropies," *Signal Process.: Image Commun.*, vol. 29, no. 8, pp. 856–863, 2014.
[49] R. Hassen, Z. Wang, and M. M. A. Salama, "Image sharpness assessment based on local phase coherence," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2798–2810, Jul. 2013.
[50] 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.
[51] N. Carlevaris-Bianco, A. Mohan, and R. M. Eustice, "Initial results in underwater single image dehazing," in *Proc. MTS/IEEE SEATTLE*, 2010, pp. 1–8.
[52] W. Song, Y. Wang, D. Huang, and D. Tjondronegoro, "A rapid scene depth estimation model based on underwater light attenuation prior for underwater image restoration," in *Proc. Adv. Multimedia Inform. Process.*, R. Hong, W.-H. Cheng, T. Yamasaki, M. Wang, and C.-W. Ngo, Eds., 2018, pp. 678–688.
[53] D. Akkaynak and T. Treibitz, "A revised underwater image formation model," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 6723–6732.
[54] K. Panetta, S. Agaian, Y. Zhou, and E. J. Wharton, "Parameterized logarithmic framework for image enhancement," *IEEE Trans. Systems, Man, Cybernetics, Part B*, vol. 41, no. 2, pp. 460–473, Apr. 2011.
[55] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
[56] P. G. Freitas, S. Alamgeer, W. Y. Akamine, and M. C. Farias, "Blind image quality assessment based on multiscale salient local binary patterns," in *Proc. 9th ACM Multimedia Syst. Conf.*, 2018, pp. 52–63.
[57] A. Ninassi, O. Le Meur, P. Le Callet, and D. Barba, "Does where you gaze on an image affect your perception of quality? Applying visual attention to image quality metric," in *Proc. IEEE Int. Conf. Image Process.*, 2007, vol. 2, pp. II–169–II–172.
[58] H. Liu and I. Heynderickx, "Visual attention in objective image quality assessment: Based on eye-tracking data," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 7, pp. 971–982, Jul. 2011.
[59] W. Zhang, A. Borji, Z. Wang, P. Le Callet, and H. Liu, "The application of visual saliency models in objective image quality assessment: A statistical evaluation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 6, pp. 1266–1278, Jun. 2016.
[60] W. Zhang and H. Liu, "Toward a reliable collection of eye-tracking data for image quality research: Challenges, solutions, and applications," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2424–2437, May 2017.
[61] J. Huang, S. Kumar, M. Mitra, W.-J. Zhu, and R. Zabih, "Image indexing using color correlograms," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 1997, pp. 762–768.
[62] S. Shrivastava, B. Gupta, and M. Gupta, "Optimization of image retrieval by using hsv color space, zernike moment amp; dwt technique," in *Proc. IEEE Int. Conf. Comput. Intell. Comput. Res.*, 2015, pp. 1–5.
[63] A. Liaw et al., "Classification and regression by randomforest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.
[64] A. Krogh and J. Vedelsby, "Neural network ensembles, cross validation, and active learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 1995, pp. 231–238.
[65] R. Zhu, F. Zhou, W. Yang, and J. Xue, "On hypothesis testing for comparing image quality assessment metrics [tips tricks]," *IEEE Signal Process. Mag.*, vol. 35, no. 4, pp. 133–136, Jul. 2018.
[66] K. Ma et al., "Group maximum differentiation competition: Model comparison with few samples," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 851–864, Apr. 2020.
[67] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
[68] J. Kim and S. Lee, "Fully deep blind image quality predictor," *IEEE J. Sel. Top. Signal Process.*, vol. 11, no. 1, pp. 206–220, Feb. 2017.
[69] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 206–219, Jan. 2018.
[70] D. Chen, Y. Wang, and W. Gao, "No-reference image quality assessment: An attention driven approach," *IEEE Trans. Image Process.*, vol. 29, pp. 6496–6506, 2020, doi: 10.1109/TIP.2020.2990342.
[71] X. Liu, J. Van De Weijer, and A. D. Bagdanov, "Rankiqa: Learning from rankings for no-reference image quality assessment," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 1040–1049.
[72] S. Sun, T. Yu, J. Xu, W. Zhou, and Z. Chen, "GraphiQA: learning distortion graph representations for blind image quality assessment," *IEEE Trans. Multimedia*, 2021, doi: 10.1109/TMM.2022.3152942.
[73] L. Lévêque et al., "Cuid: A new study of perceived image quality and its subjective assessment," in *Proc. IEEE Int. Conf. Image Process.*, 2020, pp. 116–120.
[74] Y.-T. Kim, "Contrast enhancement using brightness preserving bi-histogram equalization," *IEEE Trans. Consum. Electron.*, vol. 43, no. 1, pp. 1–8, Feb. 1997.
[75] J. Stark, "Adaptive image contrast enhancement using generalizations of histogram equalization," *IEEE Trans. Image Process.*, vol. 9, no. 5, pp. 889–896, May 2000.
[76] S. Juric and V. Klepac, "Gamma and gamma correction in television production," in *Proc. Int. Symp. ELMAR*, 2009, pp. 83–86.
[77] K. Iqbal, R. A. Salam, A. Osman, and A. Z. Talib, "Underwater image enhancement using an integrated colour model," *IAENG Int. J. Comput. Sci.*, vol. 34, no. 2, pp. 239–244, 2007.
[78] 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.*, 2010, pp. 1703–1709.

[79] A. S. A. Ghani and N. A. M. Isa, "Underwater image quality enhancement through composition of dual-intensity images and rayleigh-stretching," in *Proc. IEEE Fourth Int. Conf. Consum. Electron. Berlin*, 2014, pp. 219–220.
[80] P. Guo et al., "Multi-scale enhancement fusion for underwater sea cucumber images based on human visual system modelling," *Comput. Electron. Agriculture*, vol. 175, 2020, Art. no. 105608.
[81] D. Huang, Y. Wang, W. Song, J. Sequeira, and S. Mavromatis, "Shallow-water image enhancement using relative global histogram stretching based on adaptive parameter acquisition," in *MultiMedia Modeling*. Cham: Springer International Publishing, 2018, pp. 453–465.
[82] L. Chao and M. Wang, "Removal of water scattering," in *Proc. 2nd Int. Conf. Comput. Eng. Technol.*, 2010, vol. 2, pp. V2–35–V2–39.



Pengfei Guo received the Ph.D. degree from the South China University of Technology, Guangzhou, China, in 2015. He is currently an Associate Professor with the School of Computing Science, Zhongkai University of Agriculture and Engineering, Guangzhou, China. He was a Visiting Scholar of Cardiff University, South China University of Technology, and Sun Yat-sen University. He has been focusing his research in applied mathematics and its interdisciplinary application. His research interests include neural networks, robotics, control theory, optimizations, computer vision and image quality assessment.

Hantao Liu received the Ph.D. degree from the Delft University of Technology, Delft, The Netherlands in 2011. He is currently a Reader with the School of Computer Science and Informatics, Cardiff University, Cardiff, U.K. He is an Associate Editor of the IEEE Signal Processing Letters and the IEEE Transactions on Circuits and Systems for Video Technology. His research interests include machine learning, computer vision and image quality assessment.

Delu Zeng received his Ph.D. degree in electronic and information engineering from South China University of Technology, China, in 2009. He is now a Full Professor in the School of Mathematics in South China University of Technology, China. He has been the visiting scholar of Columbia University, University of Oulu, University of Waterloo. He has been focusing his research in applied mathematics and its interdisciplinary applications. His research interests include numerical calculations, applications of partial differential equations, optimizations, machine learning and their applications in image processing, and data analysis.

Tao Xiang (Senior Member, IEEE) received the B.Eng., M.S., and Ph.D. degrees in computer science from Chongqing University, China, in 2003, 2005, and 2008, respectively, where he is currently a Professor with the College of Computer Science. He has published over 100 articles on international journals and conferences. He also served as a referee for numerous international journals and conferences. His research interests include multimedia security, cloud security, data privacy, and cryptography.

Leida Li (Member, IEEE) received the B.S. and Ph.D. degrees from Xidian University, Xi'an, China, in 2004 and 2009, respectively. In 2008, he was a Research Assistant with the Department of Electronic Engineering, Kaohsiung University of Science and Technology, Taiwan. From 2014 to 2015, he was a Visiting Research Fellow with the Rapid-Rich Object Search (ROSE) Laboratory, School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, where he was a Senior Research Fellow from 2016 to 2017. From 2009 to 2019, he worked in the School of Information and Control Engineering, China University of Mining and Technology, as Assistant Professor, Associate Professor and Professor, respectively. Currently, he is a Professor with the School of Artificial Intelligence, Xidian University. His research interests include multimedia quality assessment, affective computing, information hiding, and image forensics. He has served as SPC for IJCAI 2019-2020, Session Chair for ICMR 2019 and PCM 2015, and TPC for AAAI 2019, ACM MM 2019-2020, ACM MM-Asia 2019, ACII 2019, PCM 2016. He is now an Associate Editor of the Journal of Visual Communication and Image Representation and the EURASIP Journal on Image and Video Processing.

Ke Gu (Member, IEEE) is currently a Professor at the Beijing University of Technology, Beijing, China. His research interests include environmental perception, image processing, quality assessment, and machine learning. He was the Leading Special Session Organizer in the VCIP 2016 and the ICIP 2017. He received the Best Paper Award from the IEEE TRANSACTIONS ON MULTIMEDIA and the Best Student Paper Award at the IEEE International Conference on Multimedia and Expo (ICME) in 2016. He is an Associate Editor for Computer Animation and Virtual Worlds (CAVW) and IET Image Processing (IET-IPR), an Area Editor for Signal Processing: Image Communication (SPIC), and an Editor for Applied Sciences, Displays, and Entropy. He is a Reviewer for 20 top SCI journals.