

REVIEW OPEN ACCESS

Underwater Image Quality Evaluation: A Comprehensive Review

 Mengjiao Shen¹  | Miao Yang¹  | Jinyang Zhong¹ | Hantao Liu² | Can Pan¹
¹School of Electronic Engineering, Jiangsu Ocean University, Lianyungang, China | ²School of Computer Science and Informatics, Cardiff University, Cardiff, UK

Correspondence: Miao Yang (lemonmiao@gmail.com)

Received: 15 October 2024 | **Revised:** 9 March 2025 | **Accepted:** 22 March 2025

Funding: This work was supported in part by National Natural Science Foundation of China (NSFC) under Grant 62271236, National Key R&D Program Project 2023YFC3108205, Key Country-Specific Industrial Technology R&D Cooperation Project 23GH002, Graduate Research and Practice Innovation Program under Grants KYCX2023-67 and KYCX23-3462.

ABSTRACT

Underwater image quality evaluation (UIQE) is crucial in improving image processing techniques and optimizing the design of the imaging system to obtain object information more accurately. However, existing UIQE methods are designed based on limited images or consider only a few natural scene statistics (NSS) metrics, lacking consideration for generalization across various underwater imaging applications. In this paper, an in-depth review of the existing UIQE methods based on evaluation operations is provided, emphasizing the bias present when evaluating UIQE methods using individual metrics. To address this, a novel metric called quadrilateral datum evaluation (QDE) is designed for UIQE methods. It comprehensively considers robustness across different datasets, as well as correlation and ranking consistency with mean opinion scores (MOS). This is the first solution to measure an UIQE method from an all-encompassing visual perspective. By using QDE, UIQE methods characterized by greater feature strength and small imbalance demonstrate good consistency and robustness across multiple aspects, providing a basis for the design of UIQE methods.

1 | Introduction

Underwater vision technology is crucial in subsea exploration, marine ecological protection, and underwater engineering monitoring [1–5]. Image quality assessment (IQA) evaluates the quality of an image, helping to determine whether the image meets the expected visual standards. Many algorithms such as blind/referenceless image spatial quality evaluator (BRISQUE) [6], natural image quality evaluator (NIQE) [7], distortion identification and label transfer (DILT) [8], integrated local natural image quality evaluator (ILNIQE) [9], structure naturalness and perception NIQE (SNPNIQE) [10], and deep learning-based methods like deep learning IQA (DLIQA) [11], RankIQA [12], deep bilinear convolutional neural network (DB-CNN)

[13], HyperIQA [14], unified no-reference image quality and uncertainty evaluator (UNIQUE) [15], multibranch convolutional neural network (MB-CNN) [16] are proposed recently. However, measuring the quality of underwater images is challenged because of the absorption and scattering of light, non-uniform illumination, and noise [17–19]. The underwater image quality evaluation (UIQE) methods can be categorized as subjective methods and objective methods. Subjective UIQE methods provide the quality benchmark but are manpower and time-consuming. In contrast, objective UIQE offers advantages such as repeatability and automation, making them highly attractive to researchers. The early research on objective UIQE mainly focused on the quality evaluation for grayscale images [20–23]. For underwater color images, the distinct evaluation metrics tailored

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2025 The Author(s). *IET Image Processing* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology.

to the unique characteristics of the underwater environment are commonly employed enabling automated quality assessment [24–35].

As humans are the ultimate recipients of images, the performance of an IQA method is usually validated by correlation metrics such as the Pearson linear correlation coefficient (PLCC), Spearman rank order correlation coefficient (SROCC), Kendall rank order correlation coefficient (KROCC), and root mean square error (RMSE). The PLCC measures the linear correlation between the algorithm's predicted values and the mean opinion scores (MOS). At the same time, the SROCC assesses the monotonic relationship between the two, and the KROCC quantifies their ordinal relationship. The RMSE measures the difference between IQA values and the MOSs, with values closer to 0 indicating better IQA algorithm performance. In addition, the pristine/distorted image discriminability test (D-test), the listwise ranking consistency test (L-test), and the pairwise preference consistency test (P-test) [36], are commonly applied to validate the relative performance of an IQA metric. These include the ability to discern between different categories of images in space, the robustness of image ratings for images with identical content and distortion types but varying degrees of distortion, and the prediction of preferences for image pairs exhibiting significant quality disparities.

Recently, the maximum differentiation (MAD) [37] and group maximum differentiation (gMAD) [38] methods are proposed to compare the perceptual discriminability performance between two or more IQA methods. The MAD works by refuting the compared model using another model, and a model that is more challenging to refute is considered better. The gMAD quantifies attack performance by the attacker searching for the maximum difference in prediction of other models within the same mass range and quantifies defense performance by the degree to which the defender defends from being attacked. These two competitions use all possible combinations of defender-attackers within each model.

Assessing the generalization performance of an IQA algorithm also requires considering factors such as the richness of the datasets and the diversity of the evaluation perspectives. Existing UIQE methods are usually developed based on certain datasets [31, 32] and focus on a single aspect of performance, such as the consistency with the distortion level [36], ranking consistency with the enhancement results [33], or the correlation with the MOS values [27, 29]. However, in underwater tasks like surveillance, optical characteristics can significantly affect image quality and clarity, meaning that high-quality original images do not guarantee exceptional enhanced images. To facilitate this process, there is a need for a UIQE method that demonstrates strong performance in comparing the quality of enhanced images across various content under the comprehensive influence of underwater optical characteristics. Existing evaluation metrics for UIQE methods typically only assess performance under the influence of underwater optical characteristics, performance between enhanced images, or performance between images with different contents, without directly evaluating the comprehensive performance of these aspects.

Given the limitations of existing studies, numerous review studies have been conducted in the field of underwater image quality

in recent years. For instance, Raveendran et al. [39] classified and summarized various underwater image enhancement (UIE) algorithms, and listed relevant datasets. Wang et al. [40] provided a review of multiple aspects within the underwater imaging domain, including underwater image enhancement and restoration (UIER), datasets, and evaluation methods. Xu et al. [41] offered a systematic classification and analysis of UIQE methods from both underwater optical images and sonar images perspectives, delving into imaging principles, influencing factors, datasets, and evaluation methods. Despite providing valuable insights, these studies each have their own focuses and limitations. Specifically, the review by Raveendran et al. [39] primarily centers on UIE methods, while the reviews by Wang et al. [40] and Xu et al. [41], although covering a broader range of topics, lack in-depth analysis of UIQE methods for underwater optical images. This paper introduces a new classification of existing underwater image quality datasets (UIQD) and proposes a comprehensive metric to evaluate UIQE algorithms from four key perspectives. The goal is to develop a flexible evaluation framework adaptable to different datasets and metrics based on task-specific requirements. The main contributions of this paper are as follows:

- A novel dataset classification is introduced in this paper considering distortion level and the subjective evaluation mode simultaneously. With this classification, it becomes more straightforward to comprehend the accuracy of the UIQE algorithms in assessing images with different levels of optical distortion, enhanced outputs, and underwater images with diverse content.
- A comprehensive review of 37 UIQE algorithms and 25 UIQDs is provided, where the image features, the fitting methods, and evaluation operations of the UIQE algorithms are discussed, enabling researchers an in-depth understanding of the progress of each algorithm. Particularly, we conduct a multidimensional evaluation of the blind underwater image quality evaluation (BUIQE) method based on the newly proposed dataset classification method, utilizing various evaluation metrics such as PLCC, L-test, D-test, and P-test to reflect the characteristics of different BUIQE methods across various scenarios. Experimental results find that relying solely on specific datasets is insufficient for fully assessing the performance of the BUIQE algorithm.
- An intuitive QDE is proposed to comprehensively evaluate the performance of the UIQE algorithm in various underwater scenarios. By comparing different imaging distances, water turbidities, underwater enhanced images of the same content, and images of different content, the overall performance of the UIQE algorithm could be assessed. The flexibility of QDE allows for the addition or removal of dataset types and metrics to adapt to specific task requirements. Higher feature strength and small imbalance parameter values indicate superior performance of the UIQE method in terms of accuracy and robustness in underwater image quality assessment. The visibility of the quadrilateral radar charts and polar coordinate plots facilitates a deeper understanding of the UIQE algorithm's performance across various scenarios. By highlighting visual representations, researchers can identify weaknesses and make targeted improvements, enhancing the UIQE algorithm's performance and applicability.

The structure of the paper is as follows: Section 2 introduces the impact of underwater imaging on image quality and UIQE datasets. Section 3 discusses existing UIQE methods. Section 4 introduces the proposed intuitive and comprehensive measurement for the UIQE. Finally, the paper is concluded in Section 5.

2 | Subjective Underwater Image Quality Assessment

2.1 | Underwater Imaging System

Based on the commonly used Jaffe-McGlamery underwater imaging physical model [42], light primarily travels through three paths in water to reach the imaging plane. These three paths are direct reflection, forward-scattered, and backscattering. The total irradiance can be expressed as:

$$E = E_d + E_f + E_b \quad (1)$$

where E represents the total energy received by the underwater camera, E_d is the direct reflection energy, E_f and E_b represent the forward-scattered and backscatter energy, respectively.

The direct reflection component refers to the light that is directly reflected from the target object to the camera. The attenuation of this component is primarily due to the propagation loss of light in a homogeneous water medium. The forward-scattered component refers to the small-angle scattering of light due to particles on the target surface. This scattering can cause blurring of image details and edge information. The backscatter component is caused by the scattering of suspended particles in the water and usually reduces the contrast of the image. It also forms a hazy background on the original image.

Additionally, imaging under the water is influenced by factors such as imaging equipment and light sources, leading to diminished colorfulness, contrast, clarity, and loss of details in the image.

2.2 | Underwater Image Quality Assessment Dataset

Typically, image quality datasets are divided into real and synthetic image datasets [43]. However, the complexity of distortions in underwater images poses simulation challenges. Hou et al. [44] categorized underwater image datasets into UIER datasets and UIQE datasets. UIER datasets are mainly used to test UIE algorithms, while UIQE datasets are mainly used to evaluate UIQE algorithms. Nevertheless, this classification method does not fully consider the impact of various types and levels of distortion in underwater environments on image quality. To address this limitation, we categorize the UIQDs by the presence of distortion levels and further subdivide the datasets without distortion levels according to whether the image is grouped for the same content or not. This division helps assess how an UIQE algorithm performs under different quality evaluation needs and gain a deeper insight into its robustness.

2.2.1 | Underwater Image Quality Dataset Without Distortion Levels

To acquire a diverse range of underwater image samples, researchers typically gather them through public websites, diving, and tank imaging. Due to differences in the subjective evaluation methods, the intended applications of MOS vary. Here we categorized the UIQDs without distortion levels into two major classes: grouped evaluation-based and ungrouped evaluation-based, as shown in Figure 1. The underwater image datasets such as PKU-EAQA dataset [45], UIEB dataset [46], UID2021 dataset [44], Gu et al. [47], SAUD dataset [32], UIEQ dataset [48], and URankerSet dataset [49] are all based on grouped evaluation mode. Researchers use either the pairwise comparison (PC) method or the double stimulus (DS) method to evaluate groups of images with the same content. Partial images of the UID2021 dataset are illustrated in Figure 2. Each image in the UID2021 dataset is accompanied by 15 enhanced or restored versions. The values of the UIQE for this dataset are not a direct comparison of images with different contents. Hence, for assessing UIQE algorithm performance on these datasets, it's advisable to utilize the L-test instead of PLCC for evaluation. Furthermore, an effective UIQE algorithm should ensure that the quality of enhanced images falls within a higher range, while the quality of the raw images should be constrained to a lower range. Therefore, when L-test values of certain UIQE algorithms are closely clustered, prioritizing the algorithm with a superior D-test value is needed.

In contrast, ungrouped evaluation mode datasets without distortion levels, such as the KonIQ-10k dataset [50], UOQ dataset [27], UWIQA dataset [28], UEIQA dataset [30], UIFD dataset [51], UIED dataset [33], UIQAB dataset [52], UIUD dataset [53], Swimming pool dataset [54], USRD dataset [55], LSUI dataset [56], and UIQD dataset [57] evaluate images directly. Partial images from the UWIQA dataset are illustrated in Figure 3. Image scores are determined individually by observers and are not influenced by reference images. This type of dataset can be used to evaluate the algorithm's linear correlation with MOSs in predicting image quality values. For image pairs with significant differences in MOSs, we anticipate notable disparities in their IQA prediction values as well. Thus, besides reflecting the correlation between image quality predictions and MOSs with PLCC values, higher P-test values are essential to elucidate the algorithm's significant capability in distinguishing image pairs with different qualities.

2.2.2 | Underwater Image Quality Dataset with Distortion Levels

The interactions between light and water molecules, as well as suspended particles, constitute the primary factors influencing the absorption and scattering of light. These interactions are directly associated with imaging distances and water turbidities. Consequently, researchers gather sequence images by varying imaging distances or water turbidity in the tank to construct distortion level sequence datasets, aiming to evaluate the algorithm's performance in sensitivity to underwater optical distortion levels. These datasets encompass the TURBID dataset [73], NWPU

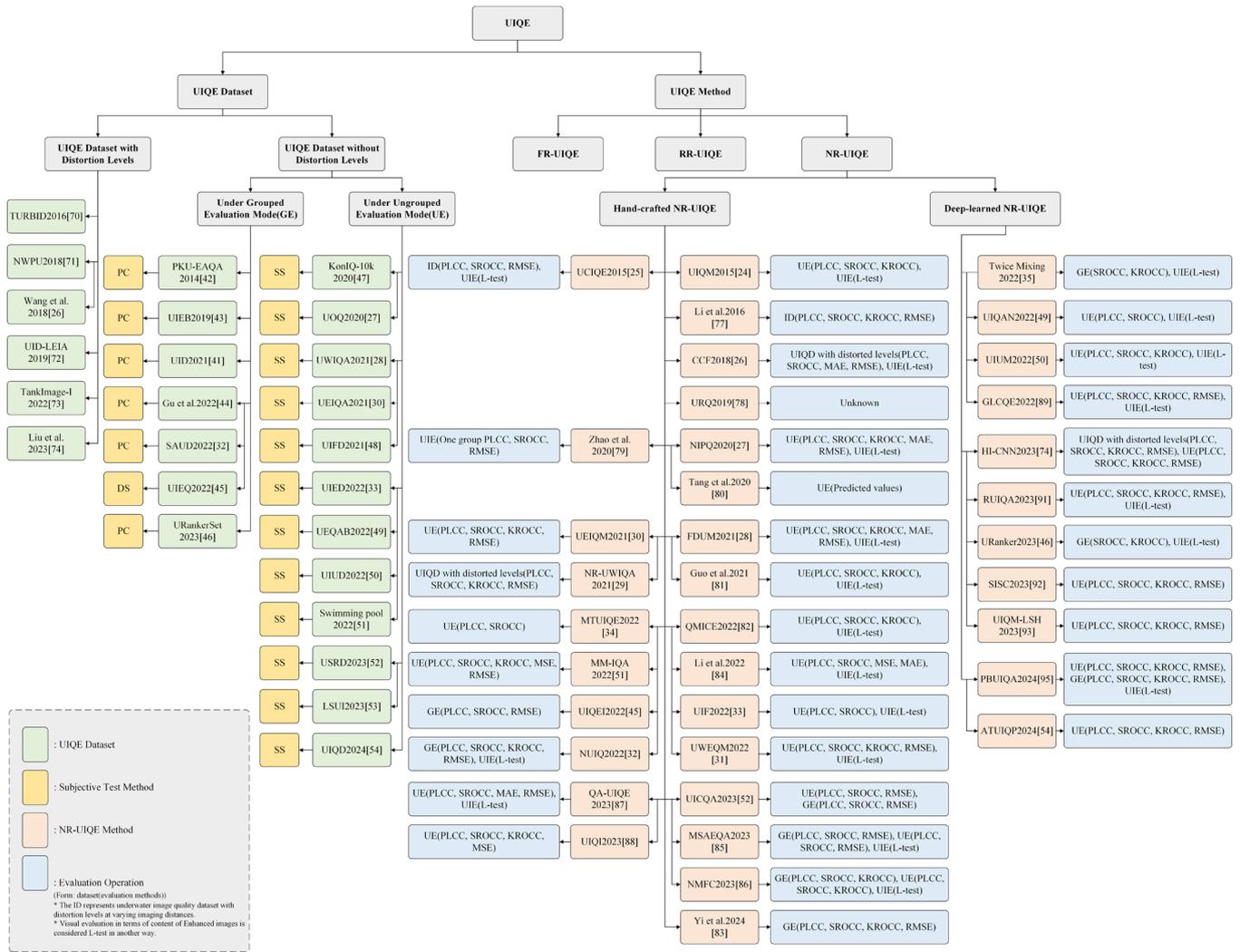


FIGURE 1 | The illustration of underwater image quality assessment. (The "ID" represents the underwater image quality dataset with distortion levels at varying imaging distances. The terms "SS", "DS", and "PC" represent the single stimulus method, double stimulus method, and pairwise comparison method, respectively. Visual evaluation in terms of the content of enhanced images is considered L-test in another way.)

dataset [74], Wang et al. dataset [26], UID-LEIA dataset [75], TankImage-I dataset [76], and Liu et al. dataset [77]. Figures 4 and 5 show images from the TankImage-I dataset and the UID-LEIA dataset, respectively. The subjective evaluation of the UID-LEIA dataset proposed by Sanchez-Ferreira et al. [75] uses the mean annoyance value (MAV). In contrast to the MOS, a larger MAV value indicates poorer image quality. It can be seen from Figure 5 that the MAV of the green object taken at the 20 cm distance, surpasses that of the blue object, while at the 30 cm distance, the MAV values of the blue object exceed that of the green object. It indicates utilizing the MOS/MAV to directly qualify the underwater image sequence with different content is not feasible. The L-test is frequently applied to such datasets to assess the algorithm's sensitivity to distortion levels. Nevertheless, it is crucial to acknowledge that the L-test for such datasets encompasses the performance of the D-test to some extent. Consequently, the D-test is employed for comparison when the values of the L-test are not comparable. Furthermore, distortion level sequences can furnish relative quality rankings for image pairs and function as a data augmentation technique for quality modeling based on deep learning [12, 78].

3 | Objective Underwater Image Quality Assessment

According to the degree of reliance on original image information, the UIQE methods can be divided into full reference underwater image quality evaluation (FR-UIQE), reduced reference underwater image quality evaluation (RR-UIQE), and no-reference underwater image quality evaluation (NR-UIQE)/BUIQE.

3.1 | Full Reference Underwater Image Quality Assessment

The FR-UIQE method aims to evaluate the quality of a target image using an ideal image as a reference. Common evaluation metrics include mean absolute error (MAE), RMSE, peak signal-to-noise ratio (PSNR), structural similarity measure (SSIM) [79], among others. Researchers typically opt for images that have undergone some degree of enhancement or processing, especially those considered to be the optimal enhanced or captured

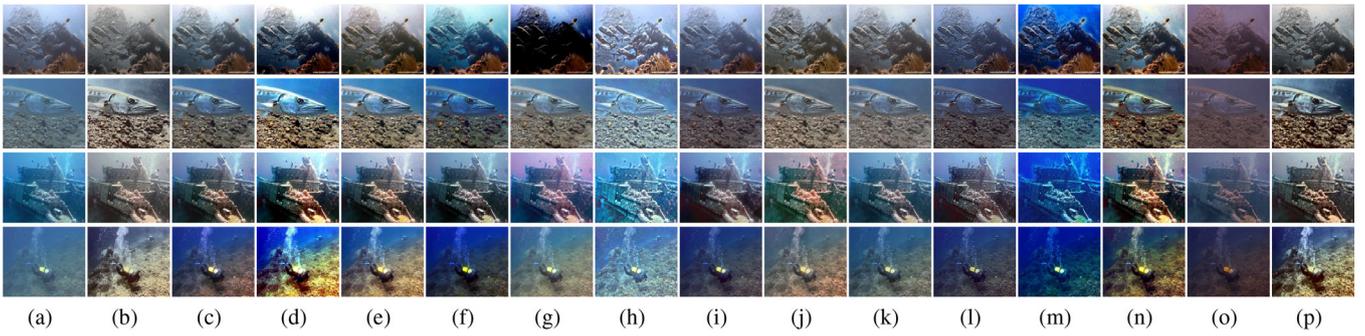


FIGURE 2 | Partial images from the UID2021 dataset (From left to right: source images (a), and the enhanced or restored results by using Bayesain-retinex [58] (b), CBF [59] (c), CHS [60] (d), GLN-CHE [61] (e), HP [62] (f), IBLA [63] (g), L^2 UWE [64] (h), RCP [65] (i), TS [66] (j), Ucolor [67] (k), UNTV [68] (l), UTV [69] (m), UWB-VCSE [70] (n), UWCNN [71] (o), and VR [72] (p), respectively).

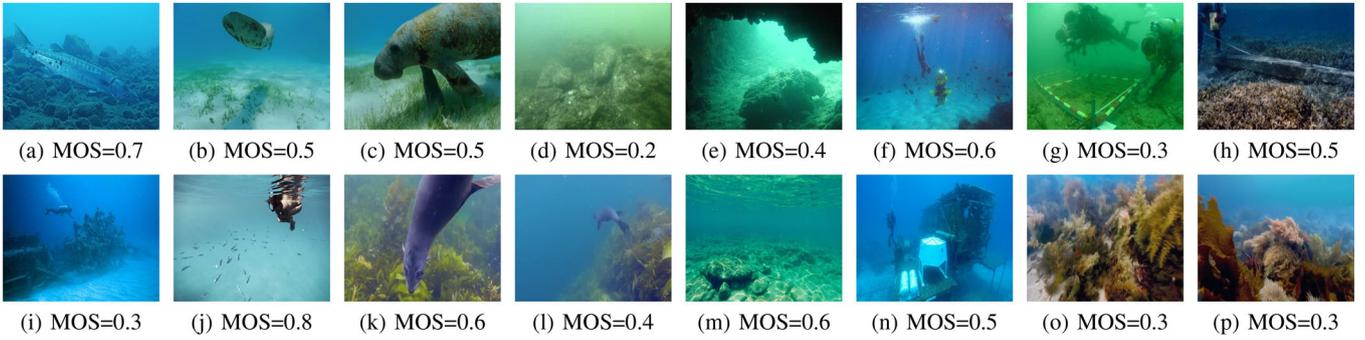


FIGURE 3 | Partial images from the UWIIQA dataset.

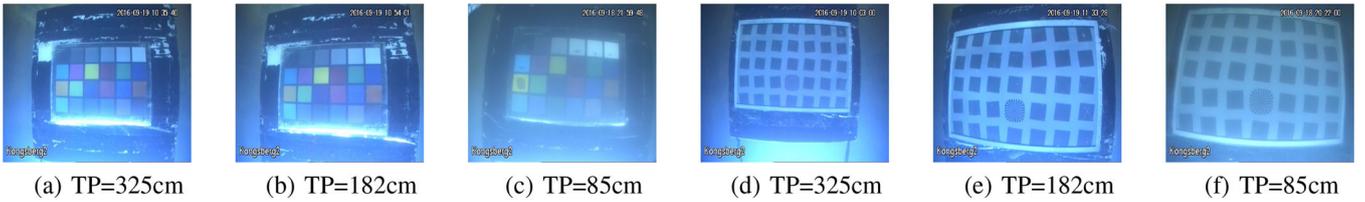


FIGURE 4 | Partial images from the TankImage-I dataset. The camera is positioned at a distance of 60 cm. TP means transparency.

at the nearest imaging distance, to serve as reference images [56]. However, different enhancement or restoration techniques may introduce varying degrees of distortion, meaning that even though certain images may appear visually optimal, they may not entirely represent the true quality of the original image. Secondly, underwater imaging distance and angle can also impact image quality, and selecting the nearest image as a reference may overlook the effects of light propagation. Moreover, the subjective evaluation of the results of the enhancement method is related to the sensitivity of the observers. Therefore, the BUIQE methods are more suitable.

3.2 | Blind Underwater Image Quality Assessment

The BUIQE methods can be divided into traditional machine learning-based and deep learning-based methods, as shown in Figure 1.

Traditional machine learning-based methods mainly consist of two modules: feature processing and feature fusion. Early researchers developed a series of methods, such as underwater image quality measure (UIQM) [24], underwater color image quality evaluation (UCIQE) [25], Li et al. [80], CCF [26], URQ [81], no-reference image quality predictor (NIPQ) [27], Zhao et al. [82], Tang et al. [83], frequency domain underwater metric (FDUM) [28], Guo et al. [84], underwater enhancement image quality metric (UEIQM) [30], no-reference quality metric for images of underwater targets in multi-colored environments (QMICE) [85], and Yi et al. [86] selecting chromaticity, contrast, clarity, fog density, visibility, and other one to three quality perception features as indicators to evaluate underwater image quality. However, these indicators are relatively limited and often neglect the deep-level visual characteristics of images. In recent years, researchers have also considered aspects such as image texture, structure, naturalness, and noise, proposing methods like no-reference image quality evaluation metric for enhanced underwater images (NR-UWIIQA) [29], Li et al. [87], underwater

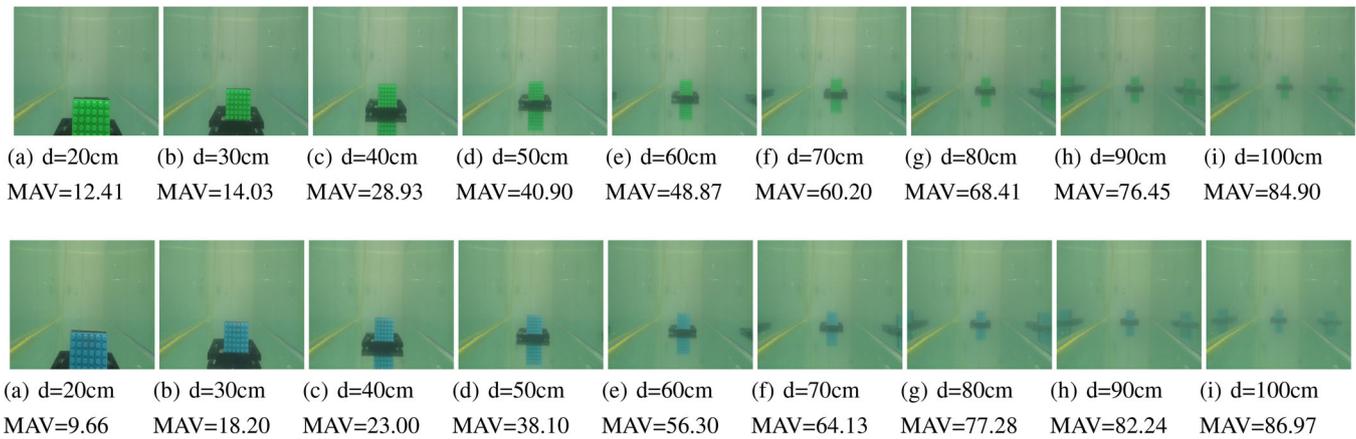


FIGURE 5 | Partial images from the UID-LEIA dataset. The d represents the imaging distance.

image fidelity (UIF) [33], UWEQM [31], multitopic underwater image quality evaluation (MTUIQE) [34], multi-feature fusion image quality assessment (MM-IQA) [54], UIQE index (UIQEI) [48], no-reference (NR) underwater image quality metric (NUIQ) [32], UICQA [55], MSAEQA [88], no-reference underwater image quality assessment based on multi-feature fusion in color space (NMFC) [89], QA-UIQE [90], and underwater image quality index (UIQI) [91] to better capture image details and deep-level features. In terms of feature fusion, as the number of features increases, traditional linear methods like multiple linear regression (MLR) are insufficient. The support vector machine regression (SVR), random forests regression (RFR), and Gaussian process regression (GPR) are applied to map the features and the MOSS.

However, traditional machine learning-based approaches often necessitate manual feature extraction and may have limitations when confronted with complex and variable underwater environments. With the emergence of deep learning in the field of computer vision, deep learning-based techniques have become a new avenue for addressing underwater image quality assessment issues. For instance, in 2022, Fu et al. [35] introduced the twice mixing, a no-reference underwater image quality assessment method based on rank learning. This method leveraged the concept of blending high-quality and low-quality images to generate medium-quality images and utilizes rank image pairs for training the network to evaluate image quality. Li et al. [52] proposed an underwater image quality assessment network (UIQAN) for UIE images, which dynamically adjusted quality perception feature maps by embedding deep image prior spatial attention blocks, thus modeling human visual attention in a data-driven manner. Additionally, Chen et al. [53] introduced an underwater image utility measure (UIUM) that links image quality with task-based practicality. This method employed transfer learning to extract task-relevant features and transfer features from fish detection to practical-oriented image quality assessment utilizing shared layers. The generation-based joint luminance-chrominance underwater image quality evaluation (GLCQE) [92] proposed by Wang et al. is a generation-based joint luminance-chrominance learning for BUIQE for evaluating the quality of underwater images. Its core concept involves utilizing the direction and extent of image enhancement to assess image quality. GLCQE employed generative adversarial networks (GAN)

[93] to generate two images: one is the unenhanced version, and the other is the optimally enhanced version. Subsequently, GLCQE separated the luminance and chrominance channels of these images and utilized them to detect image distortions. Liu et al. [77] devised an end-to-end hybrid-input convolutional neural network (HI-CNN), for predicting the degradation of underwater images. This method employed a feature extraction module to simultaneously extract features from both the original underwater images and saliency maps and designed an end-to-end model for predicting image quality scores. Subsequently, Wang et al. [94] proposed a rank-based underwater quality assessment algorithm (RUIQA) for domain-adapted underwater image enhancement, which trained the network using hidden quality information learned from rankings. Guo et al. [49] proposed an underwater image quality assessment method based on Transformer and multi-scale features, named URanker. It processed input images through multiple serial and parallel blocks and utilized a histogram prior module and dynamic connection mechanism to enhance evaluation performance. During the training process, the marginal ranking loss function was employed to learn the relative quality ranking of image pairs with sorting labels to optimize the model. Chu et al. [95] introduced a feature interaction-based underwater image quality assessment method named SISC. The method employed a self-attention-based feature refinement module to facilitate both local and non-local cross-spatial feature interactions. Additionally, the method incorporated a cross-attention-based feature fusion module to achieve cross-scale feature interactions, enhancing information complementarity across different scales and further refining the feature representation. Furthermore, Liu et al. [96] introduced an underwater image quality metric with low-level properties and selected high-level semantics (UIQM-LSH), an enhancing underwater image quality assessment with influential perceptual features. This method considered low-level image properties such as colorfulness, contrast, brightness, as well as high-level semantic features selected through vision transformer [97] training and minimum angle regression. Image quality was then predicted through SVR feature processing. In 2024, Wang et al. [98] proposed a prior-based underwater image enhancement quality assessment network (PBUIQA). The method utilized a parameter estimation network composed of two mutually supervised sub-networks, a U-Net-based [99] encoder-decoder network and a dense-based network, to predict the parameters

TABLE 1 | Performance comparisons of UIQM and UCIQE methods with different metrics.

Method	L-test (SAUD)	PLCC (UWIQA)
UIQM	0.3161	0.6315
UCIQE	0.4259	0.5680

of the underwater raw images. Subsequently, these features were fed into a quality assessment network through an underwater prior-guided module, aiding the network in accurately perceiving image distortions. Liu et al. [57] introduced an end-to-end underwater image quality assessment network named ATUIQP. This network innovatively integrated channel attention, spatial attention mechanisms, and transformer [100] to accurately characterize the degradation of image quality in terms of channel, spatial, and global aspects.

Compared to traditional methods, deep learning methods are able to learn higher-level feature representations from raw data, thus reducing the need for manual feature design, and generally have better generalization capabilities to handle more complex scenarios and data changes.

It can be seen in Figure 1, it is evident that the evaluation of the UIQE method typically relies solely on one or a few datasets. To investigate the impact of different metrics and datasets on the evaluation of UIQE, we computed the L-test values of the UCIQE [25] and UIQM [24] methods on the SAUD dataset [32]. Subsequently, we perform PLCC evaluation using the UCIQE and UIQM methods on the UWIQA dataset [28]. The outcomes are listed in Table 1, which reveals that UCIQE outperformed UIQM in terms of L-test performance on the SAUD dataset, whereas UIQM exhibited superior PLCC performance over UCIQE on the UWIQA dataset. This underscores the results of UIQE methods are dataset dependent.

3.3 | Evaluations of BUIQE Methods

A good BUIQE method should be outstanding in measuring images with varying imaging distances, and images with varying water turbidities, comparing the visual quality of different UIE outputs, and reflecting the correlation of predicted values and MOSs between different content images. Based on the division in Section 2, we utilize the UIQDs with distortion levels at varying imaging distances and water turbidities, as well as the UIQDs without distortion levels under grouped evaluation mode and ungrouped evaluation mode to analyze the performance of the existing BUIQE methods. Fifteen BUIQE methods, including UIQM [24], UCIQE [25], CCF [26], URQ [81], Zhao et al. [82], FDUM [28], UEIQM [30], NUIQ [32], QMICE [85], UIQI [91], NMFC [89], Yi et al. [86], twice mixing [35], URanker [49], and ATUIQP [57] are compared.

For UIQM, UCIQE, CCF, URQ, Zhao et al., FDUM, UEIQM, NUIQ, QMICE, UIQI, NMFC, and the Yi et al. methods, each UIQD is randomly divided into 80% and 20%, corresponding to training and test sets, respectively. The training/test split is operated 1000 times, and the average performance is reported

for the individual dataset evaluation. Given the necessity for the twice mixing method to utilize original images and their corresponding enhanced versions for training, we employ the Fu et al. dataset [35], URankerSet [49], and UIQD dataset [57], to train twice mixing, URanker, and ATUIQP. The methods are implemented using PyTorch and the experiments are conducted on an NVIDIA 2080 Ti GPU. The initial learning rate is set to $1e-5$, with a decay factor of 0.1 applied every 5 epochs. The network is optimized using the Adam optimizer, and the learning rate is progressively reduced via a StepLR scheduler.

Before assessing the model's performance, we employed the five-parameter logistic function [101] to map the predicted scores onto a discrete scale of MOS values for comparison. This transformation process can be described by the following formula:

$$f(s) = \alpha_1 \left(\frac{1}{2} - \frac{1}{1 + \exp(\alpha_2(s - \alpha_3))} \right) + \alpha_4 s + \alpha_5 \quad (2)$$

where s and $f(s)$ represent the predicted score and the fitted scores through nonlinear regression, respectively, and $\alpha_k (k = 1, 2, \dots, 5)$ are model parameters determined during the fitting process.

3.3.1 | Performance on Different Distortion Levels

To compare the accuracy of different BUIQE algorithms in predicting the degree of underwater absorption and scattering, we first conduct performance comparisons on the UID-LEIA [75] dataset. The L-test is used to evaluate the algorithms' sensitivity to underwater distortion levels. Additionally, the D-test is utilized to assess the algorithms' capability to discriminate between original and distorted images. During the testing process, images captured at various imaging distances and water turbidities are categorized, with those captured at the furthest imaging distance or exhibiting the highest level of water turbidity designated as distorted images, while those captured at the closest imaging distance or displaying the lowest level of water turbidity are marked as original images. The rationale behind opting for the UID-LEIA dataset stems from its uniqueness as the sole distortion level sequence dataset currently equipped with MOSs. The results are shown in Table 2.

As shown in Table 2, UEIQM and Yi et al. methods excel in underwater image quality assessment across different imaging distances, primarily due to their advanced feature extraction and fusion strategies, as well as deep consideration of physical models. Specifically, while UEIQM shares some basic metrics with UIQM, its superior performance in varying imaging distances can be attributed to its use of SVR for feature fusion. This non-linear modeling tool captures complex relationships brought by changes in imaging distance. In contrast, Yi et al. method uses multi-scale analysis to extract features directly from images, adapting to detail variations at different distances and focusing on characteristics like contrast and clarity, thus more accurately reflecting image quality changes.

In scenarios with varying water turbidities, Yi et al. and URanker methods show significant advantages in the L-test, mainly due to

TABLE 2 | Performance comparisons of BUIQE methods on UID-LEIA dataset.

Method	Varying the imaging distances		Varying the water turbidities	
	L-test	D-test	L-test	D-test
UIQM	0.8478	0.6667	-0.4889	0.9111
UCIQE	0.6367	0.7667	-0.1556	0.8889
CCF	0.7422	0.6333	-0.5667	0.7444
URQ	0.7711	0.8000	0.1111	0.5000
Zhao et al.	0.6278	0.5333	0.2333	1.0000
FDUM	0.6556	0.6000	-0.1347	0.6000
UEIQM	0.9967	1.0000	0.2117	0.5778
NUIQ	0.9845	0.8333	0.5111	0.5111
QMICE	0.7411	0.8000	-0.4889	0.7000
UIQI	0.9733	1.0000	0.4333	0.6222
NMFC	0.6533	0.9667	0.3222	0.5222
Yi et al.	0.9878	1.0000	0.5778	0.5333
Twice mixing	0.8122	0.9333	-0.0889	0.7667
URanker	0.9211	0.8667	0.5889	0.5333
ATUIQP	0.8205	0.9444	-0.0889	0.8333

their multi-scale analysis and rank-based learning mechanisms. These methods effectively capture quality variations under different turbidity levels and provide stable relative rankings. The training process of URanker enables it to distinguish images of varying quality, leading to outstanding performance in the L-test. In addition to demonstrating good consistency with subjective MOS values, the ability to differentiate between reference and distorted images is also a critical metric. Table 2 shows that the performance of Yi et al. and URanker in the D-test is relatively average, indicating that researchers can improve the model's performance by focusing on the ability to distinguish between reference images and those with added turbidity.

3.3.2 | Performance on Different UIE Outputs

The subjective assessment of datasets without distortion levels under grouped evaluation mode employs the PC or DS methods on image sets with identical content. As a result, MOS values obtained are only valid within the same content images. By enhancing the quality of underwater images, researchers can process and interpret image data more effectively, providing significant support for a wide range of research and practical applications, such as marine biology, underwater archaeology, and marine resource exploration [102]. We systematically evaluate the performance of the BUIQE algorithms on the SAUD [32] and UID2021 [44] datasets. These UIQDs are organized by the reference image and its corresponding enhanced counterparts. The L-test on image groups with the same content is executed to assess the algorithm's consistency in ranking the enhancement results. Simultaneously, we select the enhanced image with the highest MOS value within the same content as the optimal

enhanced image and employ the D-test to evaluate the algorithm's performance in discriminating between original and enhanced images. The results are presented in Table 3.

As shown in Table 3, NUIQ and NMFC methods exhibit relatively high values in the L-test. Both methods emphasize multi-feature fusion in color space, which helps distinguish subtle differences between images of the same content enhanced by different algorithms, particularly in aspects like color and contrast. In the D-test, NMFC demonstrates superior performance by detecting minute differences between images, especially those that are visually imperceptible. Therefore, NMFC excels in evaluating the effectiveness of different enhancement methods.

3.3.3 | Performance on Ungrouped Evaluation Outputs

Finally, the subjective assessment of datasets without distortion levels under ungrouped evaluation mode employs the single stimulus (SS) method. Such datasets are suitable for assessing the correlation between the predictive values of different BUIQE algorithms and human perception. We evaluate the performance of 15 UIQE methods in terms of the primary metric PLCC and the secondary metric P-test on both UIED [33] and UWIQA [28] datasets. When conducting the P-test, we consider image pairs where the difference in MOSs exceeded the mean plus or minus the standard deviation as image pairs with significant quality differences. The PLCC and P-test of the BUIQE method are presented in Table 4.

Based on the data of Table 4, it is evident that NUIQ and NMFC methods exhibit higher PLCC values on the UWIQA and UIED datasets. Within these two methods, the NUIQ method particularly demonstrates superior performance in the P-test, likely due to its use of carefully designed quality-aware features. By extracting features from color and luminance components, NUIQ better captures color bias, contrast issues, and detail loss in underwater images. This makes the NUIQ method especially effective in assessing quality differences between images with varying content and conditions, notably in addressing color shifts and reduced visibility.

In these experiments, we find that the performance of different methods fluctuates greatly when performed on different application scenarios. This indicates that relying solely on the specific dataset is insufficient to evaluate a BUIQE algorithm. Therefore, we introduce here the QDE method, which encapsulates multiple datasets and various correlation indicators to provide an intuitive and comprehensive evaluation of the BUIQE algorithms.

4 | Underwater Image Quality Quadrilateral Datum

4.1 | Principles of the QDE Method

To more accurately measure the performance of different algorithms under various conditions, we adopt a quadrilateral radar chart to visualize multiple performances of a BUIQE metric, as illustrated in Figure 6. In the radar chart, a_1 and a_2 , respectively, represent the performance of an algorithm on distortion level

TABLE 3 | Performance comparisons of BUIQE methods on SAUD and UID2021 datasets.

Method	L-test			D-test		
	SAUD	UID2021	\bar{X}	SAUD	UID2021	\bar{X}
UIQM	0.3161	0.5614	0.4388	0.5750	0.9167	0.7459
UCIQE	0.4259	0.6517	0.5388	0.7400	0.9750	0.8575
CCF	0.2701	0.6387	0.4544	0.7850	0.9417	0.8634
URQ	0.2595	0.6356	0.4476	0.6450	0.9250	0.7850
Zhao et al.	0.4113	0.2379	0.3246	0.7650	0.5750	0.6700
FDUM	0.5326	0.5819	0.5573	0.7000	0.9833	0.8417
UEIQM	0.8375	0.5973	0.7174	0.8800	0.9167	0.8984
NUIQ	0.8707	0.7848	0.8278	0.8950	0.9153	0.9052
QMICE	0.3546	0.5879	0.4713	0.6150	0.9417	0.7784
UIQI	0.7969	0.7434	0.7702	0.8850	0.9667	0.9259
NMFC	0.8981	0.8057	0.8519	0.9700	0.9417	0.9559
Yi et al.	0.8398	0.6833	0.7616	0.8100	0.9746	0.8923
Twice mixing	0.4098	0.5640	0.4869	0.6600	0.7583	0.7092
URanker	0.5240	0.6578	0.5909	0.8350	0.9500	0.8925
ATUIQP	0.2362	0.5832	0.4097	0.7800	0.8833	0.8317

The \bar{X} represents the mean value.

TABLE 4 | Performance comparisons of BUIQE methods on UIED and UWIQA datasets.

Method	PLCC			P-test		
	UIED	UWIQA	\bar{X}	UIED	UWIQA	\bar{X}
UIQM	0.0617	0.6315	0.3466	0.4739	0.8937	0.6838
UCIQE	0.2289	0.5680	0.3985	0.4987	0.8586	0.6787
CCF	0.2005	0.4528	0.3267	0.5020	0.7978	0.6499
URQ	0.3200	0.1723	0.2462	0.6627	0.9560	0.8094
Zhao et al.	0.1054	0.3193	0.2124	0.4922	0.6396	0.5659
FDUM	0.0560	0.6657	0.3609	0.4847	0.9163	0.7005
UEIQM	0.3138	0.6162	0.4650	0.9603	0.6125	0.7864
NUIQ	0.5822	0.7234	0.6528	0.7472	0.9820	0.8646
QMICE	0.3118	0.6542	0.4830	0.6710	0.9179	0.7945
UIQI	0.5276	0.6876	0.6076	0.6877	0.9776	0.8327
NMFC	0.5823	0.7274	0.6549	0.7409	0.7118	0.7264
Yi et al.	0.5479	0.7006	0.6243	0.7743	0.9372	0.8558
Twice mixing	0.3747	0.4502	0.4125	0.7688	0.7820	0.7754
URanker	0.1963	0.5140	0.3552	0.6091	0.8081	0.7086
ATUIQP	0.1505	0.5490	0.3498	0.5430	0.8561	0.6996

sequence datasets under different imaging distances and water turbidities. The a_3 and a_4 reflect the algorithm's overall performance on undistorted level sequence datasets under grouped and ungrouped evaluation models, respectively. The farther these four indicators are from the origin, the better the performance. When all four vertices lie on the circumference of the circle, it indicates that performance has reached an optimal level. To quantify this

performance, we map these metrics into a four-dimensional space, where a_1 , a_2 , a_3 , and a_4 correspond to components in four dimensions, forming a four-dimensional vector $\vec{a} = (a_1, a_2, a_3, a_4)$. The proposed QDE involves two factors:

$$QDE = \begin{bmatrix} v_{norm} \\ \sigma_{norm} \end{bmatrix} \quad (3)$$

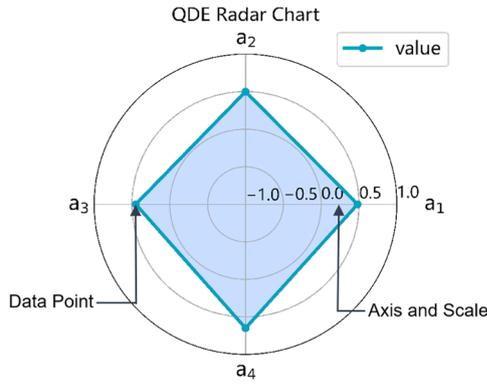


FIGURE 6 | The illustration of the QDE radar chart.

$$v_{\text{norm}} = \frac{v - v_{\min}}{v_{\max} - v_{\min}} \quad (4)$$

$$\sigma_{\text{norm}} = \frac{\sigma - \sigma_{\min}}{\sigma_{\max} - \sigma_{\min}} \quad (5)$$

$$v = \|\vec{a}\| = \sqrt{\sum_{i=1}^4 (a_i + 1)^2} \quad (6)$$

$$\sigma = \sqrt{\frac{1}{4} \sum_{i=1}^4 (a_i - \mu)^2} \quad (7)$$

$$\mu = \frac{1}{4} \sum_{i=1}^4 a_i \quad (8)$$

where the feature intensity (v) reflects the length of a four-dimensional vector, which means calculating the norm $\|\vec{a}\|$ of the vector. The imbalance (σ) is measured by the standard deviation of the four-dimensional vector. To ensure comparability of algorithm performance across different datasets and better reflect consistency with subjective evaluation, both the feature intensity v and imbalance σ need to be normalized. Specifically, by calculation, v_{\max} and v_{\min} are found to be 4 and 1, respectively, while σ_{\max} and σ_{\min} are 0.75 and 0. For a QDE where the v_{norm} is larger and σ_{norm} is smaller, it indicates that the algorithm performs well in terms of consistency with subjective evaluation when it comes to image quality at various distortion levels, comparing image quality from different UIE outputs, and assessing the quality of different content images.

Considering that the PLCC and L-test values range from -1 to 1, we set the origin of the QDE map to -1, as shown in Figure 6. μ is the mean of the four indicators. Each a_i , $i = 1, 2, 3, 4$, is computed from two sub-indicators, b_{i1} and b_{i2} , as follows:

$$a_i = w_1 \times b_{i1} + w_2 \times |b_{i1}| \times b_{i2} \quad (9)$$

where w_1, w_2 are weighted coefficients.

4.2 | Applicability in Real-World Scenarios

The core advantage of the QDE method lies in its integrated evaluation and dynamic adjustment capabilities. QDE can adjust the types of datasets and performance metrics according to the specific task, ensuring that the evaluation results are targeted

and accurate. This is especially important for complex application scenarios that involve diverse types of images. To illustrate the applicability of the QDE method, we use an underwater archaeology project as an example. At the same archaeological site, archaeologists may find various types of artifacts, such as bronze and porcelain. To observe the details more clearly, they often enhance the original images to improve clarity and visibility. By comparing the quality of enhanced images with different contents (e.g., bronze and porcelain), researchers can select the most suitable images for further analysis and documentation, thus improving the efficiency and accuracy of archaeological work. In this case, the evaluation can be carried out using two types of datasets: the undistorted grade sequence dataset (for different content images) and the undistorted grade sequence dataset (for the same content enhanced images). To adapt to the requirements of specific tasks, the QDE calculation method has been adjusted accordingly, changing from evaluating the feature intensity and imbalance of four-dimensional vectors to that of two-dimensional vectors. This adjustment enables QDE to better suit specific application scenarios, ensuring that the evaluation results are more accurate and reliable.

4.3 | Analysis of Weight Allocation Schemes

To comprehensively evaluate the performance of the BUIQE algorithms on various datasets, we assign different w_1 and w_2 , ensuring that when the main indicator b_{i1} is the same, algorithms with higher performance on the secondary indicator b_{i2} are favored. However, in cases where b_{i1} is small and b_{i2} is large, assigning a high weight to b_{i1} and a low weight to b_{i2} significantly affects the criterion. To address this situation, we opted for using the $|b_{i1}| \times b_{i2}$ instead of b_{i2} as the secondary indicator.

We employ analysis of variance (ANOVA)[103] to assess whether significant performance differences exist among different weight allocation methods. The null hypothesis (H_0) assumes no significant differences among these methods, while the alternative hypothesis (H_1) suggests the presence of a significant difference. By calculating the $F_{\text{statistic}}$ [104], ANOVA compares differences both between and within groups. The formula is shown in Equation (8). A higher $F_{\text{statistic}}$ supports the H_1 , indicating a significant difference. Simultaneously, the P_{value} is utilized to assess the significance of observed differences. Typically, we set the significance level at 0.05. If the calculated P_{value} is below this threshold, we reject the H_0 and accept the H_1 , signifying a noteworthy difference in weight allocation methods. Conversely, if the P_{value} exceeds the significance level, we cannot reject the H_0 and we cannot conclude that there is a significant difference in group means.

$$F_{\text{statistic}} = \frac{SSB/p - 1}{SSW/(n - p)} \quad (10)$$

where $F_{\text{statistic}}$ mainly includes the sum of squares between groups (SSB), the sum of squares within groups (SSW), and their respective degrees of freedom. The p and n represent the number of groups and the total number of samples, respectively.

The weight allocation preselection schemes are presented in Table 5. Our experiments are divided into two parts. Firstly, in

TABLE 5 | Weight allocation preselection schemes.

Serial number	Indicator 1	w_1	Indicator 2	w_2
Scheme 1	b_{i1}	0.5	b_{i2}	0.5
Scheme 2	b_{i1}	0.5	$ b_{i1} \times b_{i2}$	0.5
Scheme 3	b_{i1}	0.6	$ b_{i1} \times b_{i2}$	0.4
Scheme 4	b_{i1}	0.7	$ b_{i1} \times b_{i2}$	0.3
Scheme 5	b_{i1}	0.8	$ b_{i1} \times b_{i2}$	0.2
Scheme 6	b_{i1}	0.9	$ b_{i1} \times b_{i2}$	0.1

TABLE 6 | Experiment results of the ANOVA.

Experiment number	Serial number	$F_{statistic}$	P_{value}
Experiment 1	Schemes 1-2	6.9865	0.0177
Experiment 2	Schemes 2-6	0.0246	0.9988

Experiment 1, one-way ANOVA is conducted on Scheme 1 and Scheme 2 to demonstrate that $|b_{i1}| \times b_{i2}$ is more reasonable than b_{i2} as a secondary indicator when testing the merits of the BUIQE method. We deliberately choose small values for b_{i1} and large values for b_{i2} , and opt for two specific weight allocation schemes: Schemes 1 and 2. Additionally, to comprehensively evaluate the performance of different weight allocation methods, Experiment 2 is carried out, conducting one-way ANOVA for Schemes 2–6 to analyze the rationality of these weight allocation methods. The results of the two experiments are shown in Table 6.

It can be observed from Table 6 that the $F_{statistic}$ of Experiment 1 is large, and the P_{value} is less than 0.5. This supports H1, suggesting a significant performance difference between the two weight allocation methods. As mentioned earlier, using $|b_{i1}| \times b_{i2}$ as a complement to b_{i2} helped to avoid unfavorable impacts when b_{i1} is small and b_{i2} is large in a composite value. Therefore, Scheme 2 was superior to Scheme 1. In Experiment 2, the $F_{statistic}$ is relatively small, and the P_{value} is large. We are unable to reject the H_0 , indicating that in this study, there is not sufficient statistical evidence to support significant differences in performance among the five weight allocation methods.

Therefore, we conclude that these weight allocation methods exhibited similar performance, and it can not be determined that any one method is statistically significantly superior to the others. In the context of experimental data, we opt for Scheme 2, with weight settings of 0.5 for b_{i1} and 0.5 for $|b_{i1}| \times b_{i2}$. The rationale behind this choice lay in emphasizing the importance of the primary indicator b_{i1} , maintaining relative balance in weight settings in the absence of rejecting the H_0 , and simultaneously accounting for the influence of the supplemental term $|b_{i1}| \times b_{i2}$.

4.4 | Experiment Results

We conduct experiments comparing 15 state-of-the-art BUIQE methods. The QDE values and the four indicators are shown in Tables 7 and 8. To enhance the interpretability of the data and provide an intuitive visualization of the performance of each BUIQE

TABLE 7 | Four indicators of the QDE radar charts.

Method		α_1	α_2	α_3	α_4
Head-crafted BUIQE	UIQM	0.7065	−0.0217	0.3831	0.2918
	UCIQE	0.5624	−0.0086	0.5004	0.3345
	CCF	0.6061	−0.0724	0.4234	0.2695
	URQ	0.6940	0.0833	0.3995	0.2227
	Zhao et al.	0.4813	0.2333	0.2710	0.1663
	FDUM	0.5245	−0.0269	0.5132	0.3069
	UEIQM	0.9967	0.1670	0.6810	0.4153
	NUIQ	0.9024	0.3862	0.7886	0.6086
	QMICE	0.6670	−0.0733	0.4191	0.4334
	UIQI	0.9733	0.3514	0.7417	0.5568
Deep learned BUIQE	NMFC	0.6424	0.2452	0.8331	0.5653
	Yi et al.	0.9878	0.4430	0.7206	0.5793
	Twice Mixing	0.7851	−0.0104	0.4161	0.3662
	URanker	0.8597	0.4515	0.5591	0.3034
	ATUIQP	0.7977	−0.0074	0.3752	0.2973

TABLE 8 | Results of the QDE method.

Number	Method	v_{norm}	σ_{norm}
1	UIQM [24]	0.5765	0.3460
2	UCIQE [25]	0.5769	0.2956
3	CCF [26]	0.5535	0.3323
4	URQ [81]	0.5793	0.3041
5	Zhao et al. [82]	0.5289	0.1569
6	FDUM [28]	0.5654	0.2976
7	UEIQM [30]	0.7301	0.4113
8	NUIQ [32]	0.7885	0.2603
9	QMICE [85]	0.5920	0.3596
10	UIQI [91]	0.7811	0.3060
11	NMFC [89]	0.7238	0.2828
12	Yi et al. [86]	0.7965	0.2689
13	Twice mixing [35]	0.6117	0.3757
14	URanker [49]	0.7046	0.2719
15	ATUIQP [57]	0.5971	0.3833

method in terms of QDE measurement, we employ two graphical representation approaches: radar charts, which comprehensively reflect the overall characteristics of each method, and polar coordinate plots, where the radius represents the normalized feature strength v_{norm} , the angle is determined by $((1 - \sigma_{norm}) \times 360^\circ)$, and the angles of all methods are proportionally distributed in the overall polar coordinate diagram for unified visualization. This design allows for a clearer comparison of the performance across different methods. The corresponding figures are located in Figures 7 and 8. Observational data clearly indicate that the

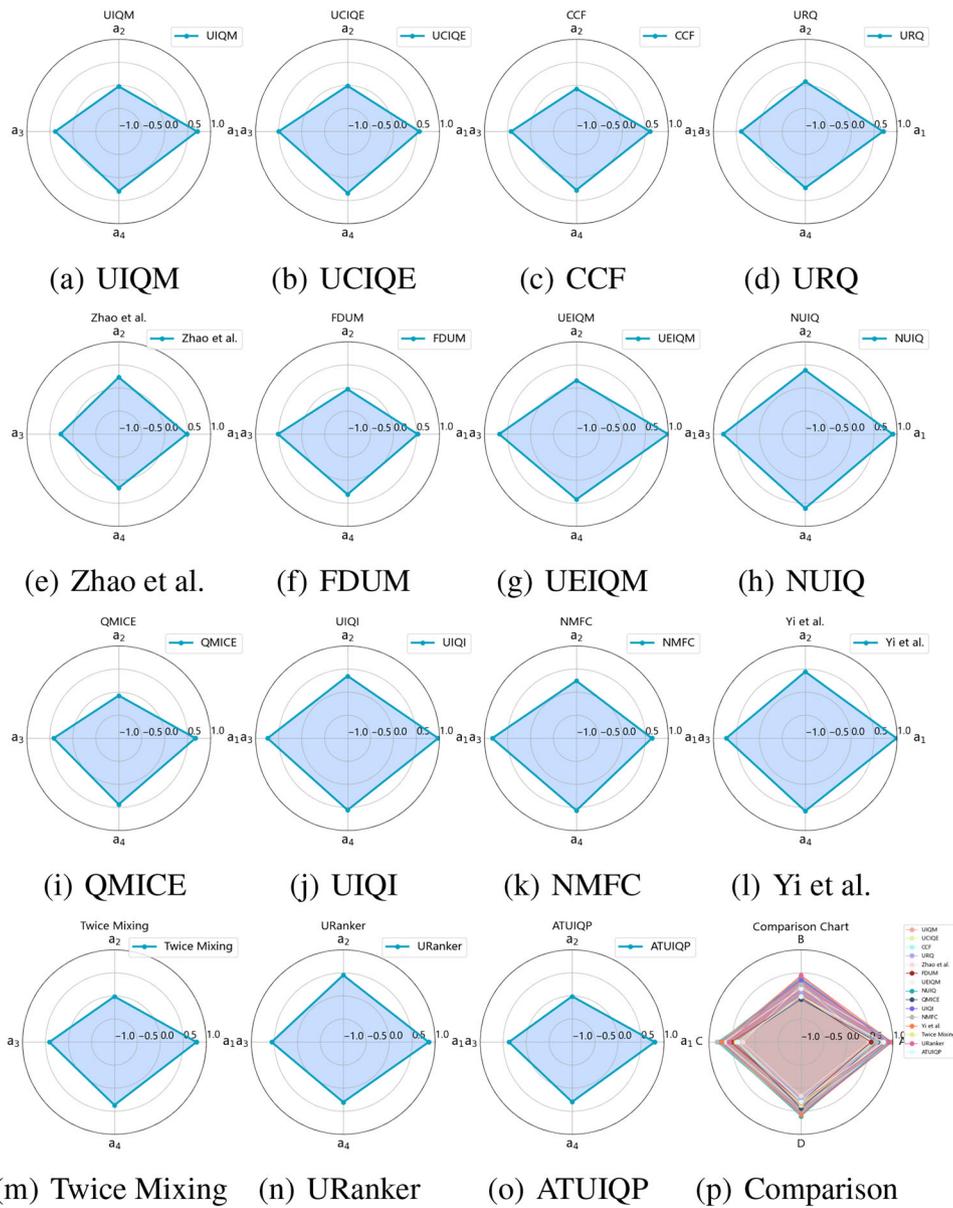


FIGURE 7 | The QDE radar charts of BUIQE methods.

Yi et al. [86] and NUIQ [32] methods excel in feature strength v , achieving scores of 0.7965 and 0.7885, respectively, ranking them top among all evaluated UIQA methods. Notably, while Zhao et al. [82] method excels in addressing data imbalance, the Yi et al. and NUIQ methods closely follow, demonstrating their relatively balanced performance across various aspects. Comprehensive analysis of data from Figure 7, Figure 8, Table 7, and Table 8 reveals that Yi et al. and NUIQ methods not only rank at the forefront in performance evaluations across four different datasets but also this superior performance resonates with their high QDE values.

Considering all UIQE methods based on the findings in Table 7 and Figure 7, it is evident that the current UIQE methods generally perform well in evaluating image sets with varying imaging distances, but exhibit poorer accuracy in assessing image sets with changes in water turbidities and image sets with different content. Further examinations of Tables 3 and 4 reveal that existing UIQE

methods demonstrate good consistency in assessing the ranking of enhanced image sets across different datasets. However, there is a significant disparity in performance in the D-test, suggesting insufficient stability in accurately distinguishing between original and enhanced images. When evaluating image sets with different content, while the correlation between UIQE-predicted image quality values and MOSs is relatively high, the accuracy in assessing image pairs with significant quality differences remains inconsistent. This further confirms the ability to assess image quality using QDE across various scenarios.

While the experimental results demonstrate the effectiveness of the QDE method in evaluating underwater image quality, it is not without limitations. For instance, the method's performance can be influenced by the specific characteristics of the datasets, potentially affecting its generalizability to other scenarios. Additionally, the computational complexity of the QDE framework may pose challenges for real-time applications. These findings

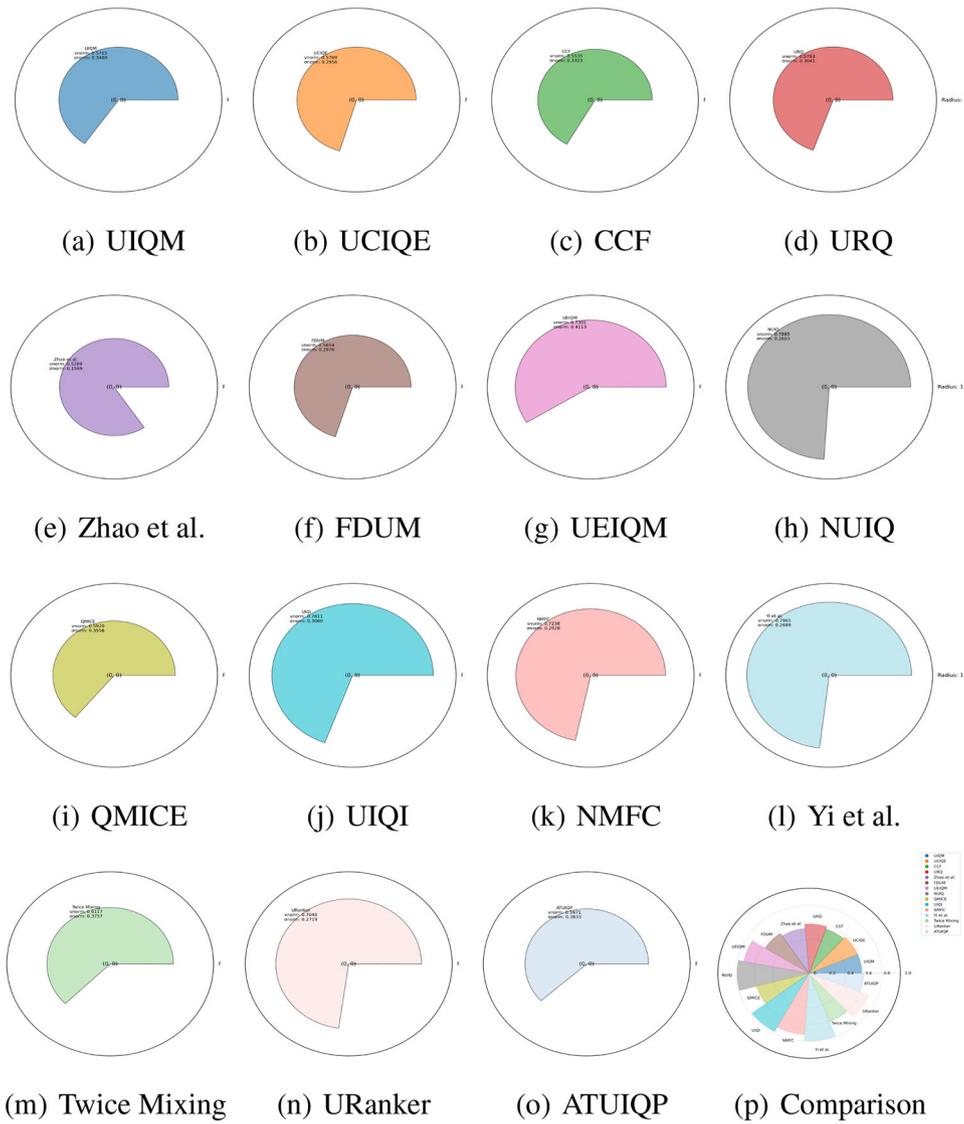


FIGURE 8 | The polar coordinate plots of BUIQE methods.

suggest potential directions for future improvement, including reducing dependency on dataset-specific features and enhancing computational efficiency.

5 | Conclusion

This work presents a comprehensive overview of the latest developments in the field of UIQE, particularly delving into the progress of UIQE algorithms from the perspective of novel dataset classification. By discussing the image features, fitting methods, and evaluation procedures of 37 UIQE algorithms, researchers gain deeper insights into the advancements of each algorithm. Additionally, the introduced QDE performance metric offers a holistic evaluation of UIQE algorithm performance across various underwater scenarios, which is significant for domains such as underwater surveillance deployment and underwater archaeology, providing guidance and reference for relevant applications. This metric integrates performance indicators from multiple scenarios and validates the rationality of weight distribution through variance analysis. Moreover, the visibility of

the quadrilateral radar charts and polar coordinate plots aids in better understanding the performance of UIQE algorithms in diverse scenarios, facilitating targeted improvements to address algorithmic weaknesses and further enhancing performance and applicability.

The QDE method further offers a comprehensive and adaptable framework for UIQE across diverse underwater scenarios. Its ability to integrate task-specific datasets and metrics enhances its practical utility, making it a significant contribution to underwater image quality evaluation research. In the future, this method can be further extended to more complex underwater environments, ensuring its robustness and scalability in real-world applications.

Author Contributions

Mengjiao Shen: conceptualization, investigation, methodology, validation, visualization, writing – original draft. **Miao Yang:** supervision, validation, writing – review & editing. **Jinyang Zhong:** investigation,

supervision, validation. **Hantao Liu:** supervision, validation. **Can Pan:** supervision, validation.

Acknowledgements

This work was supported in part by National Natural Science Foundation of China (NSFC) under Grant 62271236, National Key R&D Program Project 2023YFC3108205, Key Country-Specific Industrial Technology R&D Cooperation Project 23GH002, Graduate Research and Practice Innovation Program under Grants KYCX2023-67 and KYCX23-3462.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data available on request from the authors.

References

1. J. Zhou, B. Li, and D. Zhang, et al., "UGIF-net: An Efficient Fully Guided Information Flow Network for Underwater Image Enhancement," *IEEE Transactions on Geoscience and Remote Sensing* 61 (2023): 4206117.
2. M. Yang, Z. Sheng, G. Yin, and H. Wang, "A Recurrent Neural Network Based Fuzzy Sliding Mode Control for 4-DOF ROV Movements," *Ocean Engineering* 256 (2022): 111509.
3. M. Yang, H. Wang, K. Hu, G. Yin, and Z. Wei, "IA-Net: An Inception-Attention-Module-Based Network for Classifying Underwater Images from Others," *IEEE Journal of Oceanic Engineering* 47, no. 3 (2022): 704–717.
4. J. Zhou, D. Zhang, and W. Zhang, "Cross-View Enhancement Network for Underwater Images," *Engineering Applications of Artificial Intelligence* 121 (2023): 105952.
5. J. Fan, X. Wang, C. Zhou, Y. Ou, F. Jing, and Z. Hou, "Development, Calibration, and Image Processing of Underwater Structured Light Vision System: A Survey," *IEEE Transactions on Instrumentation and Measurement* 72 (2023): 1–18, <https://doi.org/10.1109/TIM.2023.3235420>.
6. A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-Reference Image Quality Assessment in the Spatial Domain," *IEEE Transactions on Image Processing* 21, no. 12 (2012): 4695–4708.
7. A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "Completely Blind" Image Quality Analyzer," *IEEE Signal Processing Letters* 20, no. 3 (2012): 209–212.
8. Q. Wu, H. Li, K. N. Ngan, B. Zeng, and M. Gabbouj, "No Reference Image Quality Metric via Distortion Identification and Multi-Channel Label Transfer," in *2014 IEEE International Symposium on Circuits and Systems (ISCAS)* (IEEE, 2014), 530–533.
9. L. Zhang, L. Zhang, and A. Bovik, "A Feature-Enriched Completely Blind Image Quality Evaluator," *IEEE Transactions on Image Processing* 24, no. 8 (2015): 2579–2591.
10. Y. Liu, K. Gu, and Y. Zhang, et al., "Unsupervised Blind Image Quality Evaluation via Statistical Measurements of Structure, Naturalness, and Perception," *IEEE Transactions on Circuits and Systems for Video Technology* 30, no. 4 (2019): 929–943.
11. W. Hou, X. Gao, D. Tao, and X. Li, "Blind Image Quality Assessment via Deep Learning," *IEEE Transactions on Neural Networks and Learning Systems* 26, no. 6 (2014): 1275–1286.
12. X. Liu, J. Van De Weijer, and A. Bagdanov, "Rankiq: Learning from Rankings for No-Reference Image Quality Assessment," in *Proceedings of the IEEE International Conference on Computer Vision* (IEEE, 2017), 1040–1049.
13. W. Zhang, K. Ma, J. Yan, D. Deng, and Z. Wang, "Blind Image Quality Assessment Using a Deep Bilinear Convolutional Neural Network," *IEEE Transactions on Circuits and Systems for Video Technology* 30, no. 1 (2018): 36–47.
14. S. Su, Q. Yan, and Y. Zhu, et al., "Blindly Assess Image Quality in the Wild Guided by a Self-Adaptive Hyper Network," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE, 2020), 3667–3676.
15. W. Zhang, K. Ma, G. Zhai, and X. Yang, "Uncertainty-Aware Blind Image Quality Assessment in the Laboratory and Wild," *IEEE Transactions on Image Processing* 30 (2021): 3474–3486.
16. Z. Pan, F. Yuan, X. Wang, L. Xu, X. Shao, and S. Kwong, "No-Reference Image Quality Assessment via Multibranch Convolutional Neural Networks," *IEEE Transactions on Artificial Intelligence* 4, no. 1 (2022): 148–160.
17. P. Zhao, X. Chen, V. Chung, and H. Li, "DeLFIQE—A Low-Complexity Deep Learning-Based Light Field Image Quality Evaluator," *IEEE Transactions on Instrumentation and Measurement* 70 (2021): 1–11.
18. L. Shen, B. Zhao, Z. Pan, B. Peng, S. Kwong, and J. Lei, "Channel Recombination and Projection Network for Blind Image Quality Measurement," *IEEE Transactions on Instrumentation and Measurement* 71 (2022): 1–12.
19. M. Yang, Z. Xie, J. Dong, H. Liu, H. Wang, and M. Shen, "Distortion-Independent Pairwise Underwater Image Perceptual Quality Comparison," *IEEE Transactions on Instrumentation and Measurement* 72 (2023): 5024415.
20. Y. Y. Schechner and N. Karpel, "Recovery of Underwater Visibility and Structure by Polarization Analysis," *IEEE Journal of Oceanic Engineering* 30, no. 3 (2005): 570–587.
21. M. Arredondo and K. Lebart, "A Methodology for the Systematic Assessment of Underwater Video Processing Algorithms," in *Europe Oceans 2005 1* (IEEE, 2005), 362–367.
22. W. Hou, A. D. Weidemann, D. J. Gray, and G. Fournier, "Imagery-Derived Modulation Transfer Function and Its Applications for Underwater Imaging," in *Applications of Digital Image Processing XXX* (SPIE, 2007), 707–714.
23. W. Hou and A. Weidemann, "Objectively Assessing Underwater Image Quality for the Purpose of Automated Restoration," in *Visual Information Processing XVI* (SPIE, 2007), 171–177.
24. K. Panetta, C. Gao, and S. Agaian, "Human-Visual-System-Inspired Underwater Image Quality Measures," *IEEE Journal of Oceanic Engineering* 41, no. 3 (2015): 541–551.
25. M. Yang and A. Sowmya, "An Underwater Color Image Quality Evaluation Metric," *IEEE Transactions on Image Processing* 24, no. 12 (2015): 6062–6071.
26. Y. Wang, N. Li, and Z. Li, et al., "An Imaging-Inspired No-Reference Underwater Color Image Quality Assessment Metric," *Computers & Electrical Engineering* 70 (2018): 904–913.
27. D. Wu, F. Yuan, and E. Cheng, "Underwater No-Reference Image Quality Assessment for Display Module of ROV," *Scientific Programming* 2020 (2020): 1–15.
28. N. Yang, Q. Zhong, K. Li, R. Cong, Y. Zhao, and S. Kwong, "A Reference-Free Underwater Image Quality Assessment Metric in Frequency Domain," *Signal Processing: Image Communication* 94 (2021): 116218.
29. M. Irshad, Sanchez-C. Ferreira, S. Alamgeer, C. H. Llanos, and M. Farias, "No-Reference Image Quality Assessment of Underwater Images Using Multi-Scale Salient Local Binary Patterns," *Electronic Imaging* 2021, no. 9 (2021): 265-1–265-7.
30. P. Guo, L. He, S. Liu, D. Zeng, and H. Liu, "Underwater Image Quality Assessment: Subjective and Objective Methods," *IEEE Transactions on Multimedia* 24 (2021): 1980–1989.

31. P. Guo, H. Liu, D. Zeng, T. Xiang, L. Li, and K. Gu, "An Underwater Image Quality Assessment Metric," *IEEE Transactions on Multimedia* 25 (2023): 5093–5106.
32. Q. Jiang, Y. Gu, C. Li, R. Cong, and F. Shao, "Underwater Image Enhancement Quality Evaluation: Benchmark Dataset and Objective Metric," *IEEE Transactions on Circuits and Systems for Video Technology* 32, no. 9 (2022): 5959–5974.
33. Y. Zheng, W. Chen, R. Lin, T. Zhao, and P. Le Callet, "UIF: An Objective Quality Assessment for Underwater Image Enhancement," *IEEE Transactions on Image Processing* 31 (2022): 5456–5468.
34. M. Yang, G. Yin, Y. Du, and H. Wang, "Multitopic Underwater Image Quality Assessment With Visual Attention Factors," *Journal of Electronic Imaging* 31, no. 2 (2022): 023020.
35. Z. Fu, X. Fu, Y. Huang, and X. Ding, "Twice Mixing: A Rank Learning Based Quality Assessment Approach for Underwater Image Enhancement," *Signal Processing: Image Communication* 102 (2022): 116622.
36. K. Ma, Z. Duanmu, and Q. Wu, et al., "Waterloo Exploration Database: New Challenges for Image Quality Assessment Models," *IEEE Transactions on Image Processing* 26, no. 2 (2016): 1004–1016.
37. Z. Wang and E. Simoncelli, "Maximum Differentiation (MAD) Competition: A Methodology for Comparing Computational Models of Perceptual Quantities," *Journal of Vision* 8, no. 12 (2008): 8–8.
38. K. Ma, Z. Duanmu, and Z. Wang, et al., "Group Maximum Differentiation Competition: Model Comparison With Few Samples," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42, no. 4 (2018): 851–864.
39. S. Raveendran, M. D. Patil, and G. Birajdar, "Underwater Image Enhancement: A Comprehensive Review, Recent Trends, Challenges and Applications," *Artificial Intelligence Review* 54 (2021): 5413–5467.
40. M. Wang, K. Zhang, H. Wei, W. Chen, and T. Zhao, "Underwater Image Quality Optimization: Researches, Challenges, and Future Trends," *Image and Vision Computing* 146 (2024): 104995.
41. X. Xu, H. Cai, M. Wang, W. Chen, R. Zhang, and T. Zhao, "Exploring Underwater Image Quality: A Review of Current Methodologies and Emerging Trends," *Image and Vision Computing* 154 (2024): 105389.
42. J. Jaffe, "Computer Modeling and the Design of Optimal Underwater Imaging Systems," *IEEE Journal of Oceanic Engineering* 15, no. 2 (1990): 101–111.
43. H. Lin, V. Hosu, and D. Saupe, "KADID-10k: A Large-Scale Artificially Distorted IQA Database," in *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)* (IEEE, 2019), 1–3.
44. G. Hou, Y. Li, H. Yang, K. Li, and Z. Pan, "UID2021: An Underwater Image Dataset for Evaluation of No-Reference Quality Assessment Metrics," *ACM Transactions on Multimedia Computing, Communications and Applications* 19, no. 4 (2023): 1–24.
45. Z. Chen, T. Jiang, and Y. Tian, "Quality Assessment for Comparing Image Enhancement Algorithms," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2014), 3003–3010.
46. C. Li, C. Guo, and W. Ren, et al., "An Underwater Image Enhancement Benchmark Dataset and Beyond," *IEEE Transactions on Image Processing* 29 (2019): 4376–4389.
47. Y. Gu, Q. Jiang, F. Shao, and W. Gao, "A Real-World Quality Evaluation Dataset for Enhanced Underwater Images," *Journal of Image and Graphics* 27 (2022): 1467–1480.
48. W. Li, C. Lin, T. Luo, H. Li, H. Xu, and L. Wang, "Subjective and Objective Quality Evaluation for Underwater Image Enhancement and Restoration," *Symmetry* 14, no. 3 (2022): 558.
49. C. Guo, R. Wu, and X. Jin, et al., "Underwater Ranker: Learn Which is Better and How to be Better," in *Proceedings of the AAAI Conference on Artificial Intelligence* 37 (2023): 702–709.
50. V. Hosu, H. Lin, T. Sziranyi, and D. Saupe, "KonIQ-10k: An Ecologically Valid Database for Deep Learning of Blind Image Quality Assessment," *IEEE Transactions on Image Processing* 29 (2020): 4041–4056.
51. R. Lin, T. Zhao, W. Chen, Y. Zheng, and H. Wei, "Underwater Image Quality Database Towards Fish Detection," in *2021 IEEE/CIC International Conference on Communications in China (ICCCWorkshops)* (IEEE, 2021), 205–210.
52. M. Li, Y. Lin, L. Shen, Z. Wang, K. Wang, and Z. Wang, "Human Perceptual Quality Driven Underwater Image Enhancement Framework," *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022): 1–15.
53. W. Chen, R. Lin, H. Liao, T. Zhao, K. Gu, and P. Callet, "Utility-Oriented Underwater Image Quality Assessment Based on Transfer Learning," arXiv preprint, arXiv:2205.03574 (2022).
54. F. Lei, S. Li, S. Xie, and J. Liu, "Subjective and Objective Quality Assessment of Swimming Pool Images," *Frontiers in Neuroscience* 15 (2022): 766762.
55. X. Li, H. Xu, and G. Jiang, et al., "Underwater Image Quality Assessment from Synthetic to Real-World: Dataset and Objective Method," *ACM Transactions on Multimedia Computing, Communications and Applications* 20, no. 3 (2023): 1–23.
56. L. Peng, C. Zhu, and L. Bian, "U-Shape Transformer for Underwater Image Enhancement," *IEEE Transactions on Image Processing* 32 (2023): 3066–3079.
57. Y. Liu, B. Zhang, R. Hu, K. Gu, G. Zhai, and J. Dong, "Underwater Image Quality Assessment: Benchmark Database and Objective Method," *IEEE Transactions on Multimedia* 26 (2024): 7734–7747.
58. P. Zhuang, C. Li, and J. Wu, "Bayesian Retinex Underwater Image Enhancement," *Engineering Applications of Artificial Intelligence* 101 (2021): 104171.
59. C. O. Ancuti, C. Ancuti, C. D. Vleeschouwer, and P. Bekaert, "Color Balance and Fusion for Underwater Image Enhancement," *IEEE Transactions on Image Processing* 27 (2018): 379–393.
60. K. Iqbal, R. A. Salam, A. Osman, and A. Z. Talib, "Underwater Image Enhancement Using an Integrated Colour Model," *IAENG International Journal of Computer Science* 34, no. 2 (2007).
61. X. Fu and X. Cao, "Underwater Image Enhancement With Global-Local Networks and Compressed-Histogram Equalization," *Signal Processing: Image Communication* 86 (2020): 115892.
62. G. Hou, Z. Pan, B. Huang, G. Wang, and X. Luan, "Hue Preserving-Based Approach for Underwater Colour Image Enhancement," *IET Image Process* 12 (2018): 292–298.
63. Y. T. Peng and P. Cosman, "Underwater Image Restoration Based on Image Blurriness and Light Absorption," *IEEE Transactions on Image Processing* 26 (2017): 1579–1594.
64. T. P. Marques and A. Albu, "L²UWE : A Framework for the Efficient Enhancement of Low-Light Underwater Images Using Local Contrast and Multi-Scale Fusion," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (IEEE, 2020), 2286–2295.
65. A. Galdran, D. Pardo, A. Picón, and A. Alvarez-Gila, "Automatic Red-Channel Underwater Image Restoration," *Journal of Visual Communication and Image Representation* 26 (2015): 132–145.
66. X. Fu, Z. Fan, M. Ling, Y. Huang, and X. Ding, "Two-Step Approach for Single Underwater Image Enhancement," in *2017 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)* (IEEE, 2017), 789–794.
67. C. Li, S. Anwar, J. Hou, R. Cong, C. Guo, and W. Ren, "Underwater Image Enhancement via Medium Transmission-Guided Multi-Color Space Embedding," *IEEE Transactions on Image Processing* 30 (2021): 4985–5000.

68. J. Xie, G. Hou, G. Wang, and Z. Pan, "A Variational Framework for Underwater Image Dehazing and Deblurring," *IEEE Transactions on Circuits and Systems for Video Technology* 32 (2021): 3514–3526.
69. G. Hou, J. Li, G. Wang, H. Yang, B. Huang, and Z. Pan, "A Novel Dark Channel Prior Guided Variational Framework for Underwater Image Restoration," *Journal of Visual Communication and Image Representation* 66 (2020): 102732.
70. X. Li, G. Hou, L. Tan, and W. Liu, "A Hybrid Framework for Underwater Image Enhancement," *IEEE Access* 8 (2020): 197448–197462.
71. C. Li, S. Anwar, and F. Porikli, "Underwater Scene Prior Inspired Deep Underwater Image and Video Enhancement," *Pattern Recognition* 98 (2020): 107038.
72. X. Fu, P. Zhuang, Y. Huang, Y. Liao, X. Zhang, and X. Ding, "A Retinex-Based Enhancing Approach for Single Underwater Image," in *2014 IEEE International Conference on Image Processing (ICIP)* (IEEE, 2014), 4572–4576.
73. A. Duarte, F. Codevilla, J. D. O. Gaya, and S. Botelho, "A Dataset to Evaluate Underwater Image Restoration Methods," in *OCEANS 2016-Shanghai* (IEEE, 2016), 1–6.
74. Y. Ma, X. Feng, L. Chao, D. Huang, Z. Xia, and X. Jiang, "A New Database for Evaluating Underwater Image Processing Methods," in *2018 Eighth International Conference on Image Processing Theory, Tools and Applications (IPTA)* (IEEE, 2018), 1–6.
75. C. Sánchez-Ferreira, L. Coelho, H. V. Ayala, M. C. Farias, and C. Llanos, "Bio-Inspired Optimization Algorithms for Real Underwater Image Restoration," *Signal Processing: Image Communication* 77 (2019): 49–65.
76. M. Yang, G. Yin, H. Wang, J. Dong, Z. Xie, and B. Zheng, "A Underwater Sequence Image Dataset for Sharpness and Color Analysis," *Sensors* 22, no. 9 (2022): 3550.
77. W. Liu, R. Cui, Y. Li, and S. Zhang, "Hybrid-Input Convolutional Neural Network-Based Underwater Image Quality Assessment," *IEEE Transactions on Neural Networks and Learning Systems* 36, no. 1 (2025): 1790–1798.
78. F. Ou, Y. Wang, J. Li, G. Zhu, and S. Kwong, "A Novel Rank Learning Based No-Reference Image Quality Assessment Method," *IEEE Transactions on Multimedia* 24 (2021): 4197–4211.
79. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing* 13, no. 4 (2004): 600–612.
80. Z. Li, Z. Gu, H. Zheng, B. Zheng, and J. Liu, "Underwater Image Sharpness Assessment Based on Selective Attenuation of Color in the Water," in *OCEANS 2016-Shanghai* (IEEE, 2016), 1–4.
81. L. Li, "Underwater Color Image Quality Assessment," (2019).
82. X. Zhao, G. Hou, Z. Pan, J. Li, and G. Wang, "A No-reference Underwater Image Quality Assessment Method Based on HVS," *Computer Engineering* 46, no. 07 (2020): 235–242, <https://doi.org/10.19678/j.issn.1000-3428.0055324>.
83. S. Tang, C. Li, and Q. Tian, "Underwater Image Quality Assessment Based on Human Visual System," in *2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)* (IEEE, 2020), 378–382.
84. J. Guo, "Underwater Image Quality Evaluation based on Human Visual System," in *2021 International Conference on Big Data, Artificial Intelligence and Risk Management (ICBAR)* (IEEE, 2021), 115–119.
85. Z. Luo, Z. Tang, L. Jiang, and C. Wang, "An Underwater-Imaging-Model-Inspired No-Reference Quality Metric for Images in Multi-Colored Environments," *Expert Systems With Applications* 191 (2022): 116361.
86. X. Yi, Q. Jiang, and W. Zhou, "No-Reference Quality Assessment of Underwater Image Enhancement," *Displays* 81 (2024): 102586.
87. Y. Li, G. Hou, Y. Hao, M. Wang, S. Zhang, and Z. Pan, "A Training-Based Blind Underwater Image Quality Evaluation Metric," in *Proceedings of the 5th International Conference on Big Data Technologies* (Springer, 2022), 236–241.
88. X. Li, H. Xu, and G. Jiang, et al., "Underwater Image Quality Assessment Based on Multiscale and Antagonistic Energy," *IEEE Transactions on Instrumentation and Measurement* 73 (2024): 5000814.
89. T. Chen, X. Yang, N. Li, T. Wang, and G. Ji, "Underwater Image Quality Assessment Method Based on Color Space Multi-Feature Fusion," *Scientific Reports* 13, no. 1 (2023): 16838.
90. S. Zhang, Y. Li, L. Tan, H. Yang, and G. Hou, "A No-Reference Underwater Image Quality Evaluator via Quality-Aware Features," *Journal of Visual Communication and Image Representation* 97 (2023): 103979.
91. Y. Liu, K. Gu, and J. Cao, et al., "UIQI: A Comprehensive Quality Evaluation Index for Underwater Images," *IEEE Transactions on Multimedia* 26 (2024): 2560–2573.
92. Z. Wang, L. Shen, Z. Wang, Y. Lin, and Y. Jin, "Generation-Based Joint Luminance-Chrominance Learning for Underwater Image Quality Assessment," *IEEE Transactions on Circuits and Systems for Video Technology* 33, no. 3 (2022): 1123–1139.
93. I. Goodfellow, P.-J. Abadie, M. Mirza, et al., "Generative Adversarial Nets," in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., 2014).
94. Z. Wang, L. Shen, M. Xu, M. Yu, K. Wang, and Y. Lin, "Domain Adaptation for Underwater Image Enhancement," *IEEE Transactions on Image Processing* 32 (2023): 1442–1457.
95. X. Chu, R. Hu, Y. Liu, J. Cao, and L. Xu, "SISC: A Feature Interaction-Based Metric for Underwater Image Quality Assessment," *IEEE Journal of Oceanic Engineering* 49, no. 2 (2024): 637–648.
96. F. Liu, Z. Huang, T. Xie, R. Hu, and B. Qi, "Enhancing Underwater Image Quality Assessment With Influential Perceptual Features," *Electronics* 12, no. 23 (2023): 4760.
97. A. Dosovitskiy, L. Beyer, and A. Kolesnikov, et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," arXiv preprint, arXiv:2010.11929 (2020).
98. Z. Wang, L. Shen, Z. Wang, Y. Lin, and J. Chen, "Prior-Based Underwater Enhanced Image Quality Assessment Network," *IEEE Journal of Oceanic Engineering* 49, no. 2 (2024): 592–605.
99. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference* (Springer, 2015), 234–241.
100. A. Vaswani, N. Shazeer, and N. Parmar, et al., "Attention is All You Need," in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., 2017).
101. H. R. Sheikh, M. F. Sabir, and A. Bovik, "A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms," *IEEE Transactions on Image Processing* 15, no. 11 (2006): 3440–3451.
102. J. Zhou, T. Yang, and W. Zhang, "Underwater Vision Enhancement Technologies: A Comprehensive Review, Challenges, and Recent Trends," *Applied Intelligence* 53, no. 3 (2023): 3594–3621.
103. M. Larson, "Analysis of Variance," *Circulation* 117, no. 1 (2008): 115–121.
104. R. Henson, "Analysis of Variance (ANOVA)," in *Brain Mapping: An Encyclopedic Reference* (Elsevier, 2015), 477–481.