Signal Processing: Image Communication 94 (2021) 116218

Contents lists available at ScienceDirect

Signal Processing: Image Communication

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

A reference-free underwater image quality assessment metric in frequency domain

Ning Yang [a,b,1], Qihang Zhong [a,b,1], Kun Li [c], Runmin Cong [a,b,d,*], Yao Zhao [a,b], Sam Kwong [d]

[a] Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China
[b] Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing 100044, China
[c] School of Software Engineering, Beijing Jiaotong University, Beijing 100044, China
[d] Department of Computer Science, City University of Hong Kong, Hong Kong Special Administrative Region, China

ARTICLE INFO

Keywords:
Underwater image
Reference-free image quality assessment
Frequency domain
Dark channel prior weighting
New dataset

ABSTRACT

Owing to the complexity of the underwater environment and the limitations of imaging devices, the quality of underwater images varies differently, which may affect the practical applications in modern military, scientific research, and other fields. Thus, achieving subjective quality assessment to distinguish different qualities of underwater images has an important guiding role for subsequent tasks. In this paper, considering the underwater image degradation effect and human visual perception scheme, an effective reference-free underwater image quality assessment metric is designed by combining the colorfulness, contrast, and sharpness cues. Specifically, inspired by the different sensibility of humans to high-frequency and low-frequency information, we design a more comprehensive color measurement in spatial domain and frequency domain. In addition, for the low contrast caused by the backward scattering, we propose a dark channel prior weighted contrast measure to enhance the discrimination ability of the original contrast measurement. The sharpness measurement is used to evaluate the blur effect caused by the forward scattering of the underwater image. Finally, these three measurements are combined by the weighted summation, where the weighed coefficients are obtained by multiple linear regression. Moreover, we collect a large dataset for underwater image quality assessment for testing and evaluating different methods. Experiments on this dataset demonstrate the superior performance both qualitatively and quantitatively.

1. Introduction

The ocean covers 71% of the earth's surface and is an important part of the earth's resources, but its exploration is only 5%, and the development is less than 1%. As an important carrier and expression form of underwater information, underwater images play an irreplaceable and significant role in the underwater environment perception and marine monitoring. However, due to the influence of the special physical and chemical underwater environment, the quality of the obtained underwater images is a mixed bag, and the degradation such as color distortion, low contrast, blurry details, and poor clarity often occurs, which makes the underwater image "invisible", "inaccurate", and "incomplete", and seriously affects the interpretation of the image content. Therefore, in the underwater application system, how to effectively evaluate and predict the quality of different underwater images has important guiding significance for subsequent tasks, such as underwater image enhancement [1–4], detection [5], and classification [6].

The image quality assessment algorithm for in-air images has been investigated comprehensively and achieved good performance. In fact, the quality assessment for in-air images and underwater images both belong to the category of image quality assessment. Many generic image IQA techniques can also be appropriately extended to underwater IQA scenario, such as frequency domain analysis. However, due to the unique underwater shooting environment, the obtained underwater images have different attributes and characteristics compared with the in-air images, thus it is usually impossible to obtain reliable results by directly transplanting the general IQA method to the underwater image. As shown in Fig. 1, the right image is much clearer than the left image, and has a higher MOS value, but the results obtained by general IQA measurements (*i.e.* CODE [7] and EMBM [8]) are just the opposite. First, the water usually contains suspended solids and various active organisms, which causes the light reflected by the underwater scene to be absorbed and scattered by the particles suspended in the medium before reaching the camera, resulting in low contrast and foggy

* Corresponding author at: Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China.
E-mail addresses: ningyang@bjtu.edu.cn (N. Yang), 17281109@bjtu.edu.cn (Q. Zhong), 18301007@bjtu.edu.cn (K. Li), rmcong@bjtu.edu.cn (R. Cong), yzhao@bjtu.edu.cn (Y. Zhao), cssamk@cityu.edu.hk (S. Kwong).
URL: https://rmcong.github.io/ (R. Cong).
[1] Equal contributions.

https://doi.org/10.1016/j.image.2021.116218
Received 20 July 2020; Received in revised form 15 January 2021; Accepted 4 March 2021
Available online 17 March 2021

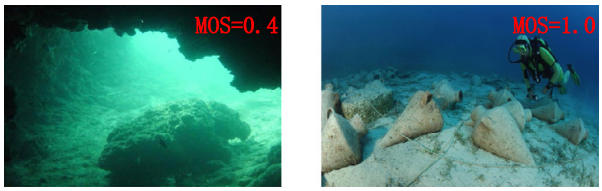

Fig. 1. The general IQA results of underwater images. (a) EMBM: 0.58; CODE: 5.18, (b) EMBM: 0.55; CODE: 4.84.

effects. Specifically, the backward scattering effect generally minifies the gray range and induces the low contrast of underwater image. The forward scattering effect generally leads to point spread phenomenon and underwater image blurring. Second, the attenuation of light depends on the wavelength of the light, the dissolved organic compounds and so on. The selective absorption of light will cause different degrees of color distortion. In general, blue and green lights have shorter wavelengths and higher frequencies, and thus their penetration ability is stronger. This is why underwater images usually exhibit blue and green. However, such degradation factors are not common in in-air images. With these in mind, we specialize in an image quality assessment method for underwater image in this paper. Starting from the degradation characterizations, the colorfulness, contrast, and sharpness can be used to evaluate the overall underwater image according to the different properties of the water medium. In this paper, considering the human visual perception scheme, we propose an effective reference-free underwater image quality assessment metric by combining these three attributes. Our main contributions lie in the upgradation and refinement of the colorfulness and contrast metrics.

For the colorfulness metric, the existing methods mainly evaluated the image quality in the spatial domain by calculating the standard deviation or the average value of chroma and saturation. In reality, the human visual system has different perception capabilities for the high-frequency and low-frequency components of an image. Inspired by this, we design a colorfulness metric in frequency domain to supplement the spatial domain metric, thereby obtaining a more comprehensive color evaluation measurement for underwater image. For the contrast metric, the existing methods used the gray-scale intensity of the image to calculate the contrast. However, the ability to distinguish different image qualities is limited. Considering the dark channel prior (DCP) can be used to remove the fogging effect of the input image, the underwater images with different qualities can be distinguished according to different DCP values. Thus, we designed a DCP weighed contrast measurement, utilizing the DCP value to refine the existing contrast metric and further enlarge its distinguishable range. In addition, the lack of datasets limits the development of this field to a certain extent. To this end, we conduct quality evaluation experiments on the Underwater Image Enhancement Benchmark (UIEB) [9] to obtain the corresponding quality ground truth, and form an Underwater Image Quality Assessment (UWIQA) dataset that can bridge the gap between the algorithm and data, thereby enriching the diversity of underwater image processing tasks.

All in all, the main contributions of this paper are summarized as follows:

- Inspired by the different sensibility of humans to high-frequency and low-frequency information, we design a color measurement in frequency domain to supplement the spatial domain metric.
- In order to improve the distinguishing ability of the contrast measurement, we propose a dark channel prior weighted contrast measure.
- An Underwater Image Quality Assessment (UWIQA) dataset is constructed by conducting the subjective evaluation experiments on the UIEB, which bridges the gap between the algorithm design and data validation.

The rest of this paper is organized as follows. The related work of underwater image quality assessment algorithms are introduced in Section 2. Then, the proposed evaluation metrics and the UWIQA dataset are presented in Sections 3 and 4, respectively. The experimental comparisons and analyses are conducted in Section 5. Finally, the conclusion is drawn in Section 6.

2. Related works

In this section, we will briefly summarize the image quality assessment (IQA) models and methods, and then introduce the existing IQA methods for underwater images.

Image quality assessment plays an important role in image processing, mainly by analyzing and studying the characteristics of the image to achieve the assessment of the image quality (*i.e.*, the degree of image distortion) [10]. According to the amount of information required from the original image, image quality assessment can be classified into full-reference (FR) IQA [10–13], no-reference (NR) IQA [14–16], and reduced-reference (RR) IQA [17–20]. FR-IQA models has both the distorted images and original images (*i.e.*, undistorted/reference image), and the quality score of the image is determined by directly comparing the information or feature similarity of the two images, which is a relatively mature research direction. However, in many practical scenarios, it is difficult to obtain the reference images, so more and more work is concentrated on RR-IQA and NR-IQA. RR-IQA models only require part of the features extracted from the reference image, while the NR-IQA models get rid of the constraints of the reference image, and only the distorted image is needed. As the most challenging issue in IQA, NR-IQA is closer to the actual situation and has gradually become a research hotspot in recent years. Deep learning technology has already demonstrated strong learning capabilities and performance advantages in many computer vision tasks, such as saliency detection [21–27], image enhancement [28–31], object tracking [32], *etc.* Recently, some neural networks have been successfully applied to image quality assessment [33–40].

However, due to different imaging principles, directly applying the in-air IQA [10,14,17,41–44] to underwater IQA issue often fails to achieve good performance. Therefore, some specially designed IQA models for underwater image are proposed based on the specific characteristics of underwater images. At the same time, for the underwater image, it is almost impossible to obtain reference images, so studying the NR-IQA methods has become the only way out. Miao et al. [45] proposed an underwater color image quality evaluation metric (UCIQE) based on standard deviation of chroma, luminance contrast, and average of saturation, which also can be applied to underwater video. Inspired by human visual system, Panetta et al. [46] proposed an underwater image quality method (UIQM) by considering the colorfulness, sharpness, and contrast, which is well related to perceived underwater image quality. Analyzing the underwater absorption and scattering characteristics, Wang et al. [47] proposed the underwater image quality assessment metric named CCF, which is based on colorfulness measure, contrast measure, and fog density measure.

3. Proposed method

In this section, an underwater image quality assessment metric is proposed by combining the colorfulness, contrast, and sharpness indexes. For the different representations of the underwater images degradation, we design three sub-measurements: (1) Enlightened by the different perception of humans to high-frequency and low-frequency components, we design a more comprehensive color measurement in both spatial domain and frequency domain to quantify the color distortion of underwater image caused by the absorption effect. (2) Based on the observation that different qualities of underwater images can be distinguished by the dark channel prior (DCP) values, we design a DCP weighed contrast measurement to evaluate the low contrast of

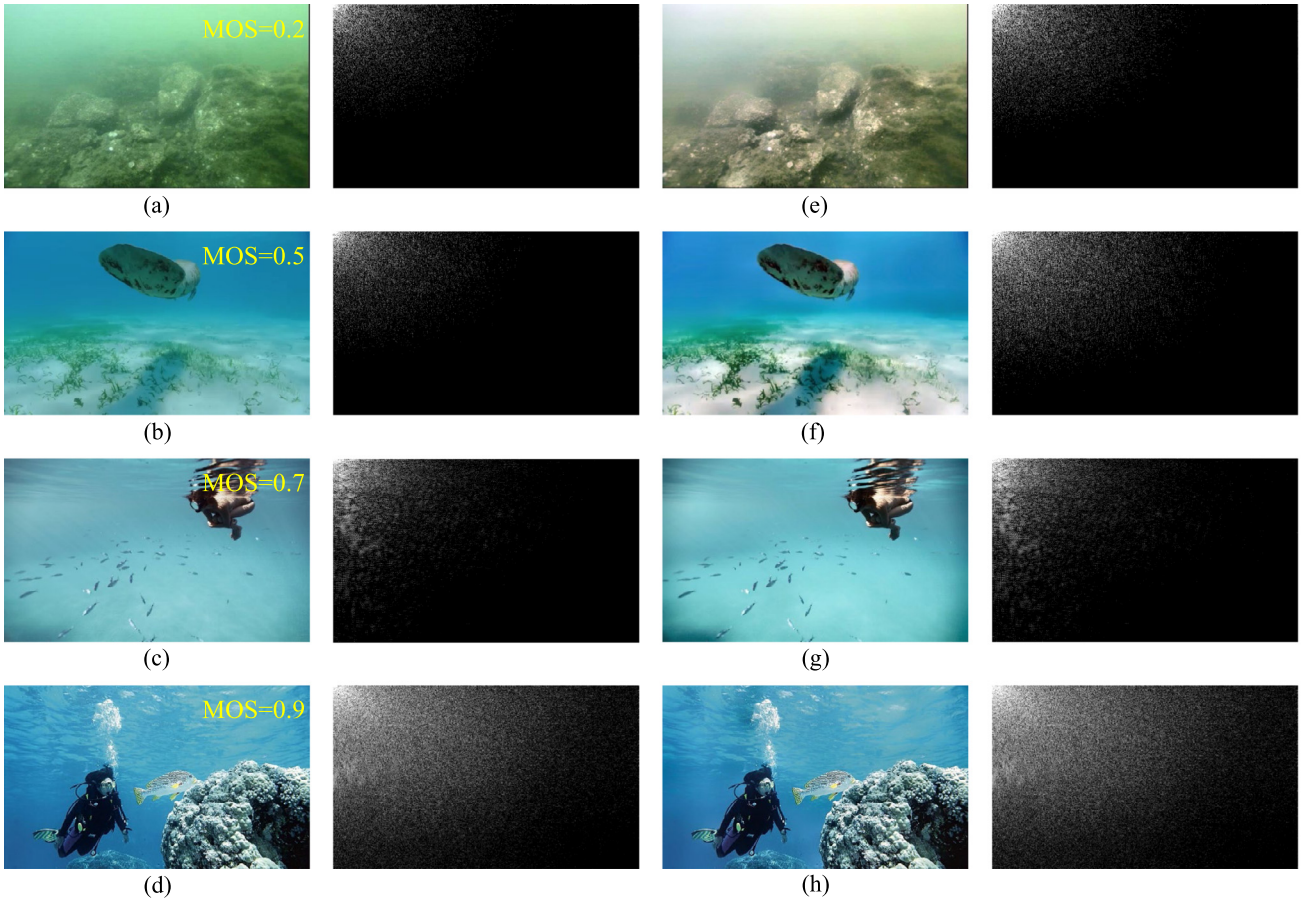

Fig. 2. The DCT coefficients of different underwater images with different qualities. (a)–(d) represent the original underwater images and their corresponding DCT coefficients. (e)–(h) represent the underwater image enhanced by [48] and the corresponding DCT coefficients.

underwater image caused by the forward scattering. (3) For the blurring effect of the underwater image caused by the backward scattering, we assess its quality using the sharpness measurement. Finally, these three indexes are weighted combined into a holistic metric, where the multiple linear regression (MLR) method is applied to obtain the weighted coefficients. We will provide more details in the following subsections.

3.1. Colorfulness measure

The selective absorption effect of light in the underwater medium will cause serious color deviation of the underwater images. Generally speaking, when quantizing the influence of colorfulness distortion, the traditional colorfulness measurements mainly analyze the image in the spatial domain. In fact, the human visual system has different perception capabilities for high-frequency and low-frequency components of images. Moreover, underwater images with different qualities should contain different high-frequency and low-frequency components. In other words, the recorded colorfulness changes are not only reflected in the spatial domain, but also in the frequency domain.

Through the discrete cosine transform (DCT), the underwater image can be transformed from the spatial domain to the frequency domain, in which the energy of the image is mainly concentrated in a certain local area, and the image characteristics can be described by a small amount of DCT coefficients. Theoretically, the low-frequency coefficients in the frequency domain reflect the flat area information in the image, mainly concentrated in the upper left corner of the frequency domain map, and the high-frequency coefficients usually describe the image boundary and texture information, corresponding to the other areas. For underwater images with different qualities, we calculate the corresponding DCT coefficients and show them in Fig. 2. By observing the DCT coefficients, we can see that the distribution of DCT coefficients of different quality images has obvious differences. For example, the DCT coefficients of poor quality underwater images (*e.g.*, the mean opinion score (MOS) of Fig. 2(a) is 0.2) are mainly concentrated in the upper left corner, while the other areas are almost zero. By contrast, the distribution range of DCT coefficients for better quality underwater images (*e.g.*, the MOS of Fig. 2(d) is 0.9) is obviously wider. Inspired by this, on the basis of colorfulness measurement in the spatial domain, we define a colorfulness measurement in the frequency domain by using the DCT coefficients.

In order to evaluate the colorfulness of underwater images of different qualities more comprehensively, the final colorfulness measure is defined as the product of colorfulness metric in frequency domain and the colorfulness metric in the spatial domain. The formula is as follows:

$$Colorfulness = Colorfulness_s \times Colorfulness_f, \tag{1}$$

where $Colorfulness_s$ and $Colorfulness_f$ are the colorfulness metric in the spatial domain and frequency domain, respectively. Similar to the UCIQE [45], the variance of chroma is used to define the colorfulness metric in spatial domain because it has a good correlation with human perception of underwater color images:

$$\begin{aligned} Colorfulness_s &= \sigma_{chr} \\ &= \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(I_{ij}-\overline{I})^2}, \end{aligned} \tag{2}$$

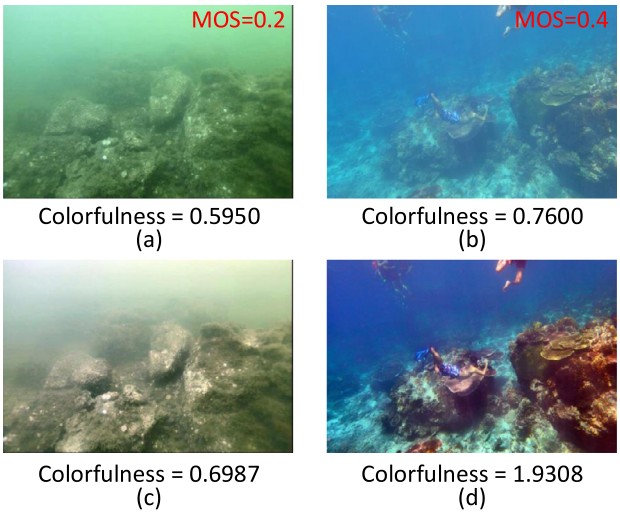

Fig. 3. (a)–(b) Original underwater images. (c)–(d) Underwater images enhanced by [48].

where σ_{chr} denotes the standard deviation of chroma, I_{ij} is the chroma component of the image in the CIELab space, $\overline{I}$ represents the mean value of the chroma component in the CIELab space, M and N indicate the height and width of the underwater image, respectively. For the colorfulness metric in frequency domain, we calculate the standard deviation of the DCT map, which is defined as:

$$Colorfulness_f = \sigma_{DCT} = \sqrt{\frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(DCT_{ij} - \overline{DCT})^2}{MN}}, \tag{3}$$

where DCT_{ij} represents the DCT coefficient at the position (i, j) of the input image, and $\overline{DCT}$ represents the mean value of the DCT map calculated by the discrete cosine transform of the input underwater image.

In Fig. 3, we provide some comparisons of different underwater images. For the different original images, Fig. 3(b) is visually better than the Fig. 3(a), and the consistent result is obtained by using our colorfulness metric. When comparing the image before and after enhanced by the method [48], the enhanced images have richer and more realistic colors than the original underwater images. Consistent with human subjective visual perception, through our proposed color measurement, the enhanced underwater image also obtains a higher score, which demonstrates the effectiveness of our proposed measurement.

3.2. Contrast measure

In addition to the color distortion of the underwater image caused by the selective absorption effect, the light scattering in the underwater medium can also cause the low contrast and blurring in the image. Backscattering is an optical effect caused by non-target scattered light that encounters suspended particles in water and randomly scattered at a small angle before reaching the camera lens. This causes the haze-like phenomenon in the underwater images, greatly lowers the scene contrast, and affects the image quality [49]. Therefore, the contrast measurement is also an important part of underwater image quality evaluation.

In [47], Wang et al. used the sum of contrast metric values from edge image blocks to represent the blurring of an underwater image and define the contrast measurement. First, the underwater image is divided into some blocks with the size of 64 × 64, and Sobel operator is used to generate the edge map. With the obtained edge map, we can judge whether the block belongs to the flat or edge block. Specifically, if the number of edge pixels in a block is greater than 0.2% of the total pixels, the block is selected as an edge block [50]. Then, the contrast measure of the underwater image is defined as the sum of root mean square (RMS) contrast values of all the edge blocks:

$$Contrast_{CCF} = \sum_{k=1}^{K}\sqrt{\frac{1}{XY}\sum_{i=1}^{X}\sum_{j=1}^{Y}\left(Ir_{ij}^{k} - \bar{I}r\right)^2}, \tag{4}$$

where Ir_{ij}^{k} is the red intensity value of the pixels in the kth edge block with size X by Y, K is the number of edge blocks, and $\bar{I}r$ represents the average red intensity of pixels in the underwater image.

However, directly using the contrast measure in [47] is not good for distinguishing images of different qualities. For example, the MOSs of Fig. 4(a) and (c) are 0.1 and 0.4, respectively, but the measured contrast value [47] is 0.0315 and 0.0211 respectively. Thus, in this paper, we aim at refining the contrast measure and enlarging the distinguishable range. He et al. [50] proposed a defogging algorithm based on the observation that some pixels always have at least one color channel with a very low value in the vast majority of non-sky local areas. This prior is named as dark channel prior (DCP). Thus, the DCP map of a fog-free image should have a lower intensity value, while fog image should have a significantly higher intensity. In fact, the underwater image and fog image have certain similarities, and both can be modeled by the atmospheric scattering model. Therefore, underwater images with different qualities should also have different dark channel prior maps. In Fig. 4, we provide some examples of DCP maps under different quality underwater images. Fig. 4(a) has a poor quality (MOS = 0.1), so its DCP map has a high intensity. However, Fig. 4(d) is relatively clear (MOS = 1), and its DCP map has a low intensity. In summary, for underwater images with different contrast degrees, the greater the contrast, the darker the DCP map. Thus, we design a DCP weighted contrast measurement to refine the existing CCF contrast metric.

First, the DCP method [50] is applied to the input underwater image and generate the DCP map I^{dark}.

$$I^{dark}(x) = \min_{y\in\Psi(x)}\left(\min_{c\in(r,g,b)} I^{c}(y)\right), \tag{5}$$

where I^{c} represents each channel of the underwater image, $c \in r, g, b$, $\Psi(x)$ denotes a window centered on the pixel x, and the window width is set to 15.

Then, considering that the intensity of the DCP map is inversely proportional to the contrast of the underwater image, we calculate the DCP weighting coefficient through a exp function of the mean value of the DCP map, which is defined as:

$$W_{DCP} = \exp\left(-\frac{I^{dark}}{\sigma^2}\right), \tag{6}$$

where σ is a parameter, which is fixed to 10.

Finally, the overall contrast measurement weighted by the DCP coefficient is demonstrated in

$$Contrast = W_{DCP} * Contrast_{CCF}. \tag{7}$$

The contrast measurement values of different underwater images before and after enhancement are shown in Fig. 5. The used image enhancement algorithm [48] can well eliminate the influence of underwater scattering. As visible, the enhanced underwater image has better contrast than the original image. For underwater images with different qualities, the better the image quality, the greater the contrast measurement score and the better the discrimination. Specifically, in Fig. 5, the MOS values of the original images (a)–(b) are 0.4 and 0.2, respectively. Before DCP weighting, their contrast values are 0.0686 and 0.0719, respectively. The contrast measure without DCP weighting has unconspicuous discrimination. After DCP weighting, the contrast values update to 0.0497 and 0.0336, respectively. For these two underwater images with different MOSs, the discriminability of the weighted contrast measure is improved, which is closer to the real quality difference. This also illustrates the effectiveness of our DCP weighting design.

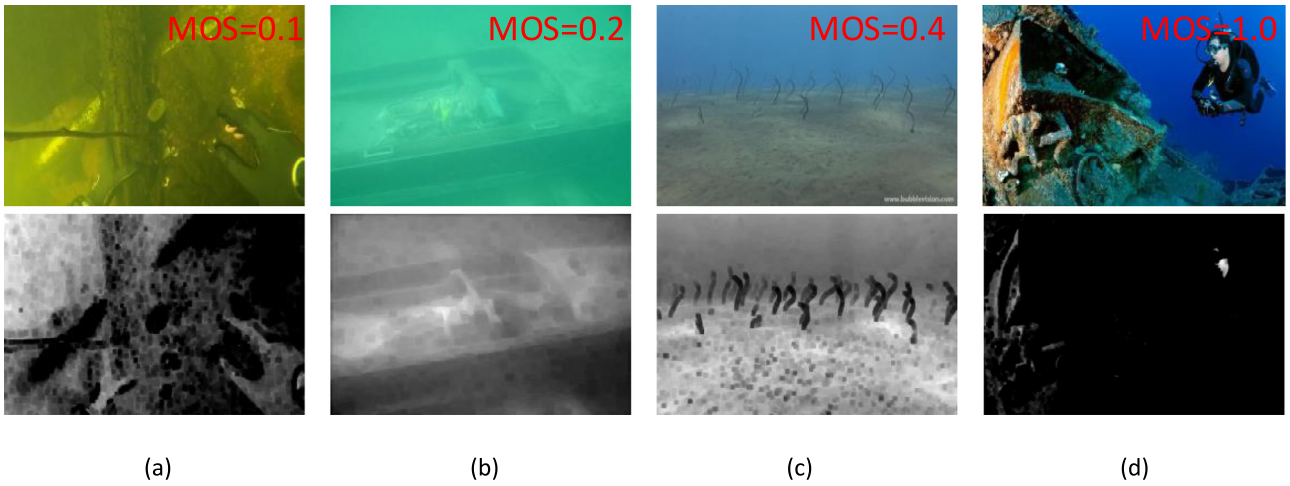

Fig. 4. In (a)–(d), the first row is the underwater images of different quality, and the second row is the corresponding DCP map [50].

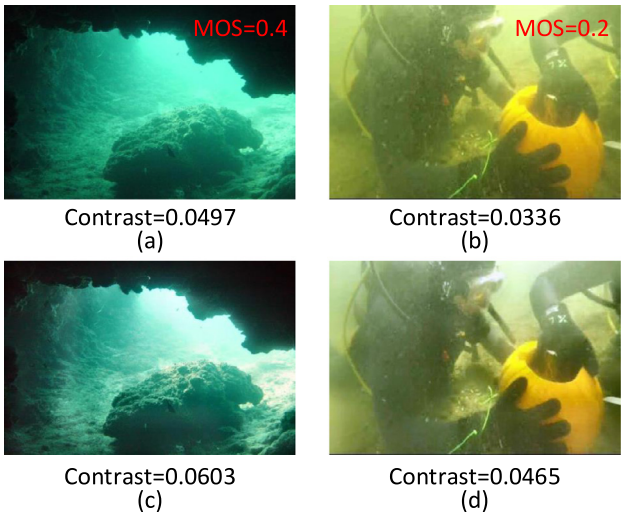

Fig. 5. (a)–(b) Original underwater images. (c)–(d) Underwater images enhanced by [48].

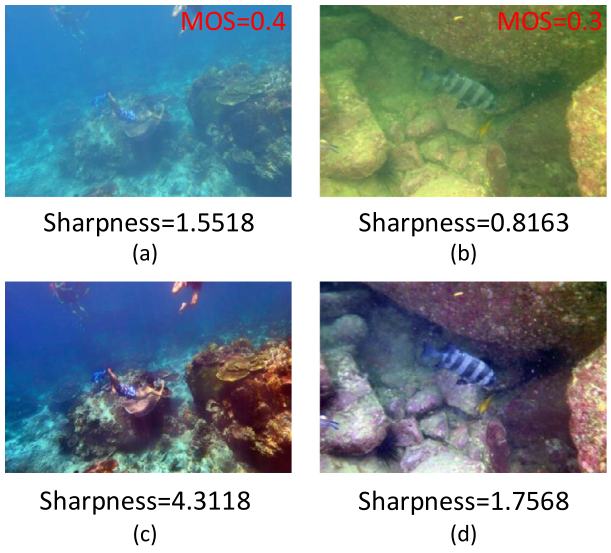

Fig. 6. (a)–(b) Original underwater images. (c)–(d) Underwater images enhanced by [48].

3.3. Sharpness measure

In the influence of the underwater scattering effect, forward scattering makes the underwater image severely blurred, thereby reducing the clarity of the image and losing some details and edges of the image. Therefore, sharpness describes changes in the details of the image, which is a more important parameter in the evaluation of underwater image quality. Similar to [51], the enhancement measure estimation (EME) measure is used to measure the sharpness of the grayscale edge map, which is defined as:

$$Sharpness = \sum_{c=1}^{3} \lambda_c EME(grayscale\ edge_c), \tag{8}$$

$$EME = \frac{2}{k_1 k_2} \sum_{l=1}^{k_1} \sum_{k=1}^{k_2} log(\frac{I_{max,k,l}}{I_{min,k,l}}) \tag{9}$$

where the underwater image is divided into $k_1 k_2$ blocks, $(I_{max,k,l})/(I_{min,k,l})$ represents the relative contrast of each block, and the EME measures in the three components of RGB color are linearly combined with its corresponding coefficient λ_c, $\lambda_R = 0.299$, $\lambda_G = 0.587$, and $\lambda_B = 0.114$ are used in conjunction with the relative visual response of the red, green and blue channels.

We compare the sharpness measurement of different underwater images before and after enhancement in Fig. 6. As can be seen from Fig. 6, the enhanced underwater image has better clarity than the original image, and the enhanced image obtain a larger sharpness score.

3.4. FDUM measure

Underwater images are affected by absorption and scattering in the underwater environment, resulting in different image qualities. Furthermore, the underwater image can be modeled as a linear correlation between the absorption and scattering parts, where the absorption effect corresponds to colorfulness part, the backward scattering corresponds to the contrast part, and the forward scattering corresponds to the sharpness part. Therefore, the colorfulness, contrast, and the sharpness measures are fused by weighted summation to obtain the final FDUM metric. The formula is as follows:

$$FDUM = \omega_1 \times Colorfulness + \omega_2 \times Contrast + \omega_3 \times Sharpness, \tag{10}$$

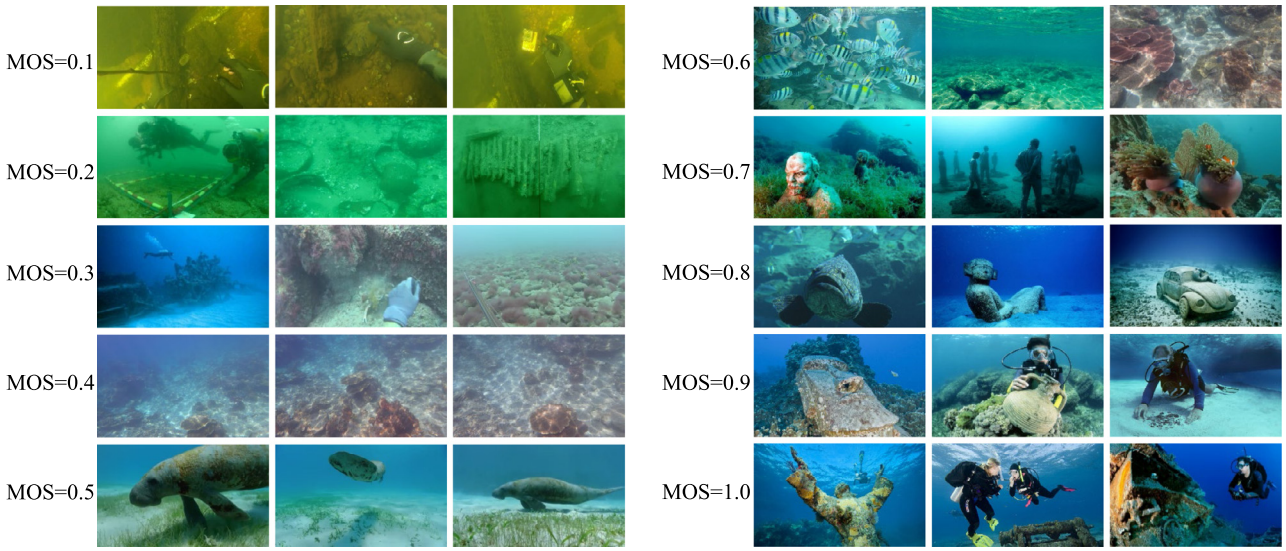

Fig. 7. Visual examples in different MOS groups in the UWIQA dataset.

Table 1
Distribution of observers.

Attributes		Number
Gender	Male	11
	Female	10
Age range	18–25	15
	25–30	6

where ω_1, ω_2 and ω_3 are the weighted coefficients corresponding to the colorfulness, contrast, and sharpness measurements, respectively. In our method, we use the multiple linear regression (MLR) to obtain the coefficients, which is fixed as [0.2982, 0.4439, 0.028].

4. The UWIQA dataset

Existing underwater image quality evaluation methods only conduct performance testing on a small number of underwater images. The finite underwater image quality evaluation benchmark dataset limits the development of this direction to a certain extent, especially for the deep learning based methods. Based on this, we annotate a large-scale underwater image quality evaluation benchmark dataset, named UWIQA dataset, which lays the data basis for subsequent algorithm research and performance evaluation. Specifically, we directly use Underwater Image Enhancement Benchmark (UIEB) [9] including 890 underwater images as our original images, and obtain the mean opinion score (MOS) for each image as the ground truth quality value. In order to obtain a more reasonable and effective MOS, we invited 21 observers to rate the quality of the underwater image dataset, where the observers include different professional backgrounds, different age ranges, different genders, and so on. In Table 1, we show the distribution statistics of all observers. The ratio of male to female is close to 1:1, and the age of observers is mainly concentrated in the people of 18–25 years old with relevant professional backgrounds.

In order to reduce the evaluation differences caused by different display devices, the evaluation process uses a unified high-definition device to display, and restricts the observer to look at the screen at eye level, as far as possible to avoid the impact of different external factors. In order to avoid the randomness of the evaluation process, the observers were required to perform three-times evaluations. The images were displayed from the beginning at the first time, but the observer does not need to score, only to get an overall understanding of the whole dataset. At the second and third times, the observer was asked to rate each image in the dataset, with scores from 1 to 5 containing 5 integer score levels, respectively corresponding to the poor quality to the good quality. Finally, the average value of the two evaluations was taken as the original score of the observer.

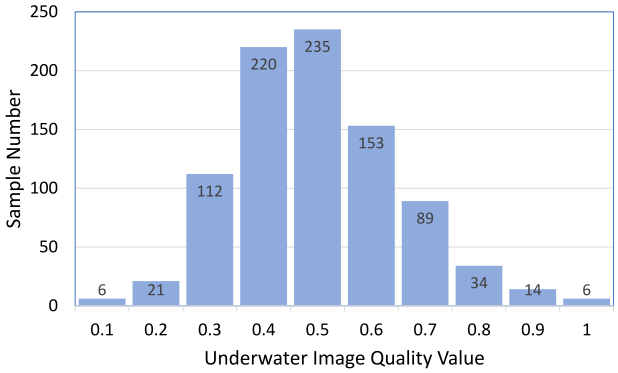

Fig. 8. MOS distribution statistics of the UWIQA dataset.

After obtaining the original scores of all reviewers, we need to clean the data. First, in order to ensure the validity of the scoring data, the outlier data needs to be deleted based on the review average of all reviewers. During this process, seven evaluation results were discarded because the large differences with others. Then, the effective scores are averaged and mapped to a total of 10 score levels from 0.1 to 1 to obtain the final MOS (*i.e.*, ground truth quality score) of each image. Fig. 7 provides some visual examples in different MOS groups of the UWIQA dataset. From it, we can see that the underwater image quality in each group has a higher consistency.

In Fig. 8, we provide the MOS distribution statistics of the UWIQA dataset. In this dataset, the number of images with a MOS score of 0.5 is the largest, accounting for 26.4% of the overall, followed by images with a MOS score of 0.4, accounting for 24.7%. By contrast, good quality images (MOS scores over 0.7) only account for 6.1%. The average MOS score of this dataset is 0.499.

5. Experiments

In this section, the experiments on the UWIQA dataset are classified into two parts. The first part is the objective quantitative performance evaluation, including the comparisons between objective evaluation

Table 2
Quantitative results of different methods on the UWIQA dataset.

	PRCC	KRCC	SRCC
FDUM	**0.638**	**0.530**	**0.683**
UIQM [45]	0.608	0.473	0.618
UCIQE [46]	0.595	0.474	0.622
CCF [47]	0.409	0.351	0.479
CODE [7]	0.545	0.435	0.575
EMBM [8]	0.161	0.073	0.100

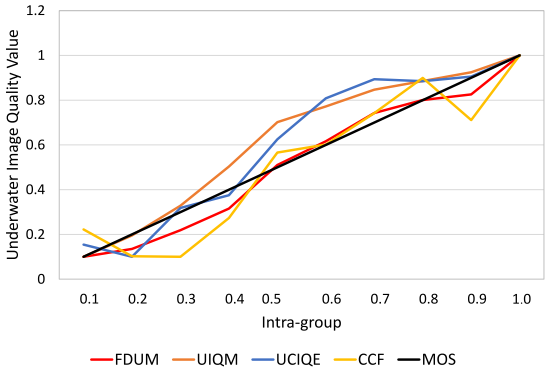

Fig. 9. The curves of the intra-group consistency.

results obtained by different methods and MOS, and the intra-group consistency comparisons. The second part is used to evaluate the effectiveness by comparing different enhancement methods. The experiments are conducted on 1.6 GHz frequency Intel i5 CPU and 8 GB of RAM using Matlab 2017a, and the average running time for one image is 3.85 s.

5.1. Objective quantitative performance evaluation

For the quantitative evaluation, the Pearson Product-moment Correlation Coefficient (PRCC), Kendall Rank Correlation Coefficient (KRCC), and Spearman Rank Correlation Coefficient (SRCC) are used to evaluate the correlation between the predicted quality score and MOS. PRCC measures the linear correlation between the evaluation score and MOS, with values ranging between −1 and 1. KRCC and SRCC compare rank correlation between evaluation score and MOS.

For comparison, five types of image quality assessment are introduced. Among them, there are three underwater IQA methods (*i.e.*, UIQM [45], UCIQE [46], and CCF [47]), and two in-air IQA methods (*i.e.*, CODE [7] and EMBM [8]). Table 2 provides the quantitative results of different methods on the UWIQA dataset. Compared with other methods, our proposed FDUM metric achieves the best performance in terms of the PRCC, KRCC, and SRCC. To be specific, when compared with the underwater IQA methods, the SRCC metric of our proposed method reaches 0.683, which has the percentage gain of 10% against the UIQM method, 9.8% against the UCIQE method, and 42.6% against the CCF method. Similarly, compared with the **second best** method, our FDUM achieves the minimum percentage gain of 11.8% in terms of KRCC, and 4.9% in terms of PRCC. It is clear that compared with the existing underwater IQA methods, our method has a stronger correlation with the subjective assessment of MOS. When compared with the CODE method [7] for in-air image, the proposed FDUM method achieves the minimum percentage gain of 17.1% in terms of PRCC, 21.8% in terms of KRCC, and 18.8% in terms of SRCC. Obviously, it is necessary to design an IQA method specifically for underwater images.

When evaluating the performance, the PRCC, KRCC, and SRCC directly calculate the correlation between the prediction and the MOS, which are the global measurements. In fact, in the UWIQA dataset, each

Table 3
Quantitative results of intra-group consistency.

	MAE	RMSE
FDUM	**0.037**	**0.051**
UIQM [45]	0.077	0.105
UCIQE [46]	0.082	0.108
CCF [47]	0.094	0.115

Table 4
Quantitative results of different enhancement methods.

Target	Metric	Image1	Image2	Image3	Image4
Original image	UIQM	0.653	0.796	1.169	1.747
	UCIQE	0.423	0.503	0.520	0.534
	CCF	8.055	22.433	20.238	48.680
	FDUM	0.124	0.253	0.356	0.623
Chiang [52]	UIQM	0.957	1.023	1.485	1.977
	UCIQE	0.537	0.652	0.599	0.572
	CCF	26.230	58.837	40.244	71.512
	FDUM	0.252	0.451	0.569	0.811
Ancuti [48]	UIQM	1.106	1.004	1.454	1.862
	UCIQE	0.596	0.652	0.600	0.588
	CCF	14.247	35.164	28.525	50.284
	FDUM	0.303	0.475	0.593	0.850

MOS group contains many underwater images, and the underwater images in each MOS group have the same MOS value and similar visual quality. A good underwater IQA algorithm should make the IQA score in each MOS group closer and close to the MOS score of the group, which we call it as intra-group consistency. The purpose of analyzing the intra-group consistency is to investigate whether the algorithm has good consistency evaluation results for underwater images in the same MOS group. First, we use different algorithms obtain the evaluation scores in each MOS group. Then, in each group, we calculate the mean value of all samples in this group and normalize the average values into [0.1, 1], generating the final quality score of each group. Finally, based on the above-mentioned the final quality score of each group, we draw the intra-group consistency curves, where the X axis represents the real MOS of each group, and the Y axis represents the final quality score of each group. Ideally, the black curve in Fig. 9 is the real intra-group consistency curve, and the closer to the black curve, the better the performance. Compared with other methods, from an intuitive point of view, our FDUM curve (the red curve) is closer to the MOS curve (the black curve). For the quantitative evaluation, we list the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) in Table 3. We can draw a conclusion consistent with Fig. 9. Our FDUM method achieves the smallest error against other method in all indexes. Specifically, for the MAE score, our method reaches 0.037, which reduces the error by 4% compared to the second best algorithm (*i.e.*, UIQM). Compared to the second best algorithm (*i.e.*, UIQM), the RMSE score of our method is reduced by 5.4%. All these measurements demonstrate the superior performance of our method in terms of the overall accuracy and intra-group consistency.

5.2. Image enhancement performance evaluation

In order to improve the quality of underwater image in different aspects, many underwater image enhancement algorithms were proposed. In this subsection, we test the effectiveness of different IQA methods for different enhancement methods. Two underwater image enhancement algorithms are used for comparisons: Ancuti et al. [48] proposed underwater video image enhancement method based on fusion principle. John et al. [52] proposed an enhancement algorithm combining techniques of (WCID). A group of underwater degraded images and corresponding enhanced results are shown in the Fig. 10. It is not difficult to see that the image enhancement method proposed in [48] achieves better visualization result. For each image in Fig. 10,

Fig. 10. (a) Original underwater images. (b) Underwater images enhanced by [52]. (c) Underwater images enhanced by [48].

we use different underwater IQA methods to evaluate them, and the objective evaluation scores are listed in Table 4. From the top to bottom, the MOSs of the original images are 0.1, 0.3, 0.4, and 0.6, respectively. Comparing the results shown in the first row block in Table 4, the quality score of our method is closer to the MOS than others. For the same underwater IQA metric, our method can obtain the same quantitative result as the subjective visual experience, that is to say, the method of [48] achieves the best enhancement effect among the two methods. But for other quality measures, this conclusion does not always hold. For example, for the second image, the CCF metric demonstrates the method of [52] has the best performance, but this is obviously inconsistent with our visual perception mechanism. Likewise, for the fourth image, the UIQM also indicates the best performance of method [52]. To sum up, for the underwater image enhancement task, our proposed metric is more in line with the human visual perception mechanism, and is more robust and stable.

6. Conclusion

In this paper, a new underwater image quality assessment metric without reference is proposed, named FDUM. This method combines spatial domain and frequency domain in colorfulness metric, and refines the contrast metric by using the dark channel prior. The final evaluation measurement of FDUM are obtained by combining the colorfulness, contrast, and sharpness. In addition, an Underwater Image Quality Assessment (UWIQA) dataset is constructed to bridge the gap between the algorithm design and data validation. Comprehensive experiments, including the comparisons with the state-of-the-art method, intra-group consistency analysis, and the quantitative validation of different enhancement methods, are conducted, demonstrating that our method can achieve significant performance improvements and more consistent visual perception.

CRediT authorship contribution statement

Ning Yang: Completed the design and implementation of the algorithm, Revised this paper several times to make it more clear and professional. **Qihang Zhong:** Completed the design and implementation of the algorithm, Revised this paper several times to make it more clear and professional. **Kun Li:** UWIQA dataset, Revised this paper several times to make it more clear and professional. **Runmin Cong:** Initiated the project, Collaboration, Revised this paper several times to make it more clear and professional. **Yao Zhao:** UWIQA dataset, Revised this paper several times to make it more clear and professional. **Sam Kwong:** Revised this paper several times to make it more clear and professional.

Declaration of competing interest

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

Acknowledgments

This work was supported by the Beijing Nova Program under Grant Z201100006820016, in part by the National Key Research and Development of China under Grant 2018AAA0102100, in part by the National Natural Science Foundation of China under Grant 62002014, Grant 61532005, Grant U1936212, Grant 61772344, Grant 61672443, in part by the Hong Kong RGC General Research Funds under Grant 9042816 (CityU 11209819), in part by the Fundamental Research Funds for the Central Universities under Grant 2019RC039, in part by Elite Scientist Sponsorship Program by the China Association for Science and Technology, in part by Elite Scientist Sponsorship Program by the Beijing Association for Science and Technology, in part by Hong Kong Scholars Program, in part by CAAI-Huawei MindSpore Open Fund, and in part by China Postdoctoral Science Foundation under Grant 2020T130050, Grant 2019M660438.

References

[1] C. Li, J. Guo, R. Cong, Y. Pang, B. Wang, Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior, IEEE Trans. Image Process. 25 (12) (2016) 5664–5677.
[2] C. Li, J. Guo, B. Wang, R. Cong, Y. Zhang, J. Wang, Single underwater image enhancement based on color cast removal and visibility restoration, J. Electron. Imaging 25 (3) (2016) 1–16.
[3] C. Li, J. Guo, C. Guo, R. Cong, J. Gong, A hybrid method for underwater image correction, Pattern Recognit. Lett. 94 (2017) 62–67.
[4] C. Li, S. Anwar, F. Porikli, Underwater scene prior inspired deep underwater image and video enhancement, Pattern Recognit. 98 (2019) 1–11.
[5] X. Chen, J. Yu, S. Kong, Z. Wu, X. Fang, L. Wen, Towards real-time advancement of underwater visual quality with GAN, IEEE Trans. Ind. Electron. 66 (12) (2019) 9350–9359.
[6] Y. Li, H. Lu, J. Li, X. Li, Y. Li, S. Serikawa, Underwater image de-scattering and classification by deep neural network, Comput. Electr. Eng. 54 (54) (2016) 68–77.
[7] K. Gu, W. Lin, G. Zhai, X. Yang, W. Zhang, C.W. Chen, No-reference quality metric of contrast-distorted images based on information maximization, IEEE Trans. Cybern. 47 (12) (2017) 4559–4565.
[8] J. Guan, W. Zhang, J. Gu, H. Ren, No-reference blur assessment based on edge modeling, J. Vis. Commun. Image Represent. 29 (5) (2015) 1–7.
[9] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, D. Tao, An underwater image enhancement benchmark dataset and beyond, IEEE Trans. Image Process. 29 (2020) 4376–4389.
[10] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[11] Z. Wang, E.P. Simoncelli, A.C. Bovik, Multiscale structural similarity for image quality assessment, in: The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003, pp. 1398–1402.
[12] L. Zhang, L. Zhang, X. Mou, D. Zhang, FSIM: A feature similarity index for image quality assessment, IEEE Trans. Image Process. 20 (8) (2011) 2378–2386.
[13] W. Xue, L. Zhang, X. Mou, A.C. Bovik, Gradient magnitude similarity deviation: A highly efficient perceptual image quality index, IEEE Trans. Image Process. 23 (2) (2014) 684–695.
[14] L. Liu, H. Dong, H. Huang, A.C. Bovik, No-reference image quality assessment in curvelet domain, Signal Process., Image Commun. 29 (4) (2014) 494–505.
[15] J. Wu, M. Zhang, L. Li, W. Dong, W. Lin, No-reference image quality assessment with visual pattern degradation, 504, 2019, pp. 487–500,
[16] W. Xue, L. Zhang, X. Mou, Learning without human scores for blind image quality assessment, in: IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 995–1002.
[17] L. Tang, K. Sun, L. Liu, G. Wang, Y. Liu, A reduced-reference quality assessment metric for super-resolution reconstructed images with information gain and texture similarity, Signal Process., Image Commun. 79 (2019) 32–39.
[18] Y. Fang, J. Liu, Y. Zhang, W. Lin, Z. Guo, Reduced-reference quality assessment of image super-resolution by energy change and texture variation, J. Vis. Commun. Image Represent. 60 (2019) 140–148.
[19] S. Mahmoudpour, P. Schelkens, Reduced-reference quality assessment of multiply-distorted images based on structural and uncertainty information degradation, J. Vis. Commun. Image Represent. 57 (2018) 125–137.
[20] J. Wu, W. Lin, Y. Fan, L. Li, G. Shi, I. Niwas, S, visual structural degradation based reduced-reference image quality assessment, Signal Process., Image Commun. 47 (2016) 16–27.
[21] Z. Chen, Q. Xu, R. Cong, Q. Huang, Global context-aware progressive aggregation network for salient object detection, in: Thirty-Fourth AAAI Conference on Artificial Intelligence, 2020, pp. 10599–10606.
[22] R. Cong, J. Lei, H. Fu, M.-M. Cheng, W. Lin, Q. Huang, Review of visual saliency detectioin with comprehensive information, IEEE Trans. Circuits Syst. Video Technol. 29 (10) (2019) 2941–2959.
[23] C. Li, R. Cong, J. Hou, S. Zhang, Y. Qian, S. Kwong, Nested network with two-stream pyramid for salient object detection in optical remote sensing images, IEEE Trans. Geosci. Remote Sens. 57 (11) (2019) 9156–9166.
[24] C. Li, R. Cong, S. Kwong, J. Hou, H. Fu, G. Zhu, D. Zhang, Q. Huang, ASIF-Net: Attention steered interweave fusion network for RGB-D salient object detection, IEEE Trans. Cybern. 50 (1) (2021) 88–100.
[25] Z. Chen, R. Cong, Q. Xu, Q. Huang, DPANet: Depth potentiality-aware gated attention network for RGB-D salient object detection, IEEE Trans. Image Process. PP (99) (2021) 1–13.
[26] Q. Zhang, R. Cong, C. Li, M.-M. Cheng, Y. Fang, X. Cao, Y. Zhao, S. Kwong, Dense attention fluid network for salient object detection in optical remote sensing images, IEEE Trans. Image Process. 30 (2021) 1305–1317.
[27] R. Cong, J. Lei, H. Fu, J. Hou, Q. Huang, S. Kwong, Going from RGB to RGBD saliency: A depth-guided transformation model, IEEE Trans. Cybern. 50 (8) (2020) 3627–3639.
[28] F. Li, R. Cong, H. Bai, Y. He, Deep interleaved network for image super-resolution with asymmetric co-attention, in: International Joint Conference on Artificial Intelligence, 2020, pp. 534–543.
[29] C. Guo, C. Li, J. Guo, R. Cong, H. Fu, P. Han, Hierarchical features driven residual learning for depth map super-resolution, IEEE Trans. Image Process. 28 (5) (2018) 2545–2557.
[30] C. Guo, C. Li, J. Guo, C.-C. Loy, J. Hou, S. Kwong, R. Cong, Zero-reference deep curve estimation for low-light image enhancement, in: IEEE Conference on Computer Vision and Pattern Recognition, 2020, pp. 1780–1789.
[31] C. Li, C. Guo, J. Guo, P. Han, H. Fu, R. Cong, PDR-net: Perception-inspired single image dehazing network with refinement, IEEE Trans. Multimed. 22 (3) (2019) 704–716.
[32] J. Choi, H.J. Chang, T. Fischer, S. Yun, Y.C. Jin, Context-aware deep feature compression for high-speed visual tracking, in: IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 479–488.
[33] K. Le, Y. Peng, L. Yi, D. Doermann, Convolutional neural networks for no-reference image quality assessment, IEEE Conference on Computer Vision and Pattern Recognition (2014) 1733–1740.
[34] X. Liu, V.D.W. Joost, A.D. Bagdanov, RankIQA: Learning from rankings for no-reference image quality assessment, in: IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1040–1049.
[35] S. Bianco, L. Celona, P. Napoletano, R. Schettini, On the use of deep learning for blind image quality assessment, Signal, Image Video Process. 12 (2) (2018) 355–362.
[36] S. Bosse, D. Maniry, K.-R. Muller, T. Wiegand, W. Samek, Deep neural networks for no-reference and full-reference image quality assessment, IEEE Trans. Image Process. 27 (1) (2018) 206–219.
[37] K. Gu, G. Zhai, X. Yang, W. Zhang, Deep learning network for blind image quality assessment, in: IEEE International Conference on Image Processing, 2014, pp. 511–515.
[38] S. Bosse, D. Maniry, T. Wiegand, W. Samek, A deep neural network for image quality assessment, in: IEEE International Conference Image Processing, 2016, pp. 3773–3777.
[39] L. Xin, L. Zhe, X. Shen, R. Mech, J.Z. Wang, Deep multi-patch aggregation network for image style, aesthetics, and quality estimation, in: IEEE International Conference on Computer Vision, 2015, pp. 990–998.
[40] B. Jin, M.V.O. Segovia, S. Susstrunk, Image aesthetic predictors based on weighted CNNs, in: IEEE International Conference on Image Processing, 2016, pp. 2291–2295.
[41] Z. Ni, L. Ma, H. Zeng, J. Chen, C. Cai, K.K. Ma, ESIM: Edge similarity for screen content image quality assessment, IEEE Trans. Image Process. 26 (10) (2017) 4818–4831.
[42] Z. Ni, H. Zeng, L. Ma, J. Hou, K.K. Ma, A gabor feature-based quality assessment model for the screen content images, IEEE Trans. Image Process. 27 (9) (2018) 4516–4528.
[43] Z. Ni, L. Ma, H. Zeng, C. Cai, K.K. Ma, Gradient direction for screen content image quality assessment, IEEE Signal Process. Lett. 23 (10) (2016) 1394–1398.
[44] Q. Jiang, W. Gao, S. Wang, G. Yue, S. Kwong, Blind image quality measurement by exploiting high order statistics with deep dictionary encoding network, IEEE Trans. Instrum. Meas. 69 (10) (2020) 7398–7410.
[45] M. Yang, A. Sowmya, An underwater color image quality evaluation metric, IEEE Trans. Image Process. 24 (12) (2015) 6062–6071.
[46] K. Panetta, C. Gao, S. Agaian, Human-visual-system-inspired underwater image quality measures, IEEE J. Ocean. Eng. 41 (3) (2016) 541–551.
[47] Y. Wang, N. Li, Z. Li, Z. Gu, H. Zheng, B. Zheng, M. Sun, An imaging-inspired no-reference underwater color image quality assessment metric, Comput. Electr. Eng. 70 (2017) 904–913.

[48] T.H.C. Ancuti, C.O. Ancuti, P. Bekaert, Enhancing underwater images and videos by fusion, in: IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 81–88.
[49] R. Schettini, S. Corchs, Underwater image processing: State of the art of restoration and image enhancement methods, EURASIP J. Adv. Signal Process. 2010 (2010) 1–14.
[50] K. He, J. Sun, X. Tang, Single image haze removal using dark channel prior, IEEE Trans. Pattern Anal. Mach. Intell. 33 (12) (2011) 2341–2353.
[51] K. Panetta, A. Samani, S.S. Agaian, Choosing the optimal spatial domain measure of enhancement for mammogram images, Int. J. Biomed. Imaging (2014) 937849.
[52] J.Y. Chiang, Y.C. Chen, Underwater image enhancement by wavelength compensation and dehazing, IEEE Trans. Image Process. 21 (4) (2012) 1756–1769.