


Article

Towards Reliable Evaluation of Underwater Image Enhancement Using Subjective and Objective Analysis

Stella Palazari ¹ and Emil Dumic ^{2,*} 

¹ Department of Multimedia, University North, 104. Brigade 3, 42000 Varaždin, Croatia; stella.palazari23@gmail.com

² Department of Electrical Engineering, University North, 104. Brigade 3, 42000 Varaždin, Croatia

* Correspondence: edumic@unin.hr

Abstract

This paper presents a systematic evaluation framework for underwater image enhancement (UIE), focusing on reliable quality assessment for vision applications in challenging underwater environments. The framework jointly analyzes subjective visual quality and objective image quality assessment measures. A controlled, laboratory-based subjective study following the ITU-R absolute category rating protocol is conducted on two datasets: UIEBD (with and without quasi-reference images) and the EUVP validation subset. A total of 132 images from UIEBD and 120 images from EUVP are evaluated, including enhanced images from four recent deep learning-based UIE models (CCL-Net, HUPE, GuidedHyb-SensUIR, and UDNNet). The subjective results reveal dataset-dependent behavior of the evaluated methods, highlighting the challenges of reliable perceptual evaluation in the presence of diverse degradations and quasi-reference data. Objective analysis shows that modern learning-based, no-reference image quality assessment (NR-IQA) models exhibit higher correlation with subjective mean opinion scores than traditional underwater-specific measures. In particular, TOPIQ_NR achieves a Spearman correlation of 0.80 on UIEBD and remains among the top-performing methods on EUVP, where LIQE reaches 0.87, while widely used measures such as UIQM and UCIQE show weaker alignment with human perception. These findings support the adoption of learning-based NR-IQA measures for robust underwater vision systems.

Keywords: underwater image enhancement (UIE); subjective quality assessment; image quality assessment (IQA); no-reference IQA; UIEBD; EUVP

1. Introduction

Underwater imaging is fundamental to a wide range of marine applications, such as oceanographic research, underwater robotics, environmental surveillance, and archaeological exploration. More broadly, it represents a specific case of imaging in challenging environments closely related to remote sensing scenarios where image quality degradation significantly affects downstream analysis and interpretation. Nevertheless, underwater images often suffer from significant quality degradation caused by wavelength-dependent light absorption and scattering, leading to color imbalance, diminished contrast, uneven illumination, and the loss of fine structural details. These degradations significantly limit the usage of raw underwater images for both human interpretation and downstream computer vision tasks.



Academic Editor: Xian Tao

Received: 15 April 2026

Revised: 24 May 2026

Accepted: 27 May 2026

Published: 2 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

To solve these problems, many underwater image enhancement (UIE) algorithms have been proposed. Early approaches rely on physical modeling, handcrafted priors, or heuristic color and contrast correction techniques. More recently, deep learning-based methods have shown substantial performance improvements by learning data-driven mappings from degraded to enhanced images, often incorporating physical cues, contrastive learning strategies, Transformer architectures, or unsupervised training procedures.

Comprehensive reviews of UIE techniques, recent trends, and open challenges can be found in [1–3]. A recent review further provides a systematic overview of underwater image restoration methods, covering cascaded framework, as well as physics-based and deep learning approaches, along with publicly available datasets and both reference and no-reference evaluation metrics [4]. In addition, polarimetric imaging techniques for imaging in scattering media, covering physical models, system designs, and emerging learning-based approaches, were reviewed in [5].

Despite significant progress in UIE algorithms, underwater image quality assessment (UIQA) remains a difficult task. True reference images are rarely accessible in real-world underwater environments, making fair evaluation of enhancement algorithms difficult. To address this issue, several benchmark datasets have been introduced, including the Underwater Image Enhancement Benchmark Dataset (UIEBD) [6] and the Enhancing Underwater Visual Perception (EUVP) dataset [7]. While such datasets have aided in the development and comparison of UIE approaches, the usefulness of quasi-reference images as ground truth remains restricted, emphasizing the importance of extensive subjective evaluation and reliable objective quality measures.

Along with the development of enhancement algorithms, a number of objective image quality assessment (IQA) approaches have been established to assess improvements. General-purpose full-reference and no-reference IQA measures are widely used, as well as underwater-specific quality measurements that detect color distortion, contrast loss, and visibility degradation. However, it is unclear how well these objective measures correlate with human perceptual judgments, especially for recent deep learning-based enhancement methods. Furthermore, the availability of a systematic cross-dataset analysis that jointly evaluates subjective quality and a broad range of modern NR-IQA measures for recent deep learning-based UIE models remains limited.

In this work, we conduct subjective and objective evaluation of four recent deep learning-based UIE methods: the cascaded contrastive learning network (CCL-Net) [8], prior-guided hybrid-sense underwater image restoration (GuidedHybSensUIR) [9], heuristic underwater perceptual enhancement (HUPE) [10], and the uncertainty distribution network (UDNet) [11]. The evaluation is conducted on images from both subsets of the UIEBD dataset, including images with quasi-references and challenging images without references, as well as on images from the EUVP validation subset. A controlled subjective experiment is performed to obtain Mean Opinion Scores (MOSs), which are then compared against a broad set of general-purpose and underwater-specific no-reference image quality assessment (NR-IQA) measures. The study further investigates whether modern objective IQA measures can reliably reflect human perceptual judgments across different underwater datasets and enhancement conditions. In contrast to existing evaluation studies primarily focused on comparative analysis of UIE methods and objective metrics, the proposed framework emphasizes the relationship between subjective human perception and objective IQA measures through controlled multi-dataset evaluation.

The main contributions of this paper are summarized as follows:

- We conduct a systematic multi-dataset subjective and objective evaluation of four state-of-the-art deep learning-based UIE models in laboratory conditions according to ITU-R BT.500-15 [12] using images from the UIEBD and EUVP datasets.

- We demonstrate that the perceived performance of UIE models is strongly dataset-dependent, with methods such as UDNNet and GuidedHybSensUIR achieving subjective quality comparable to or exceeding that of raw images in specific scenarios across the considered datasets.
- We show that recent general-purpose no-reference IQA measures, such as no-reference top-down image quality assessment (TOPIQ_NR) [13] and the learning-based image quality evaluator (LIQE) [14], achieve higher correlation with subjective MOS values than traditional underwater-specific measures such as the underwater image quality measure (UIQM) [15] and underwater color image quality evaluation (UCIQE) [16].

This paper is organized as follows: Section 2 presents related work, including both traditional and recent deep learning-based UIE models, as well as general-purpose and underwater-specific IQA measures. Section 3 describes the dataset construction process and the subjective assessment protocol. Section 4 presents the results of the objective quality evaluation. Section 5 discusses the subjective and objective assessment results of the evaluated UIE models. Finally, Section 6 concludes the paper.

2. Related Work

This section examines related work on UIE algorithms, classifying existing approaches as traditional methods or deep learning-based algorithms. Following that, existing general-purpose IQA and specific underwater IQA measures are discussed.

2.1. Underwater Image Enhancement Algorithms

UIE is affected by wavelength-dependent absorption and scattering, resulting in color casts, reduced contrast, and loss of fine details. The methods considered in this subsection can be grouped into traditional, model-based, or prior-driven approaches and deep learning-based approaches, including multi-branch and cross-view feature learning, contrastive learning, Transformer models, and recent generative models.

2.1.1. Traditional and Prior-Driven Methods

An early physically grounded restoration method is the dark-channel prior (DCP) [17], originally proposed for haze removal in atmospheric imaging and later adapted to underwater environments through the underwater dark-channel prior (UDCP) [18]. UDCP incorporates underwater light propagation effects, including scattering and wavelength-dependent absorption, in order to improve visibility and restore degraded underwater images. Retinex-based enhancement approaches [19] have also been explored to improve illumination consistency and perceptual contrast in underwater images. Several subsequent methods further extended DCP-based restoration frameworks. The generalized DCP approach [20] combines adaptive color correction with ambient light estimation and transmission modeling for restoration in turbid media. The red-channel method [21] exploits wavelength-dependent attenuation properties to improve underwater color correction and contrast restoration. In addition, a restoration framework based on image blurriness and light absorption [22] was proposed for more accurate underwater depth estimation under varying illumination conditions.

Optimization-based formulations continue to be useful in cases where explicit modeling of degradation is required. The authors of [23] proposed a model based on the underwater image formation model (UIFM) that simultaneously handles dehazing and deblurring. The proposed model uses a variational framework guided by the red-channel prior. The illumination-channel sparsity prior (ICSP) [24] incorporates a channel sparsity prior into a variational restoration framework to address non-uniform illumination.

Several methods focus on explicit color or contrast correction. The attenuated color channel correction (ACCC) method [25] compensates for differences between superior and inferior color channels, then applies multi-scale unsharp masking for better visual quality. Multi-interval sub-histogram perspective equalization (MSPE) [26] adapts a histogram-based model to spatially varying degradation levels to enhance underwater image quality.

Other traditional approaches address specific artifacts and challenging acquisition processes. The underwater vignetting image correction (UVIC) framework [27] addresses degradations caused by artificial illumination by separating vignetting and backscattering components and applying adaptive brightness and color correction. Color correction with multi-scale fusion (CCMF) [28] uses red-channel color compensation, contrast and adaptive gamma correction, and multi-scale feature fusion to improve the visual quality of underwater images.

Polarization-based approaches have also been explored to explicitly suppress scattering effects. Earlier polarization-based methods typically exploit polarization differences between object signals and backscattered light through Stokes-parameter analysis and polarized image pairs [29]. More recent polarization-guided Stokes descattering (PGSD) methods [30] combine physical polarization modeling with multi-parameter optimization and a degree of linear polarization (DoLP)-gated airlight estimation strategy to achieve robust descattering under varying haziness and illumination conditions.

Traditional and hybrid UIE methods such as UDCP [18], retinex-based enhancement [19], the red-channel method [21], and ACCC [25] established important physically motivated and color correction-based restoration strategies, which later motivated the development of recent deep learning-based enhancement approaches.

2.1.2. Deep Learning-Based Methods

Deep learning methods commonly learn data-driven mappings from degraded to enhanced images, often incorporating physical cues or engineered guidance for improved generalization. Ucolor [31] enhances underwater images by embedding multiple color spaces within an attention-guided network and using a medium transmission-guided decoder inspired by physical imaging models to mitigate color casts and low contrast. The underwater color correction network (UCCNet) and its knowledge-transfer variant (UCCNet-KT) [32] improve underwater images by exploiting cross-channel guidance, where each degraded color channel informs the correction of the others.

Multi-branch and cross-view feature learning has been explored to exploit complementary cues. The cross-view enhancement network (CVE-Net) [33] enhances underwater images by exploiting cross-view neighboring features via efficient feature alignment and dual-branch attention to suppress irrelevant content. The prior-guided hybrid-sense underwater image restoration method (GuidedHybSensUIR) [9] is a multi-scale underwater image restoration framework guided by a color-balance prior that combines detail restoration and contextual feature modeling to effectively correct color casts, recover blurry details, and provide a comprehensive benchmark for evaluating underwater image restoration methods.

Contrastive learning has recently been adopted as an effective regularizer for UIE. The hybrid contrastive learning regularization network (HCLR-Net) [34] uses non-paired data with locally perturbed negative samples, adaptive hybrid attention, and a detail repair branch to improve generalization and texture restoration in underwater image enhancement. The cascaded contrastive learning network (CCL-Net) [8] employs a two-stage framework with stage-wise contrastive objectives, performing color correction followed by haze removal to progressively improve underwater image visibility and contrast.

Transformer-based architectures have been introduced to better capture global dependencies and improve color consistency. The transmission-aware Swin Transformer (TAFormer) [35] was proposed for underwater image enhancement, combining physical imaging priors with convolutional and Transformer architectures to better model both local and long-range dependencies. Phaseformer [36] is a lightweight phase-based Transformer network for underwater image restoration using phase-based self-attention and optimized phase attention blocks to extract non-contaminated features and restore structural details while maintaining low model complexity.

Other model architectures have also been explored for UIE. The Transformer-based diffusion model proposed in [37] improves denoising efficiency and effectiveness by adopting skip sampling with non-uniform time-step strategies. The uncertainty distribution network (UDNet) [11] is an unsupervised underwater image enhancement framework that generates reference maps using statistically guided multi-color space stretching, eliminating the need for manual annotations. It uses uncertainty-aware feature learning through a conditional variational autoencoder and probabilistic adaptive instance normalization to effectively enhance contrast, saturation, and gamma correction with limited training data. Heuristic underwater perceptual enhancement (HUPE) [10] is a heuristic invertible network for underwater perception enhancement that uses an information-preserving reversible transformation with embedded Fourier features and semantic collaborative learning to jointly improve visual quality and support downstream task performance.

2.2. Objective Image Quality Assessment Measures

In this subsection, we discuss representative general-purpose and underwater-specific IQA measures that are later used in the experimental section. A concise overview of these measures is provided in Table 1. The table provides a representative overview of commonly used IQA measures relevant to this work and does not aim to cover all existing methods.

Table 1. Overview of representative IQA measures categorized by application domain, reference availability, and methodology.

Domain	Type	IQA Measures
General-purpose	FR	PSNR, SSIM [38], MS-SSIM [39], IQM2 [40], TOPIQ_FR [13], etc.
	NR, handcrafted	BRISQUE [41], NIQE [42], ILNIQE [43], NRQM [44], PIQE [45], PI [46], etc.
	NR, learning-based	ARNIQA [47], CNNIQA [48], DBCNN [49], HyperIQA [50], MANIQA [51], MUSIQ [52], PAQ-2-PIQ [53], TOPIQ_NR [13], TReS [54], WADIQAM [55], CLIP-IQA/CLIP-IQA+ [56], QualiCLIP/QualiCLIP+ [57], Q-Align [58], LIQE [14], etc.
Underwater-specific	NR, handcrafted	UIQM [15] (UICM, UISM, UIConM), UCIQE [16], CCF [59], CSN [60], UIF [61], UIQI [62], etc.
	NR, learning-based	URanker [63], ATUIQP [64], AMQI [65], etc.

2.2.1. General-Purpose Image Quality Assessment

Objective IQA plays a crucial role in the evaluation of image enhancement and restoration algorithms. Existing IQA methods can be broadly categorized according to the avail-

ability of a reference image. In this subsection, we review commonly used full-reference (FR) IQA and no-reference (NR) IQA measures for general-purpose IQA.

Full-reference image quality assessment (FR-IQA) methods evaluate image fidelity by directly comparing a distorted image with its reference. Among the most widely used FR-IQA measures are the structural similarity index (SSIM) [38] and its multi-scale extension, the multi-scale structural similarity index (MS-SSIM) [39], which assess image quality by modeling perceived structural information across spatial scales. Image Quality Measure 2 (IQM2) [40] is another FR-IQA measure that combines the structural similarity index (SSIM) with a steerable pyramid wavelet transform applied across multiple scales and orientations. These measures are commonly used when reference images are available, as they generally correlate better with human perception than pixel-wise error measures such as the peak signal-to-noise ratio (PSNR).

Reduced-reference image quality assessment (RR-IQA) measures require only partial information from the reference image by extracting relevant features. Although reduced-reference measures have been extensively studied in the broader IQA literature, they are less commonly adopted in UIE evaluation and are therefore not explicitly evaluated in this work.

NR-IQA methods estimate perceptual quality without access to reference images and are therefore particularly suitable for real-world applications. Early NR-IQA approaches use handcrafted features and natural scene statistics (NSSs), measuring statistical irregularities observed in natural images. Representative measures in this category include the Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [41], the Natural Image Quality Evaluator (NIQE) [42], the Integrated Local Natural Image Quality Evaluator (ILNIQE) [43], the No-Reference Quality Metric (NRQM) [44], the Perception-based Image Quality Evaluator (PIQE) [45], and the Perceptual Index (PI) (adopted in perceptual image quality benchmarks) [46].

More recent NR-IQA methods use deep learning to learn quality-aware representations directly from data. CNN-based and Transformer-based models, such as the adaptive representation-based no-reference image quality assessment method (ARNIQA) [47], convolutional neural network image quality assessment (CNNIQA) [48], the deep bilinear convolutional neural network (DBCNN) [49], hyper network-based image quality assessment (HyperIQA) [50], multi-dimension attention network for image quality assessment (MANIQA) [51], multi-scale image quality Transformer (MUSIQ) [52], patch-to-image quality ranking (PAQ-2-PIQ) [53], top-down image quality assessment–no-reference (TOPIQ_NR) [13], Transformer-based ranking strategy (TReS) [54], and weighted average deep image quality assessment model (WADIQAM) [55], show improved robustness across diverse distortion types by exploiting large-scale subjective datasets.

In addition, vision–language models have recently been introduced to NR-IQA, such as contrastive language–image pretraining-based image quality assessment (CLIP-IQA) and its improved variant, CLIP-IQA+ [56]; quality-aware CLIP (QualiCLIP) and QualiCLIP+ [57]; quality alignment (Q-Align) [58]; and the learning-based image quality evaluator (LIQE) [14]. These measures incorporate semantic priors and text-defined quality levels learned from large multimodal datasets to better align objective scores with human perceptual judgments. Despite their effectiveness for general-purpose IQA, most NR-IQA measures are not explicitly designed to handle underwater-specific degradations.

In the experimental section, FR-IQA measures implemented in Matlab 2020a are used—namely, PSNR, SSIM, and MS-SSIM. Regarding the IQM2 measure, its default Matlab implementation is used. All evaluations are conducted on grayscale images. NR-IQA measures are computed using the IQA-PyTorch toolbox (version 0.1.14.1) [66].

2.2.2. Underwater Image Quality Assessment

Objective UIQA is challenging due to the absence of reliable reference images and complex degradations caused by wavelength-dependent absorption and scattering. Existing approaches can be divided into handcrafted no-reference measures and learning-based measures.

Among earlier handcrafted measures, the underwater image quality measure (UIQM) [15] is an NR-IQA measure defined as a weighted combination of the underwater image colorfulness measure (UICM), underwater image sharpness measure (UISM), and underwater image contrast measure (UIConM). Explicitly modeling color distortion, sharpness, and contrast loss, UIQM has been widely adopted for evaluation of UIE methods. Underwater color image quality evaluation (UCIQE) [16] follows a similar no-reference formulation but emphasizes color-related degradations through a linear combination of chroma, saturation, and contrast features. To better reflect underwater imaging characteristics, the colorfulness, contrast, and fog density (CCF) index [59] measure was proposed, which extracts color and contrast features derived from physical image formation principles.

Further perceptually motivated methods include the contrast, sharpness, and naturalness (CSN) index [60], which models human color sensitivity to underwater distortions. Underwater image fidelity (UIF) [61] is designed to assess the fidelity of UIE results by evaluating the naturalness, sharpness, and structure of an enhanced underwater image (compared to the raw input image). Finally, the underwater image quality index (UIQI) [62] uses multiple perceptual dimensions, such as luminance, color cast, sharpness, contrast, fog density, and noise, to provide a comprehensive quality index for underwater images.

Recent research has increasingly focused on learning-based UIQA. Underwater ranker (URanker) [63], built on a conv-attentional image Transformer, formulates underwater quality assessment as a learning-to-rank problem, emphasizing relative quality comparisons instead of absolute scores. The attention and Transformer-driven underwater image quality predictor (ATUIQP) [64] uses channel and spatial attention modules and a Transformer. The reported paper was accompanied by a large-scale dataset, showing improved alignment with MOS values. The attention and mamba-driven quality index (AMQI) [65] employs attention mechanisms and state-space modeling to capture both local artifacts and long-range dependencies in underwater images.

In the experimental section, UIQM and its components (UICM, UISM, and UIConM), as well as UCIQE, are computed using the implementation provided in [67]. URanker is evaluated using the IQA-PyTorch toolbox [66], while the remaining UIQA measures are computed using their respective official repositories.

3. Dataset Construction and Subjective Evaluation

Several publicly available underwater image datasets have been proposed in recent years. However, their suitability for controlled subjective evaluation and fair comparison of pretrained UIE models varies. In this work, we selected the UIEBD dataset [6] and the EUVP validation dataset [7]. From the considered image pools, 24 images per dataset were selected in order to provide representative coverage of different underwater degradation characteristics, including variations in haze intensity, color distortion, illumination conditions, visibility degradation, and scene composition. The selection process also considered image resolution consistency, visual diversity, and practical constraints associated with controlled laboratory-based subjective evaluation. The final number of selected images was additionally constrained in order to maintain a manageable experimental duration and ensure reliable observer attention throughout the subjective evaluation procedure. Images containing visible watermarks, embedded method labels, or other potentially biasing visual annotations were excluded from the experiments. In particular, 105 images from the UIEBD

subset, 47 images from the UIEBD challenging subset, and 31 images from the EUVP validation subset were initially considered before selecting the final image subsets used in the experiments.

In the UIEBD dataset, we considered 105 images from the UIEBD subset and 47 images from the UIEBD challenging subset with a fixed resolution of 1280×720 pixels. UIEBD has been widely used for the benchmarking of underwater image enhancement algorithms and provides images suitable for subjective assessment. The final UIEBD selection consisted of 24 raw PNG images:

- A total of 12 images from the subset of 890 images with quasi-reference—specifically, images “3650”, “3728”, “3925”, “3947”, “9547”, “9554”, “9557”, “12290”, “12299”, “12324”, “12336”, and “15113”;
- A total of 12 images from the subset of 60 images without quasi-reference (UIEBD-Challenging)—namely, images “52”, “102”, “432”, “579”, “605”, “616”, “627”, “770”, “866”, “880”, “2575”, and “2856”.

In addition, the EUVP validation dataset was included to enable cross-dataset validation and assess the robustness of subjective and objective evaluation results under different image characteristics and acquisition conditions. In total, 31 images with a resolution of 1600×1200 pixels were considered from the EUVP validation subset. We selected a total of 24 raw JPEG images—specifically, images “n01496331_11850”, “n01496331_12025”, “n01496331_2556”, “n01496331_3153”, “n01914609_1712”, “n01914609_2552”, “n01914609_4209”, “n01914609_778”, “n01917289_1350”, “n01917289_183”, “n01917289_1864”, “n01917289_1907”, “n01917289_1916”, “n01917289_2052”, “n01917289_2068”, “n01917289_248”, “n01917289_290”, “n01917289_53”, “n01917289_79”, “n01917289_880”, “n01917289_908”, “n01917289_923”, “n01917289_971”, and “n01917289_973”.

Enhanced versions of these 24 raw images were generated using the four deep learning-based UIE algorithms described earlier. Importantly, all methods evaluated in this study have previously been tested on the UIEBD and EUVP datasets, enabling a consistent and fair comparison using publicly available pretrained models. The official repositories for the tested algorithms are listed below:

- CCL-Net [8]: ref. [68], using the HRNet model and pretrained weights;
- GuidedHybSensUIR [9]: ref. [69], using the default model and pretrained weights;
- HUPE [10]: ref. [70], using the default model and pretrained weights;
- UDNet [11]: ref. [71], using the default model and pretrained weights.

Table 2 summarizes the technical specifications of the equipment used in the experiments, as well as the demographic characteristics of the participant group. All images were displayed at their original resolution.

Table 2. Equipment information, observer statistics, and outliers.

	UIEBD	EUVP
Monitor	Dell 2407WFP-HC	Dell 2407WFP-HC
Screen diagonal	24"	24"
Resolution	1920×1200 pixels	1920×1200 pixels
Viewing distance	0.4 m	0.4 m
Male observers	6	16
Female observers	10	1
Overall	16	17
Age range (years)	22–40	20–41
Average age (years)	25	22
Number of outliers	0	0

3.1. Subjective Quality Assessment on the UIEBD Dataset

The experimental set included the 24 raw images and 12 corresponding quasi-reference images, resulting in a total of $5 \times 24 + 12 = 132$ images used for further evaluation. Before the experiment, participants confirmed that they had normal or corrected-to-normal visual acuity and normal color vision. The subjective experiment was developed using HTML, JavaScript, and PHP and is also accessible online at the following link: https://msl.unin.hr/experiment_uie_ueibd/ (accessed on 5 April 2026). Prior to the main test session, a training session was presented to the participants using five images that approximately spanned the same quality range as those in the test session but were either different raw images or derived from different raw images. The scores obtained during the training session were excluded from further analysis.

The subjective evaluation followed a five-level absolute category rating (ACR) scale (1—Bad; 2—Poor; 3—Fair; 4—Good; 5—Excellent). Special care was taken to ensure that images derived from the same raw image or processed by the same enhancement algorithm were not presented consecutively. Overall, 16 participants successfully completed the subjective evaluation.

The collected scores were independently screened for outliers using the procedure specified in ITU-R Recommendation BT.500-15 [12], and no outliers were identified. Subsequently, MOS values and their 95% confidence intervals (CIs) were computed for each enhancement algorithm, as well as for the raw images, assuming Student's *t*-distribution for statistical estimation. Inter-rater reliability was evaluated using the intraclass correlation coefficient (ICC) [72] computed with the MATLAB ICC function [73] using the absolute-agreement average-measures model (type 'A-k'). Inter-rater reliability, assessed using the absolute-agreement average-measures ICC model, yielded a value of 0.93 (95% CI: [0.91, 0.95]), indicating excellent agreement among observers. Figure 1 presents the average MOS values for the four tested enhancement algorithms, as well as for the raw images, resulting in a total of $5 \times 24 = 120$ evaluated images. The average MOS values for the quasi-reference images are not reported, as only 12 out of the 24 images have corresponding quasi-references. The results indicate that, for the UIEBD dataset, raw images achieve the highest average MOS, while UDNNet and GuidedHybSensUIR achieve comparable performance across the evaluated methods.

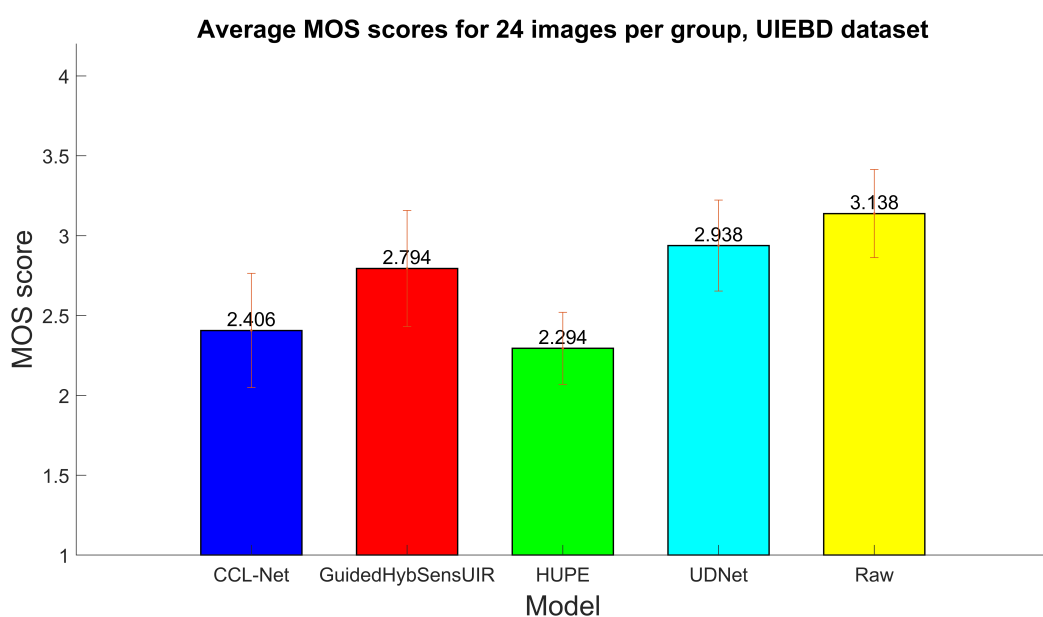


Figure 1. Average MOS values with confidence intervals for the four tested enhancement algorithms (CCL-Net, GuidedHybSensUIR, HUPE, and UDNNet) and raw images on the UIEBD dataset.

Next, we evaluated the results for the 12 images with available quasi-references, and their average MOS values are presented in Figure 2. The figure reports average MOS values for the four tested enhancement algorithms, as well as for the corresponding raw and quasi-reference images, resulting in a total of $6 \times 12 = 72$ evaluated images. Overall, the observed trends are consistent with the results obtained for the full image set; however, the quasi-reference images exhibit noticeably lower average MOS values compared to the raw images.

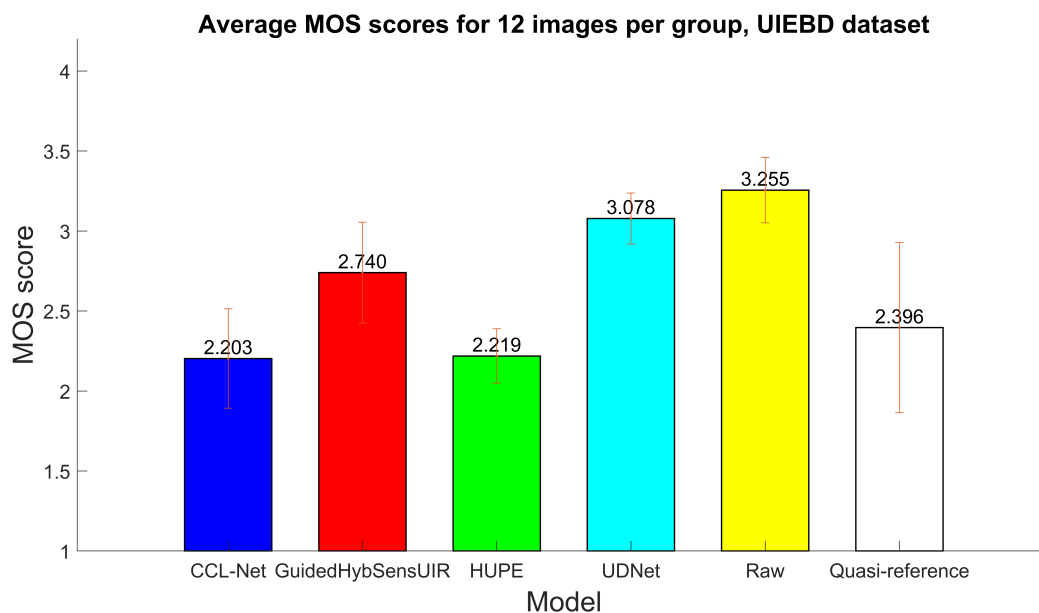


Figure 2. Average MOS values with confidence intervals for the four tested enhancement algorithms (CCL-Net, GuidedHybSensUIR, HUPE, and UDNNet) and raw and quasi-reference images on the UIEBD dataset.

The image with the highest MOS in the evaluated dataset, image “627” from the UIEBD challenging subset, is shown in Figure 3. For this sample, the GuidedHybSensUIR enhancement method achieves the best perceived quality, while the original image also attains a high MOS value. It can be observed that, for this image, the HUPE and UDNNet methods do not fully remove haze, resulting in lower perceptual quality. Although CCL-Net partially suppresses haze, it introduces an unnatural reddish color in the background, which, likewise, reduces subjective quality.

The image with the lowest MOS in the evaluated dataset, image “12290” from the UIEBD dataset with quasi-references, is shown in Figure 4. For this sample, the raw image achieves the highest MOS among the evaluated methods, while UDNNet provides comparable visual quality. Although UDNNet produces a brighter image compared to the original, it also introduces haze, which negatively affects perceptual quality. The HUPE model overexposes certain regions of the image, leading to lower MOS values, while other methods show noticeable visual distortions.

3.2. Subjective Quality Assessment on the EUVP Dataset

To further compare enhanced images, we also evaluated previously mentioned images from the EUVP validation dataset. Table 2 summarizes the technical specifications of the equipment used in the second experiment, as well as the demographic characteristics of the participant group. All images were displayed at their original resolution in full-screen mode. The experiment can be found online at the following link: https://msl.unin.hr/experiment_uie_euwp/ (accessed on 5 April 2026).



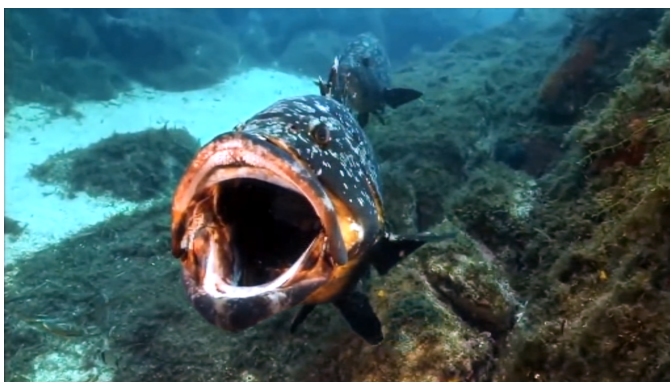
(a) Original, MOS = 4.25



(b) CCL-Net, MOS = 3.625



(c) GuidedHybSensUIR, MOS = 4.625



(d) HUPE, MOS = 3.25



(e) UDNNet, MOS = 4.0625

Figure 3. Image “627” and results: (a) original; (b) CCL-Net; (c) GuidedHybSensUIR; (d) HUPE; (e) UDNNet.

The subjective evaluation followed the same training procedure as for the UIEBD dataset evaluation, as well as the same five-level ACR scale. Overall, 17 participants successfully completed the subjective evaluation.

The previously described score screening procedure using the ITU-R Recommendation BT.500-15 [12] showed no outliers. MOS values and their 95% confidence intervals (CIs) were computed for each enhancement algorithm, as well as for the raw images, assuming Student’s *t*-distribution. Inter-rater reliability, assessed using the absolute-agreement average-measures ICC model, yielded a value of 0.94 (95% CI: [0.93, 0.96]), indicating excellent agreement among observers. Figure 5 presents the average MOS values for the four tested enhancement algorithms, as well as for the raw images, resulting in a total of $5 \times 24 = 120$ evaluated images.

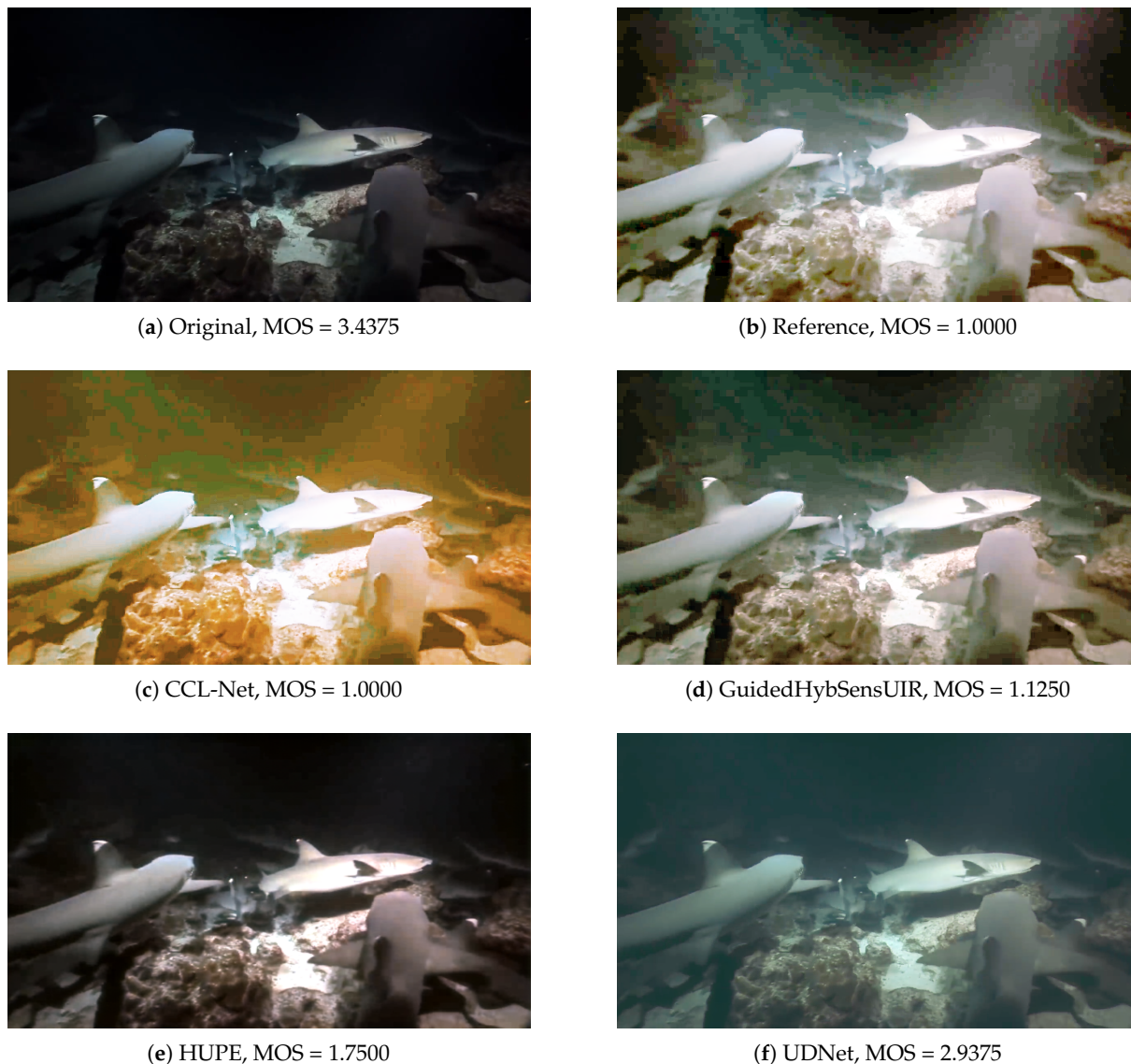


Figure 4. Image “12290” and results: (a) original; (b) reference; (c) CCL-Net; (d) GuidedHybSensUIR; (e) HUPE; (f) UDNet.

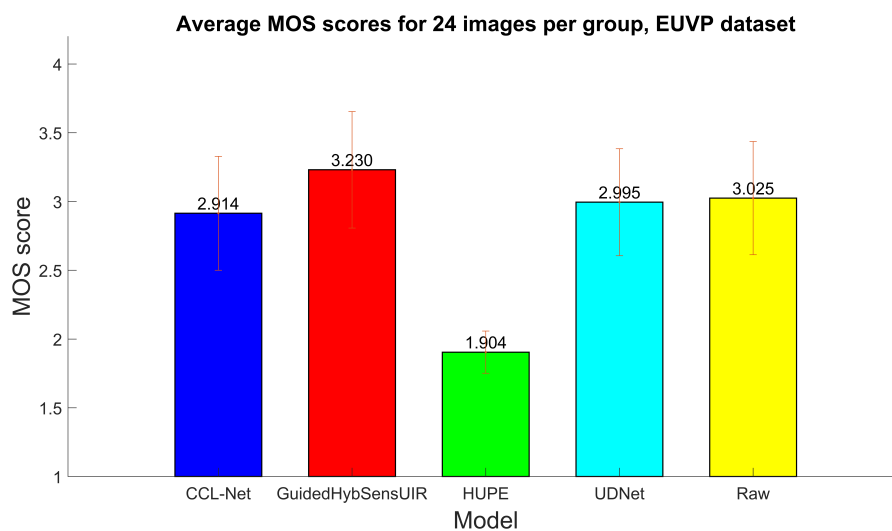


Figure 5. Average MOS values with confidence intervals for the four tested enhancement algorithms (CCL-Net, GuidedHybSensUIR, HUPE, and UDNet) and raw images on the EUVP dataset.

4. Objective Quality Assessment

In this section, we compare the obtained MOS results with the previously described NR-IQA measures, including both general-purpose and UIE-specific measures. In addition, FR-IQA results are reported for the dataset subset presented in Figure 2.

4.1. No-Reference Image Quality Assessment

In this subsection, we compare the set of general-purpose and underwater-specific no-reference image quality assessment (NR-IQA) methods introduced earlier:

- General-purpose NR-IQA methods (22): ARNIQA, BRISQUE, CLIP-IQA+, CLIP-IQA, CNNIQA, DBCNN, HyperIQA, ILNIQE, LIQE, MANIQA, MUSIQ, NIQE, NRQM, PAQ-2-PIQ, PI, PIQE, Q-Align, QualiCLIP+, QualiCLIP, TOPIQ_NR, TReS, and WADIQAM_NR.
- Underwater NR-IQA methods (11): URanker, CCF, CSN_uwiqua, CSN_uid2021, UIF, UIQI, UIQM, UICM, UISM, UIConM, and UCIQE.

The NR-IQA measures were computed for the four UIE algorithms and the corresponding raw images in both the previously described UIEBD and EUVP datasets. Quasi-reference images from the UIEBD dataset were not included, as reference information is generally unavailable and not required in NR-IQA evaluation. Overall, this procedure resulted in 120 objective scores for nearly all NR-IQA methods across both datasets. An exception is UIF, which employs raw images as implicit references; therefore, the raw images were excluded for this measure, resulting in a total of 96 scores.

The correlation between objective NR-IQA scores and subjective MOS values was then assessed using Pearson’s linear correlation coefficient (PLCC), Spearman’s rank correlation coefficient (SRCC), and Kendall’s rank correlation coefficient (KRCC). For PLCC computation, three types of nonlinear fitting were applied to the objective scores prior to correlation analysis, presented as Equations (1)–(3):

$$C_1(z) = b_1z^3 + b_2z^2 + b_3z + b_4 \tag{1}$$

$$C_2(z) = \frac{b_1 - b_2}{1 + e^{(z-b_3)/b_4}} \tag{2}$$

$$C_3(z) = b_1\left(\frac{1}{2} - \frac{1}{1 + e^{b_2(z-b_3)}}\right) + b_4z + b_5 \tag{3}$$

The correlation results for general and underwater UIE measures are reported in Tables 3 and 4 for the UIEBD and EUVP datasets, respectively.

Table 3. Correlation between objective measures and MOS values for the UIEBD dataset (the best result in each category is in bold).

	Objective Measure	PLCC_C ₁	PLCC_C ₂	PLCC_C ₃	SRCC	KRCC
General NR-IQA methods	ARNIQA	0.646	0.635	0.656	0.574	0.414
	BRISQUE	0.377	0.389	0.369	−0.378	−0.261
	CLIP-IQA+	0.686	0.683	0.685	0.676	0.488
	CLIP-IQA	0.465	0.462	0.380	0.436	0.316
	CNNIQA	0.379	0.374	0.378	0.306	0.235
	DBCNN	0.453	0.451	0.468	0.426	0.304
	HyperIQA	0.414	0.401	0.429	0.392	0.278
	ILNIQE	0.384	0.379	0.372	−0.345	−0.244
	LIQE	0.465	0.453	0.504	0.460	0.321
	MANIQA	0.405	0.407	0.409	0.373	0.271
	MUSIQ	0.720	0.715	0.714	0.682	0.506
	NIQE	0.379	0.382	0.423	−0.357	−0.255
NRQM	0.574	0.574	0.587	0.574	0.405	

Table 3. Cont.

	Objective Measure	PLCC_C ₁	PLCC_C ₂	PLCC_C ₃	SRCC	KRCC
General NR-IQA methods	PAQ-2-PIQ	0.511	0.506	0.499	0.506	0.357
	PI	0.517	0.511	0.529	−0.502	−0.358
	PIQE	0.411	0.408	0.407	−0.382	−0.265
	Q-Align	0.740	0.731	0.747	0.666	0.502
	QualiCLIP+	0.529	0.564	0.575	0.544	0.377
	QualiCLIP	0.550	0.550	0.552	0.543	0.381
	TOPIQ_NR	0.786	0.788	0.790	0.796	0.614
	TReS	0.435	0.432	0.416	0.436	0.313
	WADIQAM_NR	0.106	0.222	0.239	0.093	0.062
Underwater NR-IQA methods	Uranker	0.356	0.305	0.370	0.038	0.028
	CCF	0.227	0.321	0.333	−0.025	−0.021
	CSN_uwiqa	0.261	0.240	0.216	0.177	0.124
	CSN_uid2021	0.256	0.359	0.255	0.117	0.082
	UIF	0.564	0.564	0.577	0.567	0.400
	UIQI	0.246	0.246	0.246	0.261	0.181
	UIQM	0.263	0.140	0.262	−0.052	−0.044
	UICM	0.163	0.210	0.237	−0.131	−0.087
	UISM	0.240	0.264	0.265	0.115	0.076
	UIConM	0.300	0.359	0.359	−0.115	−0.065
	UCIQE	0.250	0.319	0.349	−0.034	−0.027

Table 4. Correlation between objective measures and MOS values for the EUVP dataset (the best result in each category is in bold).

	Objective Measure	PLCC_C ₁	PLCC_C ₂	PLCC_C ₃	SRCC	KRCC
General NR-IQA methods	ARNIQA	0.768	0.771	0.771	0.751	0.540
	BRISQUE	0.599	0.608	0.632	−0.529	−0.382
	CLIP-IQA+	0.856	0.856	0.857	0.844	0.663
	CLIP-IQA	0.806	0.806	0.806	0.787	0.603
	CNNIQA	0.775	0.774	0.775	0.758	0.553
	DBCNN	0.800	0.797	0.805	0.804	0.613
	HyperIQA	0.803	0.802	0.802	0.793	0.589
	ILNIQE	0.427	0.470	0.419	−0.375	−0.261
	LIQE	0.842	0.842	0.861	0.871	0.689
	MANIQA	0.691	0.678	0.706	0.623	0.454
	MUSIQ	0.848	0.852	0.802	0.832	0.650
	NIQE	0.743	0.744	0.750	−0.710	−0.531
	NRQM	0.607	0.683	0.685	0.597	0.418
	PAQ-2-PIQ	0.593	0.593	0.564	0.590	0.413
	PI	0.696	0.694	0.718	−0.685	−0.502
	PIQE	0.538	0.487	0.573	−0.348	−0.248
	Q-Align	0.694	0.692	0.702	0.656	0.484
	QualiCLIP+	0.777	0.779	0.792	0.744	0.557
	QualiCLIP	0.861	0.859	0.860	0.844	0.649
	TOPIQ_NR	0.819	0.819	0.819	0.804	0.608
TReS	0.759	0.757	0.751	0.743	0.554	
	WADIQAM_NR	0.599	0.598	0.598	0.579	0.412
Underwater NR-IQA methods	Uranker	0.224	0.247	0.275	0.190	0.128
	CCF	0.186	0.183	0.187	0.166	0.113
	CSN_uwiqa	0.467	0.466	0.489	0.461	0.316
	CSN_uid2021	0.437	0.437	0.443	0.364	0.254
	UIF	0.518	0.518	0.540	0.472	0.336
	UIQI	0.387	0.390	0.392	0.364	0.243
	UIQM	0.355	0.354	0.387	0.329	0.227
	UICM	0.140	0.218	0.243	0.043	0.025
	UISM	0.463	0.462	0.479	0.418	0.301
	UIConM	0.155	0.140	0.140	−0.107	−0.076
	UCIQE	0.156	0.221	0.165	0.137	0.096

4.2. Full-Reference Image Quality Assessment

In this subsection, we report the results obtained using FR-IQA measures—namely, PSNR, SSIM, MS-SSIM, and IQM2. These measures were computed on the previously described subset of the dataset containing quasi-reference images, as presented in Figure 2, resulting in a total of 60 image pairs. The image pairs were organized into five groups: four enhancement algorithms and the raw images, with 12 image pairs per group. The quantitative results are summarized in Table 5. Based on these results, the highest PSNR is achieved by the CCL-Net model, whereas the highest SSIM, MS-SSIM, and IQM2 scores are obtained by the GuidedHybSensUIR model. However, in this setting, FR-IQA measures primarily reflect similarity to the quasi-reference images, which does not necessarily correspond to a higher MOS, as presented in Figure 2.

Table 5. Average FR-IQA results computed over 12 image pairs for each UIE model and the raw images. The best values are highlighted in bold. Higher values indicate greater similarity to the quasi-reference images.

	PSNR	SSIM	MS-SSIM	IQM2
CCL-Net	23.269	0.922	0.948	0.423
GuidedHybSensUIR	22.329	0.936	0.959	0.497
HUPE	20.610	0.858	0.931	0.392
UDNet	18.816	0.866	0.873	0.196
Raw	15.758	0.724	0.815	0.134

5. Discussion

This section discusses and interprets the obtained subjective and objective evaluation results, with a particular focus on their implications for UIE assessment and evaluation protocols.

5.1. Subjective Evaluation and Statistical Analysis

In this paper, four deep learning-based UIE algorithms were compared using 24 images from the UIEBD and EUVP datasets. As shown in Figure 1, raw images achieved the highest average MOS on the UIEBD dataset in this experimental setting, while UDNet and GuidedHybSensUIR achieved comparable performance. These observations suggest that image enhancement does not necessarily lead to improved perceived visual quality under all conditions. Conversely, as shown in Figure 5, GuidedHybSensUIR achieved the highest average MOS on the EUVP dataset, while CCL-Net, UDNet, and the raw images achieved comparable performance.

The reliability of the subjective evaluation is supported by the controlled experimental conditions defined by the ITU-R BT.500-15 Recommendation, as well as the high inter-rater agreement obtained for both evaluated datasets (ICC = 0.93 and ICC = 0.94). These results confirm the consistency of observer ratings and support the validity of the reported MOS values.

To further investigate these observations, statistical analysis was performed on the overall UIEBD and EUVP datasets, comprising 120 images (Figures 1 and 5), using a one-way analysis of variance (ANOVA), followed by a Tukey–Kramer post hoc multiple-comparison test. The ANOVA revealed a statistically significant difference in average MOS values among the five evaluated groups, with $F(4, 115) = 5.796$ and $p = 276 \cdot 10^{-6}$ on the UIEBD dataset and $F(4, 115) = 8.377$ and $p = 5.75 \cdot 10^{-6}$ on the EUVP dataset. The results of the multiple-comparison test are summarized in Table 6. The analysis shows that, on the UIEBD dataset, CCL-Net yields statistically lower MOS values than the raw images, while HUPE performs significantly worse than both UDNet and the raw images. In contrast, GuidedHybSensUIR and UDNet achieve MOS values that are statistically indistinguishable

from those of the raw images. For the EUVP dataset, HUPE performs significantly worse than the other algorithms, while all others perform statistically similarly.

Table 6. Multiple-comparison test between 5 tested groups on the UIEBD and EUVP datasets with 120 images. Bold *p*-values are below the threshold of 0.05, representing statistically significant average MOS values.

Group	Control Group	UIEBD, <i>p</i> -Value	EUVP, <i>p</i> -Value
CCL-Net	GuidedHybSensUIR	0.347	0.728
CCL-Net	HUPE	0.983	0.001
CCL-Net	UDNet	0.089	0.998
CCL-Net	Raw	0.006	0.993
GuidedHybSensUIR	HUPE	0.125	0.000
GuidedHybSensUIR	UDNet	0.959	0.887
GuidedHybSensUIR	Raw	0.473	0.928
HUPE	UDNet	0.022	0.000
HUPE	Raw	0.001	0.000
UDNet	Raw	0.873	1.000

5.2. Implications for Objective Quality Assessment and Evaluation Protocols

When comparing no-reference IQA (NR-IQA) measures in Table 3, for the UIEBD dataset, the highest correlation with MOS values is achieved by the general-purpose NR metric TOPIQ_NR, with PLCC = 0.79 and SRCC = 0.80. Other recent learning-based measures, such as Q-Align and MUSIQ, also obtain PLCC values above 0.7. For the EUVP dataset (Table 4), the best results are obtained by QualiCLIP for PLCC_C₁ and PLCC_C₂, as well as by LIQE for PLCC_C₃, SRCC, and KRCC. However, several measures also obtained a PLCC correlation above 0.8—namely, CLIP-IQA+, CLIP-IQA, DBCNN, HyperIQA, MUSIQ, and TOPIQ_NR.

In contrast, underwater-specific handcrafted IQA measures, including UIQM and UCIQE, exhibit notably lower correlation with MOS. This discrepancy can be partly attributed to the evolution of underwater image enhancement algorithms, as most underwater IQA measures were originally developed for physically grounded restoration methods prior to the widespread adoption of deep learning-based enhancement approaches. The relatively higher correlation achieved by UIF (on both the UIEBD and EUVP datasets) compared to other underwater-specific measures may be explained by its training on a larger dataset that includes several deep learning-based UIE models.

A similar inconsistency is observed when comparing full-reference (FR-IQA) measures with MOS values on the UIEBD dataset. As shown in Table 5 and Figure 2, enhancement algorithms achieving higher FR-IQA scores do not necessarily achieve higher perceived quality. This behavior can be attributed to the reliance of several UIE models on quasi-reference images during training and evaluation. Since quasi-references do not represent true ground truth and may, themselves, exhibit suboptimal perceptual quality, FR metrics primarily quantify similarity to these references rather than alignment with human perception. In contrast, UDNet does not rely on quasi-reference images during training, which may reduce such bias and lead to improved subjective quality.

This work provides a systematic evaluation framework for underwater image enhancement by jointly analyzing subjective visual quality and objective IQA measures. A key strength of the proposed framework lies in the use of a controlled laboratory-based subjective experiment conducted in accordance with ITU-R recommendations, ensuring reliable MOS estimation, as confirmed by the high inter-rater agreement. In addition, this study provides a comprehensive comparison between subjective perception and a wide

range of objective IQA metrics, highlighting discrepancies that are often overlooked in UIE evaluation. Importantly, the observed variability of subjective rankings across datasets suggests that conclusions drawn from a single benchmark may not generalize reliably, emphasizing the need for multi-dataset evaluation protocols in future UIE research.

Furthermore, the effectiveness of the analyzed objective IQA measures was evaluated through correlation analysis with subjective MOS values using standard statistical measures (SRCC, PLCC, and KRCC). The obtained results show that several recent learning-based NR-IQA models achieve strong agreement with human perceptual judgments, despite not being specifically trained on underwater imagery. In particular, several modern general-purpose IQA models, including TOPIQ_NR and LIQE, demonstrated strong correlation with MOS values across the evaluated datasets, suggesting their potential suitability for perceptual quality estimation in the UIE domain. Table 7 summarizes the main strengths, limitations, and recommended application scenarios of the analyzed IQA measure categories based on the obtained results.

Table 7. Summary of the main characteristics, limitations, and recommended application scenarios of different IQA measure categories.

Metric Category	Strengths	Limitations	Recommended Usage
Underwater-specific measures	Interpretable quality indicators; designed for underwater image characteristics	Weaker correlation with subjective MOS; limited perceptual consistency	Complementary evaluation of underwater-specific degradations
Learning-based NR-IQA measures	Stronger agreement with human perceptual judgments; robust perceptual assessment	Possible domain-shift effects; reduced interpretability	Perceptual evaluation; benchmarking and optimization of UIE algorithms
FR-IQA measures	Direct comparison against reference or quasi-reference images	Dependence on reference quality; quasi-reference bias	Controlled evaluation scenarios with available reference information

Several limitations should also be acknowledged. Although the subjective experiment follows established standards, the set of evaluated images remains limited, which may affect the generalizability of the findings. This is partly due to dataset selection constraints required for controlled subjective evaluation and fair comparison of pretrained UIE models. Second, the use of quasi-reference images introduces inherent uncertainty, as such references do not represent true ground truth and may bias FR-IQA measures toward similarity rather than perceived visual quality. Furthermore, while modern NR-IQA models obtain stronger correlation with MOS, their performance may still be affected by domain shifts when applied to underwater imagery that differs from their training data, as reflected in the differences observed between the two evaluated datasets (Tables 3 and 4). This limitation is also related to the fact that most modern general-purpose NR-IQA models were not originally developed or trained specifically for underwater imaging conditions. Nevertheless, several of these models still demonstrated a strong correlation with subjective human perception across the evaluated datasets.

In addition to perceptual quality assessment, task-oriented evaluation may provide complementary insights into the practical effectiveness of UIE methods in real-world underwater applications. Although visually enhanced images may achieve higher subjective quality or improved objective IQA scores, such improvements do not necessarily translate into better performance in downstream computer vision tasks. Therefore, integrating evaluation protocols based on object detection, semantic segmentation, autonomous navigation, or feature-based matching and registration may improve the generalizability and practical

relevance of future UIE assessment frameworks. In particular, feature-preservation analysis using classical local feature extraction methods such as scale-invariant feature transform (SIFT) [74] or speeded-up robust features (SURF) [75] may provide additional insight into the structural consistency and matching robustness of enhanced underwater images. Task-oriented measures such as mean Average Precision (mAP), Intersection over Union (IoU), and Dice scores may help to further quantify the relationship between perceptual enhancement and downstream task performance.

Based on the findings of this work, we recommend that future UIE research prioritize evaluation protocols that emphasize subjective assessment and objective NR-IQA measures with better alignment with human perception, particularly learning-based approaches. Furthermore, evaluation across multiple underwater datasets with different degradation characteristics and acquisition conditions should be encouraged, since both subjective rankings and objective IQA correlations may vary considerably across datasets. Quasi-reference images should be used cautiously and primarily in controlled acquisition scenarios where corresponding in-air reference images can be reliably obtained; however, such conditions are often impractical for real underwater environments, especially for natural scenes involving vegetation or marine life. Reduced-reference IQA approaches may be particularly relevant for underwater imaging scenarios where acquiring fully reliable ground-truth reference images is impractical but partial reference information or statistical priors can still be obtained. Such approaches may represent a promising intermediate solution, since they utilize partial reference information or extracted image features while avoiding the need for complete ground-truth reference images. Downstream task-driven evaluation, such as object detection or classification performance, can provide valuable complementary insights but should be interpreted within the context of the specific task, as improvements in downstream performance do not necessarily imply enhanced perceptual quality.

6. Conclusions and Future Research

This paper presented a systematic subjective and objective evaluation of recent deep learning-based UIE algorithms. Using a controlled ITU-R subjective testing protocol, we showed that the perceived effectiveness of UIE algorithms is dataset-dependent. On the UIEBD dataset, UDNNet and GuidedHybSensUIR achieved comparable mean MOS values, suggesting that enhancement strategies not relying on quasi-reference supervision may more effectively preserve perceptual fidelity. On the EUVP dataset, GuidedHybSensUIR, CCL-Net, and UDNNet achieved comparable mean MOS values, indicating similar perceptual performance across these methods. The obtained results further suggest that reliable evaluation of UIE methods should incorporate multi-dataset analysis with diverse degradation characteristics in order to improve the robustness and generalizability of conclusions.

The objective evaluation revealed a clear mismatch between commonly used underwater-specific IQA measures and subjective human judgments, while modern general-purpose NR-IQA models showed substantially stronger alignment with MOS. These findings highlight the need to reconsider standard evaluation practices for UIE, particularly when assessing recent deep learning-based methods. The obtained results further suggest that modern learning-based NR-IQA models may provide a more reliable alternative to traditional handcrafted underwater-specific metrics for perceptual quality evaluation.

In the proposed evaluation framework, subjective assessment is considered the primary reference for perceptual quality evaluation, while objective IQA measures serve as complementary tools for scalable and repeatable analysis. Consequently, when subjective and objective evaluation results differ, subjective results should be prioritized, whereas

such discrepancies may indicate limitations of the corresponding objective measures. Nevertheless, objective IQA models that demonstrate strong correlation with human perception may still provide valuable guidance for the training, optimization, and benchmarking of future UIE algorithms.

Future research should focus on developing UIE algorithms that more directly optimize perceptual quality. At the same time, the limitations and potential biases associated with quasi-reference supervision should be carefully considered when designing and evaluating future UIE methods. In this context, reduced-reference IQA approaches may also represent a promising intermediate solution for underwater imaging scenarios where complete reference images are unavailable. In addition, future work should investigate underwater-specific IQA measures calibrated using large-scale subjective datasets. Finally, future work should investigate the relationship between perceptual image quality and downstream underwater vision tasks, such as object detection, semantic segmentation, autonomous navigation, and local feature preservation through feature-based keypoint analysis.

Author Contributions: Conceptualization, E.D. and S.P.; methodology, S.P.; writing—original draft preparation, E.D.; investigation, E.D. and S.P.; writing—review and editing, E.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was conducted within the project “Application of 3D LiDAR Point Clouds for Measuring the Efficiency of Solar Energy Systems (3D-SOLAR)” (IP-UNIN-TEH-2025-9), funded by the European Union – NextGenerationEU.

Data Availability Statement: The data presented in this study are openly available at <http://uiedataset.dynalias.com> (accessed on 5 April 2026). The subjective experiment is accessible online at the following links: https://mssl.unin.hr/experiment_uie_uiebd/ and https://mssl.unin.hr/experiment_uie_euvs/ (accessed on 5 April 2026).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

IQA	Image quality assessment
UIE	Underwater image enhancement
MOS	Mean opinion score
NR-IQA	No-reference image quality assessment
UIQA	Underwater image quality assessment
DCP	Dark-channel prior
UDCP	Underwater dark-channel prior
UIFM	Underwater image formation model
ICSP	Illumination-channel sparsity prior
ACCC	Attenuated color-channel correction
MSPE	Multi-interval sub-histogram perspective equalization
UVIC	Underwater vignetting image correction
CCMF	Color correction with multi-scale fusion
UCCNet	Underwater color correction network
UCCNet-KT	Underwater color correction network with knowledge transfer
CVE-Net	Cross-view enhancement network
GuidedHybSensUIR	Prior-guided hybrid-sense underwater image restoration
HCLR-Net	Hybrid contrastive learning regularization network
CCL-Net	Cascaded contrastive learning network
TAFFormer	Transmission-aware Swin Transformer
HUPE	Heuristic underwater perceptual enhancement

UDNet	Uncertainty distribution network
FR	Full reference
NR	No reference
FR-IQA	Full-reference image quality assessment
SSIM	Structural similarity index
MS-SSIM	Multi-scale structural similarity index
IQM2	Image quality measure 2
PSNR	Peak signal-to-noise ratio
RR-IQA	Reduced-reference image quality assessment
NSSs	Natural scene statistics
BRISQUE	Blind/Referenceless Image Spatial Quality Evaluator
NIQE	Natural Image Quality Evaluator
ILNIQE	Integrated Local Natural Image Quality Evaluator
NRQM	No-Reference Quality Metric
PIQE	Perception-based Image Quality Evaluator
PI	Perceptual Index
ARNIQA	Adaptive representation-based no-reference image quality assessment method
CNNIQA	Convolutional neural network image quality assessment
DBCNN	Deep bilinear convolutional neural network
HyperIQA	Hyper network-based image quality assessment
MANIQA	Multi-dimension attention network for image quality assessment
MUSIQ	Multi-scale image quality transformer
PAQ-2-PIQ	Patch-to-image quality ranking
TOPIQ	Top-down image quality assessment
TReS	Transformer-based ranking strategy
WADIQAM	Weighted average deep image quality assessment model
CLIP-IQA	Contrastive language-image pre-training based image quality assessment
QualiCLIP	Quality-aware contrastive language-image pre-training
Q-Align	Quality alignment
LIQE	Learning-based image quality evaluator
UIQM	Underwater image quality measure
UICM	Underwater image colorfulness measure
UISM	Underwater image sharpness measure
UIConM	Underwater image contrast measure
UCIQE	Underwater color image quality evaluation
CCF	Colorfulness index, contrast index and fog density index
AMQI	Attention and mamba-driven quality index
ATUIQP	Attention and transformer-driven underwater image quality predictor
URanker	Underwater ranker
CSN	Contrast index, sharpness index, and naturalness index
UIF	Underwater image fidelity
UIQI	Underwater image quality index
ACR	Absolute category rating
CI	Confidence interval
ICC	Intraclass correlation coefficient
SIFT	Scale-invariant feature transform
SURF	Speeded-up robust features
mAP	Mean Average Precision
IoU	Intersection over Union

References

1. Saad Saoud, L.; Elmezain, M.; Sultan, A.; Heshmat, M.; Seneviratne, L.; Hussain, I. Seeing Through the Haze: A Comprehensive Review of Underwater Image Enhancement Techniques. *IEEE Access* **2024**, *12*, 145206–145233. [[CrossRef](#)]
2. Wang, M.; Zhang, K.; Wei, H.; Chen, W.; Zhao, T. Underwater image quality optimization: Researches, challenges, and future trends. *Image Vis. Comput.* **2024**, *146*, 104995. [[CrossRef](#)]
3. Shuang, X.; Zhang, J.; Tian, Y. Algorithms for improving the quality of underwater optical images: A comprehensive review. *Signal Process.* **2024**, *219*, 109408. [[CrossRef](#)]
4. Li, B.; Chen, Z.; Lu, L.; Qi, P.; Zhang, L.; Ma, Q.; Hu, H.; Zhai, J.; Li, X. Cascaded frameworks in underwater optical image restoration *Inf. Fusion* **2025**, *117*, 102809. [[CrossRef](#)]
5. Li, X.; Han, Y.; Wang, H.; Liu, T.; Chen, S.; Hu, H. Polarimetric Imaging Through Scattering Media: A Review *Front. Phys.* **2022**, *10*, 815296. [[CrossRef](#)]
6. Li, C.; Guo, C.; Ren, W.; Cong, R.; Hou, J.; Kwong, S.; Tao, D. An Underwater Image Enhancement Benchmark Dataset and Beyond. *IEEE Trans. Image Process.* **2019**, *29*, 4376–4389. [[CrossRef](#)] [[PubMed](#)]
7. Islam, M.J.; Xia, Y.; Sattar, J. Fast Underwater Image Enhancement for Improved Visual Perception. *IEEE Robot. Autom. Lett.* **2020**, *5*, 3227–3234. [[CrossRef](#)]
8. Liu, Y.; Jiang, Q.; Wang, X.; Luo, T.; Zhou, J. Underwater Image Enhancement with Cascaded Contrastive Learning. *IEEE Trans. Multimed.* **2025**, *27*, 1512–1525. [[CrossRef](#)]
9. Guo, X.; Chen, X.; Wang, S.; Pun, C.-M. Underwater Image Restoration Through a Prior Guided Hybrid Sense Approach and Extensive Benchmark Analysis. *IEEE Trans. Circuits Syst. Video Technol.* **2025**, *35*, 4784–4800. [[CrossRef](#)]
10. Zhang, Z.; Jiang, Z.; Ma, L.; Liu, J.; Fan, X.; Liu, R. HUPe: Heuristic Underwater Perceptual Enhancement with Semantic Collaborative Learning. *Int. J. Comput. Vis.* **2025**, *133*, 3259–3277. [[CrossRef](#)]
11. Saleh, A.; Sheaves, M.; Jerry, D.; Azghadi, M.R. Adaptive deep learning framework for robust unsupervised underwater image enhancement. *Expert Syst. Appl.* **2025**, *268*, 126314. [[CrossRef](#)]
12. ITU-R BT.500-15; Methodology for the Subjective Assessment of the Quality of Television Pictures. International Telecommunications Union: Geneva, Switzerland, 2023.
13. Chen, C.; Mo, J.; Hou, J.; Wu, H.; Liao, L.; Sun, W.; Yan, Q.; Lin, W. TOPIQ: A Top-Down Approach From Semantics to Distortions for Image Quality Assessment. *IEEE Trans. Image Process.* **2024**, *33*, 2404–2418. [[CrossRef](#)]
14. Zhang, W.; Zhai, G.; Wei, Y.; Yang, X.; Ma, K. Blind Image Quality Assessment via Vision-Language Correspondence: A Multitask Learning Perspective. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 18–22 June 2023; pp. 14071–14081. [[CrossRef](#)]
15. Panetta, K.; Gao, C.; Agaian, S. Human-Visual-System-Inspired Underwater Image Quality Measures. *IEEE J. Ocean. Eng.* **2016**, *41*, 541–551. [[CrossRef](#)]
16. Yang, M.; Sowmya, A. An Underwater Color Image Quality Evaluation Metric. *IEEE Trans. Image Process.* **2015**, *24*, 6062–6071. [[CrossRef](#)]
17. He, K.; Sun, J.; Tang, X. Single Image Haze Removal Using Dark Channel Prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 2341–2353. [[CrossRef](#)] [[PubMed](#)]
18. Drews, P.L.J.; Nascimento, E.R.; Botelho, S.S.C.; Campos, M.F.M. Underwater Depth Estimation and Image Restoration Based on Single Images. *IEEE Comput. Graph. Appl.* **2016**, *36*, 24–35. [[CrossRef](#)]
19. Fu, X.; Zhuang, P.; Huang, Y.; Liao, Y.; Zhang, X.-P.; Ding, X. A Retinex-Based Enhancing Approach for Single Underwater Image. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Paris, France, 27–30 October 2014; pp. 4572–4576. [[CrossRef](#)]
20. Peng, Y.-T.; Cao, K.; Cosman, P.C. Generalization of the Dark Channel Prior for Single Image Restoration. *IEEE Trans. Image Process.* **2018**, *27*, 2856–2868. [[CrossRef](#)] [[PubMed](#)]
21. Galdran, A.; Pardo, D.; Picón, A.; Alvarez-Gila, A. Automatic Red-Channel Underwater Image Restoration. *J. Vis. Commun. Image Represent.* **2015**, *26*, 132–145. [[CrossRef](#)]
22. Peng, Y.-T.; Cosman, P.C. Underwater Image Restoration Based on Image Blurriness and Light Absorption. *IEEE Trans. Image Process.* **2017**, *26*, 1579–1594. [[CrossRef](#)]
23. Xie, J.; Hou, G.; Wang, G.; Pan, Z. A Variational Framework for Underwater Image Dehazing and Deblurring. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 3514–3526. [[CrossRef](#)]
24. Hou, G.; Li, N.; Zhuang, P.; Li, K.; Sun, H.; Li, C. Non-Uniform Illumination Underwater Image Restoration via Illumination Channel Sparsity Prior. *IEEE Trans. Circuits Syst. Video Technol.* **2024**, *34*, 799–814. [[CrossRef](#)]
25. Zhang, W.; Wang, Y.; Li, C. Underwater Image Enhancement by Attenuated Color Channel Correction and Detail Preserved Contrast Enhancement. *IEEE J. Ocean. Eng.* **2022**, *47*, 718–735. [[CrossRef](#)]
26. Zhou, J.; Pang, L.; Zhang, D.; Zhang, W. Underwater Image Enhancement Method via Multi-Interval Subhistogram Perspective Equalization. *IEEE J. Ocean. Eng.* **2023**, *48*, 474–488. [[CrossRef](#)]

27. Wang, Y.; Ma, Y.; Li, Y.; Zhang, J.; Mi, Z.; Fu, X. Underwater Vignetting Image Correction Based on Binary Polynomial Regularization and Latent Low-Rank Representation. *IEEE Trans. Circuits Syst. Video Technol.* **2025**, *35*, 3410–3425. [[CrossRef](#)]
28. Zhang, D.; He, Z.; Zhang, X.; Wang, Z.; Ge, W.; Shi, T.; Lin, Y. Underwater image enhancement via multi-scale fusion and adaptive color-gamma correction in low-light conditions. *Eng. Appl. Artif. Intell.* **2023**, *126*, 106972. [[CrossRef](#)]
29. Schechner, Y.Y.; Karpel, N. Clear underwater vision. In Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Washington, DC, USA, 27 June–2 July 2004; Volume 1, pp. 536–543. [[CrossRef](#)]
30. Chen, Z.; Wu, J.; Hu, H.; Li, X. Underwater polarimetric descattering via scene adaptation and multi-parameter optimization. *Opt. Lasers Eng.* **2026**, *196*, 109410. [[CrossRef](#)]
31. Li, C.; Anwar, S.; Hou, J.; Cong, R.; Guo, C.; Ren, W. Underwater Image Enhancement via Medium Transmission-Guided Multi-Color Space Embedding. *IEEE Trans. Image Process.* **2021**, *30*, 4985–5000. [[CrossRef](#)]
32. Lin, P.; Wang, Y.; Li, Y.; Fan, Z.; Fu, X. Underwater Color Correction Network With Knowledge Transfer. *IEEE Trans. Multimed.* **2024**, *26*, 8088–8103. [[CrossRef](#)]
33. Zhou, J.; Zhang, D.; Zhang, W. Cross-view enhancement network for underwater images. *Eng. Appl. Artif. Intell.* **2023**, *121*, 105952. [[CrossRef](#)]
34. Zhou, J.; Sun, J.; Li, C.; Jiang, Q.; Zhou, M.; Lam, K.-M.; Zhang, W.; Fu, X. HCLR-Net: Hybrid Contrastive Learning Regularization with Locally Randomized Perturbation for Underwater Image Enhancement. *Int. J. Comput. Vis.* **2024**, *132*, 4132–4156. [[CrossRef](#)]
35. Li, Y.; Mi, Z.; Wang, Y.; Jiang, S.; Fu, X. TAFormer: A Transmission-Aware Transformer for Underwater Image Enhancement. *IEEE Trans. Circuits Syst. Video Technol.* **2025**, *35*, 601–616. [[CrossRef](#)]
36. Khan, M.R.; Negi, A.; Kulkarni, A.; Phutke, S.S.; Vipparthi, S.K.; Murala, S. Phaseformer: Phase-Based Attention Mechanism for Underwater Image Restoration and Beyond. In Proceedings of the 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Tucson, AZ, USA, 28 February–4 March 2025; pp. 9618–9629. [[CrossRef](#)]
37. Tang, Y.; Kawasaki, H.; Iwaguchi, T. Underwater Image Enhancement by Transformer-based Diffusion Model with Non-uniform Sampling for Skip Strategy. In Proceedings of the 31st ACM International Conference on Multimedia, Ottawa, ON, Canada, 29 October–3 November 2023; pp. 5419–5427. [[CrossRef](#)]
38. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)]
39. Wang, Z.; Simoncelli, E.P.; Bovik, A.C. Multiscale structural similarity for image quality assessment. In Proceedings of the Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, Pacific Grove, CA, USA, 9–12 November 2003; Volume 2, pp. 1398–1402. [[CrossRef](#)]
40. Dunic, E.; Grgic, S.; Grgic, M. IQM2: New image quality measure based on steerable pyramid wavelet transform and structural similarity index. *Signal Image Video Process.* **2014**, *8*, 1159–1168. [[CrossRef](#)]
41. Mittal, A.; Moorthy, A.K.; Bovik, A.C. No-Reference Image Quality Assessment in the Spatial Domain. *IEEE Trans. Image Process.* **2012**, *21*, 4695–4708. [[CrossRef](#)] [[PubMed](#)]
42. Mittal, A.; Soundararajan, R.; Bovik, A.C. Making a “Completely Blind” Image Quality Analyzer. *IEEE Signal Process. Lett.* **2013**, *20*, 209–212. [[CrossRef](#)]
43. Zhang, L.; Zhang, L.; Bovik, A.C. A Feature-Enriched Completely Blind Image Quality Evaluator. *IEEE Trans. Image Process.* **2015**, *24*, 2579–2591. [[CrossRef](#)]
44. Ma, C.; Yang, C.-Y.; Yang, X.; Yang, M.-H. Learning a no-reference quality metric for single-image super-resolution. *Comput. Vis. Image Underst.* **2017**, *158*, 1–16. [[CrossRef](#)]
45. Venkatanath, N.; Praneeth, D.; Maruthi Chandrasekhar, B.; Channappayya, S.S.; Medasani, S.S. Blind image quality evaluation using perception based features. In Proceedings of the 2015 Twenty First National Conference on Communications (NCC), Mumbai, India, 27 February–1 March 2015; pp. 1–6. [[CrossRef](#)]
46. Blau, Y.; Michaeli, T. The Perception-Distortion Tradeoff. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6228–6237. [[CrossRef](#)]
47. Agnolucci, L.; Galteri, L.; Bertini, M.; Del Bimbo, A. ARNIQA: Learning Distortion Manifold for Image Quality Assessment. In Proceedings of the 2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 1–6 January 2024; pp. 188–197.
48. Kang, L.; Ye, P.; Li, Y.; Doermann, D. Convolutional Neural Networks for No-Reference Image Quality Assessment. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1733–1740. [[CrossRef](#)]
49. Zhang, W.; Ma, K.; Yan, J.; Deng, D.; Wang, Z. Blind Image Quality Assessment Using a Deep Bilinear Convolutional Neural Network. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *30*, 36–47. [[CrossRef](#)]
50. Su, S.; Yan, Q.; Zhu, Y.; Zhang, C.; Ge, X.; Sun, J.; Zhang, Y. Blindly Assess Image Quality in the Wild Guided by a Self-Adaptive Hyper Network. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 3664–3673. [[CrossRef](#)]

51. Yang, S.; Wu, T.; Shi, S.; Lao, S.; Gong, Y.; Cao, M.; Wang, J.; Yang, Y. MANIQA: Multi-dimension Attention Network for No-Reference Image Quality Assessment. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), New Orleans, LA, USA, 19–20 June 2022; pp. 1190–1199. [[CrossRef](#)]
52. Ke, J.; Wang, Q.; Wang, Y.; Milanfar, P.; Yang, F. MUSIQ: Multi-scale Image Quality Transformer. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 11–17 October 2021; pp. 5128–5137. [[CrossRef](#)]
53. Ying, Z.; Niu, H.; Gupta, P.; Mahajan, D.; Ghadiyaram, D.; Bovik, A.C. From Patches to Pictures (PaQ-2-PiQ): Mapping the Perceptual Space of Picture Quality. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 3572–3582. [[CrossRef](#)]
54. Golestaneh, S.A.; Dadsetan, S.; Kitani, K.M. No-Reference Image Quality Assessment via Transformers, Relative Ranking, and Self-Consistency. In Proceedings of the 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 3–8 January 2022; pp. 3989–3999. [[CrossRef](#)]
55. Bosse, S.; Maniry, D.; Müller, K.-R.; Wiegand, T.; Samek, W. Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment. *IEEE Trans. Image Process.* **2018**, *27*, 206–219. [[CrossRef](#)]
56. Wang, J.; Chan, K.C.K.; Loy, C.C. Exploring CLIP for assessing the look and feel of images. In Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence (AAAI'23/IAAI'23/EAAI'23), Washington, DC, USA, 7–14 February 2023; pp. 2555–2563. [[CrossRef](#)]
57. Agnolucci, L.; Galteri, L.; Bertini, M. Quality-Aware Image-Text Alignment for Opinion-Unaware Image Quality Assessment. *arXiv* **2024**, arXiv:2403.11176. [[CrossRef](#)]
58. Wu, H.; Zhang, Z.; Zhang, W.; Chen, C.; Liao, L.; Li, C.; Gao, Y.; Wang, A.; Zhang, E.; Sun, W.; et al. Q-ALIGN: Teaching LMMs for visual scoring via discrete text-defined levels. In Proceedings of the 41st International Conference on Machine Learning (ICML'24), Vienna, Austria, 21–27 July 2024.
59. Wang, Y.; Li, N.; Li, Z.; Gu, Z.; Zheng, H.; Zheng, B.; Sun, M. An imaging-inspired no-reference underwater color image quality assessment metric. *Comput. Electr. Eng.* **2018**, *70*, 904–913. [[CrossRef](#)]
60. Hou, G.; Zhang, S.; Lu, T.; Li, Y.; Pan, Z.; Huang, B. Underwater Image Quality Assessment via Color Sensitivity Network. *Comput. Electr. Eng.* **2024**, *118*, 109293. [[CrossRef](#)]
61. Zheng, Y.; Chen, W.; Lin, R.; Zhao, T.; Le Callet, P. UIF: An Objective Quality Assessment for Underwater Image Enhancement. *IEEE Trans. Image Process.* **2022**, *31*, 5456–5468. [[CrossRef](#)]
62. Liu, Y.; Gu, K.; Cao, J.; Wang, S.; Zhai, G.; Dong, J.; Kwong, S. UIQI: A Comprehensive Quality Evaluation Index for Underwater Images. *IEEE Trans. Multimed.* **2024**, *26*, 2560–2573. [[CrossRef](#)]
63. Guo, C.; Wu, R.; Jin, X.; Han, L.; Zhang, W.; Chai, Z.; Li, C. Underwater ranker: Learn which is better and how to be better. In Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence (AAAI'23/IAAI'23/EAAI'23), Washington, DC, USA, 7–14 February 2023; pp. 702–709. [[CrossRef](#)]
64. Liu, Y.; Zhang, B.; Hu, R.; Gu, K.; Zhai, G.; Dong, J. Underwater Image Quality Assessment: Benchmark Database and Objective Method. *IEEE Trans. Multimed.* **2024**, *26*, 7734–7747. [[CrossRef](#)]
65. Cao, J.; Zhang, B.; Liu, Y.; Hu, R.; Gu, K.; Zhai, G.; Dong, J. Attention and Mamba-Driven Quality Assessment for Underwater Images. *IEEE Trans. Multimed.* **2025**, *27*, 9761–9775. [[CrossRef](#)]
66. Chen, C.; Mo, J. IQA-PyTorch: PyTorch Toolbox for Image Quality Assessment. Available online: <https://github.com/chaofengc/IQA-PyTorch> (accessed on 1 October 2025).
67. Chen, X. Python Code for Several Metrics: PSNR, SSIM, UCIQE, and UIQM. Available online: <https://github.com/xueleichen/PSNR-SSIM-UCIQE-UIQM-Python> (accessed on 1 October 2025).
68. CCL-Net: Official implementation of the underwater image enhancement with cascaded contrastive learning. Available online: <https://github.com/lewis081/CCL-Net> (accessed on 1 October 2025).
69. GuidedHybSensUIR: Official implementation of the prior-guided hybrid-sense underwater image restoration. Available online: <https://github.com/CXH-Research/GuidedHybSensUIR> (accessed on 1 October 2025).
70. HUPE: Official implementation of the heuristic underwater perceptual enhancement. Available online: <https://github.com/ZengxiZhang/HUPE> (accessed on 1 October 2025).
71. UDNet: Official implementation of the uncertainty distribution network. Available online: <https://github.com/alzayats/UDnet> (accessed on 1 October 2025).
72. Shrout, P.E.; Fleiss, J.L. Intraclass Correlations: Uses in Assessing Rater Reliability. *Psychol. Bull.* **1979**, *86*, 420–428. [[CrossRef](#)]
73. Salarian, A. Intraclass Correlation Coefficient (ICC). Available online: <https://www.mathworks.com/matlabcentral/fileexchange/22099-intraclass-correlation-coefficient-icc> (accessed on 31 January 2026).

74. Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [[CrossRef](#)]
75. Bay, H.; Ess, A.; Tuytelaars, T.; Van Gool, L. Speeded-Up Robust Features (SURF). *Comput. Vis. Image Underst.* **2008**, *110*, 346–359. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.