Perception-Aware Underwater Image Quality Assessment: Dataset, Perceptual Quality Scores and Assessment Network

Bosen Lin, Junyu Dong, *Member, IEEE*, and Xinghui Dong, *Member, IEEE*

***Abstract*—Underwater Image Quality Assessment (UIQA) plays an important role in assess the effectiveness of Underwater Image Enhancement (UIE) algorithms or to evaluate the quality of underwater images. However, accurate UIQA that are consistent with human perception remains challenging. This dilemma on one hand is attributed to the lack of real human visual perception UIQA data, and on the other hand that the quality feature representation used by existing UIQA algorithms are inconsistent with human perceptions. To address these issues, we introduce a Large scale Underwater Image Quality Dataset (LUIQD), and propose an UIQA network named as Perception-Aware Underwater image Quality Assessment Network (PAUQA-Net). Specifically, the LUIQD includes 64,180 real and enhance underwater images covering a wide range of scenes, target and imaging conditions, with their perceptual quality scores. Based on the analysis of the mechanisms of human perception, we further design the data-driven PAUQA-Net that integrates an efficient convolutional attention vision Transformer to extract multi-scale features by a multi-path structure. Considering the specificity of human perception of underwater images, color and sharpness features from the chrominance and luminance domains are extracted and fused with local and global images features for joint feature interaction. Extensive experiments conduted on LUIQD and other datasets demonstrate that the proposed PAUQA-Net achieves superior assessment performance compared with the most popular UIQA and IQA methods. The code and dataset can be found in https://github.com/CatchACat083/PAUQA.**

***Index Terms*—Image Quality Assessment, Underwater Images, Human Visual Perception, Perceptual Quality Dataset.**

I. Introduction

The development and exploration of marine resources play a crucial role in the advancement of human society. Various tasks, such as underwater resource exploration, underwater archaeology, underwater structure inspection, and marine biology research [1] [2] [3], rely on the acquisition of high-quality underwater images. However, underwater imaging is often challenged by the scattering and absorption of light in water, leading to significant degradation in image quality. This degradation manifests as issues such as blurriness, low contrast, color cast, uneven lighting, and noise, which hinder the ability of human observers to accurately perceive and evaluate underwater objects. To improve image quality, underwater image enhancement (UIE) algorithms have been developed [4] [5]. However, to evaluate the effectiveness of UIE techniques, it is essential to assess the quality of the enhanced images in a manner that aligns with human visual perception. This process, known as Underwater Image Quality Assessment (UIQA), provides a framework for objectively quantifying and evaluating the quality of underwater images, including those enhanced by UIE algorithms, ensuring that the images meet perceptual requirements.

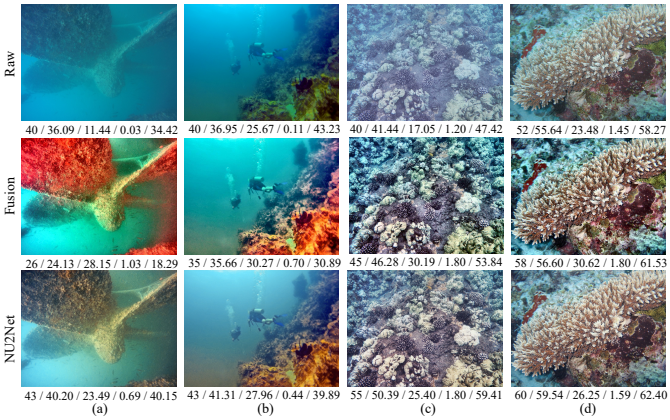

Fig. 1. Examples of underwater raw images and enhancement results using Fusion [6] and NU2Net [7] methods. The raw images exhibit typical underwater degradations, including (a) noise, (b) color distortion, and (c) blurring. Column (d) shows high-quality images. Each image is annotated with its perceptual quality score, PAUQA-Net prediction, UCIQE [8], UIQM [9], and URanker [7] values (formatted as: Perceptual Score / PAUQA-Net / UCIQE / UIQM / URanker).

Generally, UIQA is performed through no-reference or blind Image Quality Assessment (IQA) methods because original ideal images for comparison are unavailable [10] [11]. Traditional feature-based UIQA metrics focus on a set of objective indicators such as color, contrast, and sharpness [12]. However, they often fail to assess images enhanced by UIE algorithms, as these algorithms can alter colors and structures in ways that traditional metrics cannot capture, leading to results that do not match human perception. Nowadays, researchers are focusing more on user-centric UIQA methods, which aim to improve accuracy and practicality by using neural networks to extract features that describe image quality [13] [14]. With advancements in deep learning, these methods train networks using underwater images paired with perceptual quality scores,

This study was in part supported by the National Natural Science Foundation of China (NSFC) (No. 42176196). (Corresponding author: Xinghui Dong).

B. Lin is with School of Computer Science and Technology, Ocean University of China, No.238 Songling Road, Qingdao, Shangdong, China. J.Dong and X. Dong are with the State Key Laboratory of Physical Oceanography and the Faculty of Information Science and Engineering, Ocean University of China, Qingdao, 266100. (e-mail: linbosen@stu.ouc.edu.cn, dongjunyu@ouc.edu.cn, xinghui.dong@ouc.edu.cn).

DOI: 10.1109/TCSVT.2025.3574860

allowing the networks to learn the key quality features of the images [7]. This helps create consistent quality assessments that align with human perception.

Figure 1 shows four underwater images along with their enhanced results using Fusion [6] and NU2Net [7]. Some of these images are positively enhanced, while others are degraded. It is evident that traditional UIQA metrics, such as Underwater Color Image Quality Evaluation (UCIQE) [8] and Underwater Image Quality Measure (UIQM) [9], struggle to assess the perceptual quality of these images. Additionally, deep learning-based methods like URanker [7] are unable to accurately evaluate the quality of these diverse enhanced images. Therefore, given the complex and varied conditions of underwater imaging, more research is needed to design better network architectures and expand datasets that can accurately capture the unique issues in underwater images and improve the precision of UIQA methods, using human visual perception model as inspiration.

Building a large-scale, high-quality, user-centric dataset is key to training a UIQA network that can accurately evaluate the effectiveness of UIE algorithms [15]. Currently available UIQA datasets are limited by issues such as a small number of images, non-global or relative quality scores [16], and the restricted enhancement capabilities of UIE algorithms. To address these issues, we have constructed a more comprehensive and accurate UIQA dataset, which includes three key components: raw underwater images of various scenes and types of degradation, images enhanced by different UIE algorithms, and the global and absolute perceptual quality scores assigned to these images. The dataset contains 64,180 images, including raw images of varying quality and both successful and unsuccessful enhancement results. To obtain quality scores, we conducted a subjective user study to annotate the quality of each image. To balance the dataset's size with the cost of annotation, the subjective user study combines single- and multi-person annotation. The quality scores are validated through outlier elimination and statistical analysis, culminating in the formation of a Large-scale Underwater Image Quality Dataset (LUIQD).

Leveraging the LUIQD, we developed a data-driven method to effectively capture image features and facilitate accurate objective assessments of underwater image quality. This method utilizes a hybrid network that integrates multi-scale convolutional layers with a multi-path Transformer architecture, inspired by human visual perception models. Given the unique challenges presented by underwater imagery, we introduced two networks: the Color Feature Extraction Network (CFEN) and the Sharpness Feature Extraction Network (SFEN). These networks are designed to detect the inherent color and sharpness distortions prevalent in underwater images that have the greatest impact on human perception, focusing on chrominance and luminance domains, respectively.

Considering that human assessment of image quality incorporates both global composition and local details [17], we introduced an advanced convolutional attention vision transformer network. Given the interdependence of quality across different image regions [18], we designed a multi-scale Dilated Convolutional Patch Embedding (DCPE) module. This module uses dilated convolutions at varying sampling rates to extract multi-scale image tokens, enabling the extraction of features ranging from fine to coarse detail. Subsequently, a joint feature interaction layer combines these local/global features along with color/sharpness features to assess image quality comprehensively and derive the quality score. This network is termed the Perception-Aware Underwater Image Quality Assessment Network (PAUQA-Net).

The contributions of this paper are summarized as follows:

1) The LUIQD is developed to assess the quality of both real underwater images and those enhanced by various UIE methods. A total of 6,418 raw underwater images, showcasing a diverse range of scenes and degradation conditions, were collected, and a total of 64,180 images were acquired using nine different traditional and advanced deep learning-based UIE methods.

2) Based on the LUIQD, a subjective user study is designed to acquire human perception-based quality scores. We conducted a multi-person annotation experiment to validate the reliability and feasibility of single-person annotation through statistical analysis. Each image is annotated with a global and absolute quality score. To the best of our knowledge, the LUIQD dataset is the most extensive dataset capturing real human visual perception of underwater image quality available to date.

3) The PAUQA-Net is designed, employing a convolutional attention vision transformer architecture. This network optimizes the extraction of quality-related features from local to global using a multi-scale and multi-path structure, while focusing on the chrominance and luminance domains to extract critical color and sharpness features. Experimental results demonstrate that the PAUQA-Net outperforms other models in assessing the quality of underwater images.

The remainder of this paper is organized as follows. Section II provides a review of the related work. Section III details the development of the LUIQD dataset and the design of the user study for obtaining perceptual quality scores. In Section IV, we introduce the proposed PAUQA-Net network and describe its detailed structure. Experimental setup and results are discussed in Sections V and VI, respectively. Finally, Section VII presents the conclusions drawn from our work.

II. Related Works

A. Underwater Image Quality Assessment

Accurate evaluation of image quality is crucial for underwater vision perception tasks. The UIQA techniques are divided into two categories: traditional feature-based and deep learning-based approaches. Traditional feature-based UIQA methods analyze one or more image features directly related to the quality. For example, Yang et al. [8] introduced the UCIQE metric, which quantifies color deviation, blurriness, and contrast through a linear combination. Panetta et al. [9] developed UIQM, which integrates measurements of sharpness, colorfulness, and contrast. Wang et al. [12] proposed the Colorfulness, Contrast, and Fog Density (CCF) metric to assess color degradation due to water absorption, detail blurring from forward scattering, and contrast loss from backscattering.

TABLE I
COMPARISON OF DIFFERENT UIQA BENCHMARK DATASETS. THE ENHANCED METHODS (TRAD) REFER TO TRADITIONAL UNDERWATER IMAGE ENHANCEMENT METHODS, AND (DL) REFERS TO DEEP-LEARNING-BASED UNDERWATER IMAGE ENHANCEMENT METHODS. THE SUBJECTIVE METHOD (SS-ACR) REFERS TO SINGLE-STIMULUS ABSOLUTE CATEGORY RATING, WHILE (DS-PC) REFERS TO DOUBLE-STIMULUS PAIRWISE COMPARISON.

Dataset	Year	Number of Images	Enhanced Methods	Subjective Method	Score Type	Application
UWIQA [19]	2021	890 (raw)	/	SS-ACR	Global	Raw image assessment
UIED [20]	2022	1,000 (enhanced)	8 Trad + 2 DL	SS-ACR	Global	UIE result image assessment
SAUD [16]	2022	1,000 (enhanced)	7 Trad + 3 DL	DS-PC	Non-Global	UIE result image assessment
URankerSet [7]	2022	8,900 (890 raw+8,010 enhanced)	9 Trad	DS-PC	Non-Global	UIE result image assessment
UID2021 [21]	2022	960 (60 raw+900 enhanced)	/	DS-PC	Non-Global	UIE result image assessment
SOTA [13]	2023	8,000 (800 raw + 7,200 enhanced)	6 Trad + 3 DL	SS-ACR	Global	UIE result image assessment
UIQD [18]	2024	5,369 (raw)	/	SS-ACR	Global	Raw image assessment
SAUD2.0 [14]	2025	2,400 (enhanced)	7 Trad + 3 DL	SS-ACR	Global	UIE result image assessment
LUIQD (Ours)	2025	64,180 (6,418 raw + 57,762 enhanced)	3 Trad + 6 DL	SS-ACR	Global	UIE & raw image assessment

Yang et al. [19] introduced the Frequency Domain Underwater Image Quality Assessment Metric (FDUM) based on chroma, contrast, and sharpness, accounting for human sensitivity to image frequency.

To align quality scores with human perception, some studies use traditional machine learning techniques, fitting robust linear combinations based on the subjective quality score. Liu et al. [10] developed the Underwater Image Quality Index (UIQI), which integrates features like brightness, sharpness, uniformity, color bias, fog density, contrast, and noise, using Support Vector Regression (SVR) to train model parameters. Jiang et al. [16] introduced the No-reference Underwater Image Quality metric (NUIQ), which extracts perceptual features related to brightness and chromaticity, and uses support vector machines to predict the quality ranking of multiple enhanced images. Hou et al. [22] extracted features to measure contrast, sharpness, and naturalness, using SVR for quality prediction.

While multiple factors affect underwater image quality, relying solely on traditional image features limits the accuracy and comprehensiveness of UIQA. As a result, recent approaches have utilized deep learning techniques to extract image features for more effective quality evaluation. Guo et al. [7] developed "URanker," a rank-based UIQA network using a convolutional attention Transformer network, trained on human rankings of enhanced images from various algorithms. Yang et al. [23] introduced a Transformer-based network for pairwise comparison of underwater image quality. Liu et al. [18] proposed the Attention and Transformer-driven Underwater Image Quality Predictor (ATUIQP), employing a converter module to capture global features and attention modules for detecting channel-specific and local degradations. Wang et al. [24] proposed a sub-network incorporating depth, absorption coefficient, scattering coefficient, and ambient light as priors to improve UIQA accuracy. Chu et al. [25] used feature refinement and enhancement modules to extract UIQA-related features from local to global. Yang et al. [26] employed the residual difference map between enhanced images and pseudo-references to guide the UIQA network in evaluating enhanced underwater images.

Traditional UIQA methods usually focus on one or more specific quality attributes of images, which often cannot adapt to diverse underwater environments and accurately evaluate the subtle quality differences brought by UIE networks. Although deep learning-based methods have improved the ability to capture quality-related image features, existing methods cannot fully describe the global quality of both raw and enhanced underwater images, making the quality assessment results inconsistent with human visual perception. Therefore, there is a need for a UIQA network that comprehensively considers all quality attributes related to human visual assessment, ensuring a more holistic evaluation.

B. Underwater Image Quality Datasets

The UIQA datasets utilize human visual perception quality scores to assess the quality of both raw underwater images and enhanced underwater images. Depending on how the quality scores are acquired, they can be categorized into Single-Stimulus Absolute Category Rating (SS-ACR) and Double-Stimuli Pairwise Comparison (DS-PC). In SS-ACR, human observers directly assign quality scores to individual underwater images. Conversely, DS-PC involves the pairwise comparison of two images, where statistical methods are used to derive a global ranking of image quality. The comparison of different UIQA datasets is shown in Table I.

Using the SS-ACR method, Yang et al. [19] created the Underwater Image Quality Evaluation benchmark dataset (UWIQA), which consists of 890 real underwater images rated by users. Zhang et al. [20] evaluated the quality of 100 underwater images and their corresponding 1,000 enhanced images, forming the UIED dataset. Wang et al. [13] provided an advanced UIQA dataset (SOTA dataset) featuring 800 original images, each accompanied by 9 enhanced versions, together with quality scores. In [18], a UIQA database (UIQD) is constructed, comprising 5,369 real underwater images, covering a wide range of scenes and typical quality degradation conditions. Jiang et al. [14] evaluated the quality of 2,600 underwater images and their enhanced versions. This dataset also provides quality scores based on color and visibility.

Utilizing the DS-PC method, Jiang et al. [16] developed the Subjectively-Annotated UIE Benchmark Dataset (SAUD), which includes 100 real images along with 1,000 enhanced counterparts. This dataset employs the Bradley-Terry (B-T) model to generate a global ranking of image quality. Additionally, Guo et al. [7] introduced the Underwater Image Ranking Dataset (URankerSet), which features 890 underwater images along with enhancements from nine different methods.

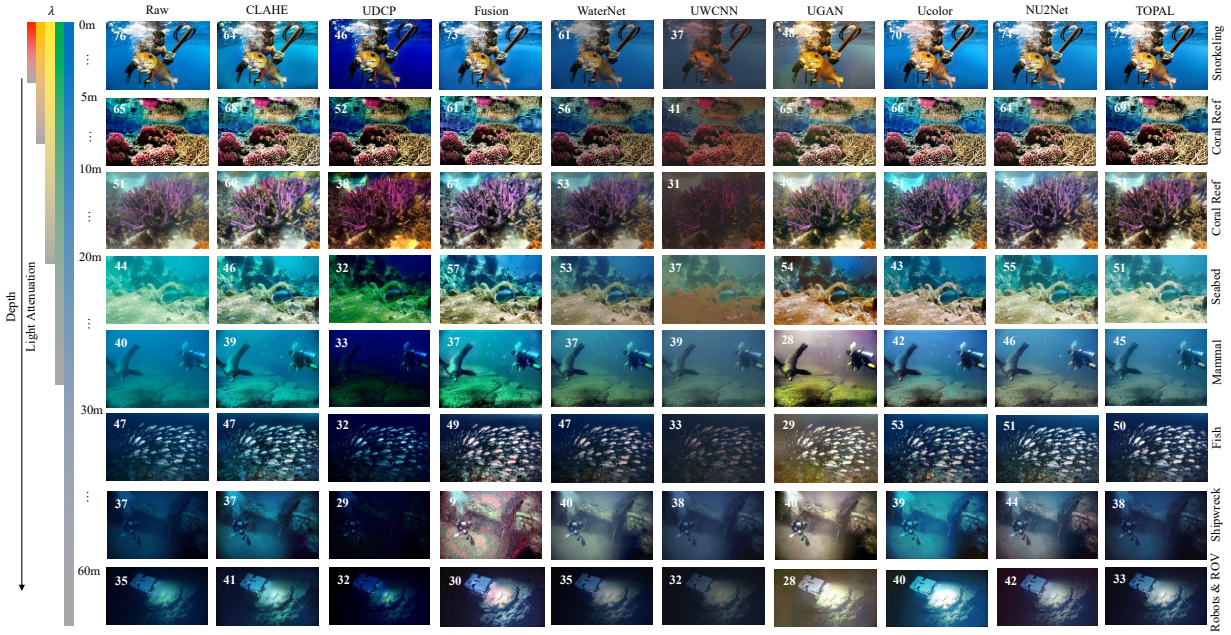

Fig. 2. Example images from the proposed LUIQD dataset. The LUIQD dataset consists of 64,180 images, including real-world underwater images and their corresponding enhancement results generated by 9 UIE methods. The raw images are organized based on the depth at which they were collected, representing a diverse range of depths, water conditions, lighting conditions, and image categories. The quality score of each image is displayed in the upper-left corner of the interface used in the subjective user study.

This dataset uses pairwise comparisons to rank the images according to their visual quality. Hou et al. [21] constructed the Underwater Image Dataset (UID2021), comprising 60 original underwater images and 15 enhanced counterparts, utilizing a paired comparison method to rank them.

However, current underwater image quality datasets exhibit several limitations. Rank-based datasets primarily represent the relative quality of homologous images, failing to provide an assessment of global image quality. Datasets using the SS-ACR method often contain a small number of images across limited scenes, which complicates fulfilling the training needs of deep learning-based UIQA networks. Additionally, some of these datasets contain only real underwater images without any scoring developed for the enhanced images. Consequently, there is a clear need for the creation of a large-scale underwater image quality dataset that can facilitate the extensive training and evaluation of advanced UIQA methods.

III. The LUIQD Dataset

In this section, we detail the specifics of the proposed LUIQD dataset, covering aspects such as data collection and enhancement, the design of a subjective user study, and the validation and processing of perceptual quality scores. The LUIQD dataset assigns a global and absolute quality score to each underwater image (raw or enhanced), which directly measures the visual quality of the images. The score can be compared across scenes, image foregrounds, and enhancement algorithms.

A. Image Collection

The quality of underwater images is influenced by several factors, including the imaging environment, lighting conditions, water quality, and the subject matter. To thoroughly characterize the diverse underwater imaging scenarios, we selected a total of 6,418 raw underwater images, collected from the Flickr website [1]. The dataset encompasses typical underwater scenes such as coral reefs, shipwrecks, diving activities, seabeds, and various underwater structures. The distribution of image categories within the dataset is detailed in Fig. 3. It should be noted that the images in the dataset are authorized for copying, distribution, display, and performance as works. However, some images may be used in derivative works, while others may not.

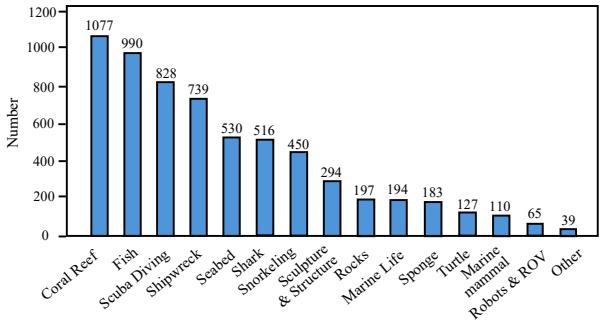

Fig. 3. Category statistics of the 6,418 real-world raw underwater images in the proposed LUIQD dataset.

With these raw images, we utilized 9 representative UIE algorithms to enhance or restore the underwater images. These UIE algorithms span traditional physics-based methods (e.g., UDCP [27]), non-physics-based methods (e.g., Fusion [6] and CLAHE [28]), and the latest deep learning-based algorithms

[1] https://www.flickr.com/

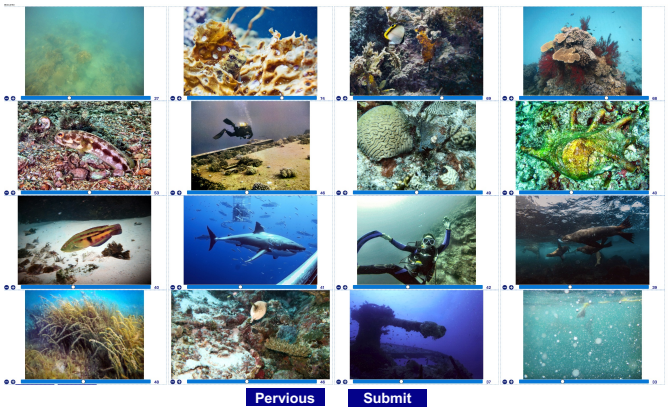

Fig. 4. The interface program in the subjective user study.

(such as UWCNN [29], UColor [30], UGAN [31], WaterNet [1], TOPAL [32], and NU2Net [7]). To ensure the effectiveness of these enhancement algorithms, we utilized the source code and training weights outlined in the respective publications for the deep learning-based methods. Consequently, the dataset comprises a total of 64,180 images.

Fig. 2 showcases examples of the raw underwater images alongside their enhancements by various UIE algorithms. It is shown that the enhancement quality of each UIE method varies across different image scenes. In some cases, the quality of the enhanced images may even be worse than that of the raw underwater images. Such significant variations in quality make these images particularly suitable for image quality assessment.

B. Subjective User Study

A subjective user study was organized to annotate the perceptual quality scores using the SS-ACR methodology. Observers were asked to rate the quality of each image on a scale from 0 to 100. Following the standard ITU-R BT.500-12 recommendation [33], we employed an ASUS ProArt Display PA32UCR monitor for the annotation of quality scores and a Datacolor Spyder X2 color calibrator to ensure true color accuracy. The monitor was set to a resolution of 3840×2160 pixels (4K UHD) with a 16:9 aspect ratio. The viewing distance was maintained at approximately twice the display height. All tests were conducted in a standard office environment with relatively consistent ambient lighting, minimizing the impact of external environmental factors on the assessment.

To ensure that the perceptual quality scores accurately represent the global and absolute quality of the images, we designed an interface program that displays multiple images simultaneously and allows observers to score them concurrently. Inspired by users observing underwater images on social media or in robotic vision tasks, the interface is built on an HTML-based website that randomly selects and displays 16 images from the total pool of 64,180 images, as illustrated in Figure 4. A single display may choose images from different original sources enhanced by various algorithms. During the scoring process, observers are encouraged to compare the images they are scoring with surrounding images to better gauge the overall quality within the entire image set. We encourage observers to evaluate the quality of underwater images in five aspects: color, contrast, texture, visibility, and foreground recognition. The detailed description of the scoring criteria is given in Table II.

TABLE II
SCORING CRITERIA FOR SUBJECTIVE USER STUDY

Score	Decsription
0-20	The color is completely distorted, contrast is completely non-existent, texture is very blurry, visibility is dim, foreground is completely unercognizable.
20-40	The color is deviated from nature, contrast is extremely low, texture is severely blurred, details is difficlut to identify, a few areas can distinguish the foreground from the background.
40-60	The color has a certain sense of nature, contrast is identified, texture is blurred, overall image is a littel burry, the foreground is partially recognizable.
60-80	The color maintains fidelity, contrast is moderate, texture is basically clear, most areas are clearly visible, outlines of foreground are clear and easy to identify.
80-100	The color is natural and true, contrast is clearly distinguished, texture is clear and accurately identified, all areas are clearly visible, details of foreground are clear and easy to identify.

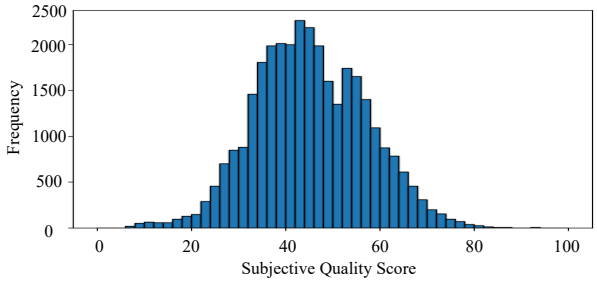

Fig. 5. Histogram of subjective quality scores in the LUIQD dataset.

Traditional subjective user studies typically require multiple observers to score the quality of an image and calculate a Mean Opinion Score (MOS) [34] [35] [36]. However, this method presents particular challenges when applied to evaluating underwater enhanced images. Firstly, the unique color properties of these images significantly impact quality assessments, necessitating the use of the same monitor models across all studies to ensure consistency. Secondly, variations in lighting conditions and display distances can influence observers' perceptions of image clarity and contrast, requiring that all evaluations be conducted within a controlled, consistent setting. This requirement can make large-scale image annotation both costly and time-consuming. To address these challenges, we adopted a modified approach: one observer was tasked with scoring all 64,180 images to create an annotation set, while 16 observers were invited to score a randomly selected subset of 2,000 images individually for validation purposes (the validation set). All observers were novices in underwater image quality assessment. Informed consent was obtained from all observers for our subjective experiment.

C. Quality Score Validation and Processing

In this subsection, we outline the methodology used to verify the reliability and feasibility of the LUIQD dataset.

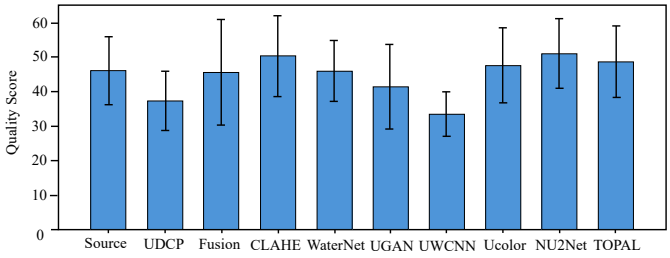

Fig. 6. Mean and standard deviation of subjective quality scores corresponding to each UIE method in the LUIQD dataset.

Initially, we performed outlier removal on the raw data collected from the validation set, following the ITU-R BT.500-12 recommendation [33]. We set the confidence interval for detecting and eliminating outliers at 95%. Specifically, for the i-th image, the 95% confidence interval for the user's subjective score is calculated as:

$$\left[\mu_i - 1.96 \times \frac{\sigma_i}{\sqrt{N}}, \mu_i + 1.96 \times \frac{\sigma_i}{\sqrt{N}}\right] \tag{1}$$

where N refers to the number of observers, and μ_i and σ_i represent the mean and standard deviation of the subjective scores, respectively. Any subjective scores that fall outside this confidence interval are considered outliers and are subsequently removed from the set. The final quality score for the i-th image is then calculated as the mean of the remaining subjective quality scores after outlier removal. Building on this refined data, we further verify the trustworthiness of the subjective scores in the annotation set by comparing them to the validation set scores using correlation and consistency analyses.

The intra-class correlation coefficient (ICC) [37] [38] is a widely used statistical measure for evaluating the reliability or consistency of subjective quality scores provided by multiple observers assessing the same images [34]. This metric is essential for determining the reliability of perceptual quality scores. Therefore, we calculated the $ICC(2,k)$ between the validation and annotation sets, taking into account the distribution of observers who rated the subjective scores. The obtained $ICC(2,k)$ value was 0.8200, indicating excellent reliability. Additionally, the 95% confidence interval for this measure suggests that the true ICC value likely falls between 0.80 and 0.84, affirming a high degree of certainty about the consistency and reliability of the subjective assessments.

Furthermore, we also employed the Spearman rank-order correlation coefficient (SROCC) [39] to verify the correlation between the validation and annotation sets. The obtained SROCC achieved a value of 0.7609, demonstrating a strong correlation between the two sets. This robust correlation supports the plausibility of the data within the annotation set. Based on the above analysis, we accept the plausibility of the annotation set and finalize it as the proposed LUIQD dataset.

The subjective quality distribution of the entire LUIQD dataset is illustrated in Fig. 5, showing a spread across the entire axis with a distribution that approximates a normal curve. This distribution ensures a thorough evaluation of image quality under various conditions. Fig. 6 presents the mean and standard deviation of the perceptual quality scores for both the raw images and different UIE methods. We will provide the complete original annotation set and validation set data to support further research and analysis.

IV. The PAUQA Network

Inspired by human visual perception models, we propose the Perception-Aware Underwater Image Quality Assessment Network (PAUQA-Net). This network leverages an efficient convolutional attention-based vision Transformer architecture, incorporating a multi-scale dilated convolutional patch embedding module inspired by the human visual system's multi-scale processing mechanism [40]. Additionally, we introduce a multi-path Transformer encoder module, which facilitates the fusion of local and global information in vision perception [41]. Given the sensitivity of human perception to both clarity and color in underwater images [13], the network extracts features from both the luminance and chrominance channels for joint quality score estimation.

A. Overall Architecture

The architecture of the PAUQA-Net is illustrated in Figure 7, which processes the input image in both the RGB and YCbCr color spaces. The input RGB image undergoes processing through a four-stage convolutional attention-based vision Transformer network, designed to extract quality-related features. Each stage of this network consists of three main components: (1) a multi-scale dilated convolutional patch embedding module that captures tokens at multiple scales, (2) a multi-path local and global feature encoder block that processes the tokens in parallel, and (3) a joint feature interaction module that aggregates the extracted features. This structure allows the network to capture and combine both fine and coarse details for a comprehensive quality assessment.

In addition to the RGB space, the YCbCr color space is also utilized. Here, the luminance channel (Y channel) is processed by the SFEN to extract sharpness features, while the chrominance channels (Cb and Cr channels) are processed by the CFEN to extract color features. Both SFEN and CFEN are convolutional neural networks, with CFEN further incorporating self-attention mechanisms. These sharpness and color features are then merged in the joint feature interaction layer to enhance the prediction of overall image quality. The network ultimately outputs a global quality score using a final linear layer.

B. Multi-Scale Image Feature Extraction

An efficient conv-attentional vision Transformer network is utilized for extracting quality-related features and performing joint predictions using both luminance and chrominance information. This network processes multi-scale image features, from local to global, in parallel through a multi-path structure.

Initially, the RGB input image $I \in \mathbb{R}^{C\times H\times W}$ is processed by a convolutional stem block, which consists of two convolution modules. Each convolution module includes a normalization layer, a Hardswish activation function, and a 3×3

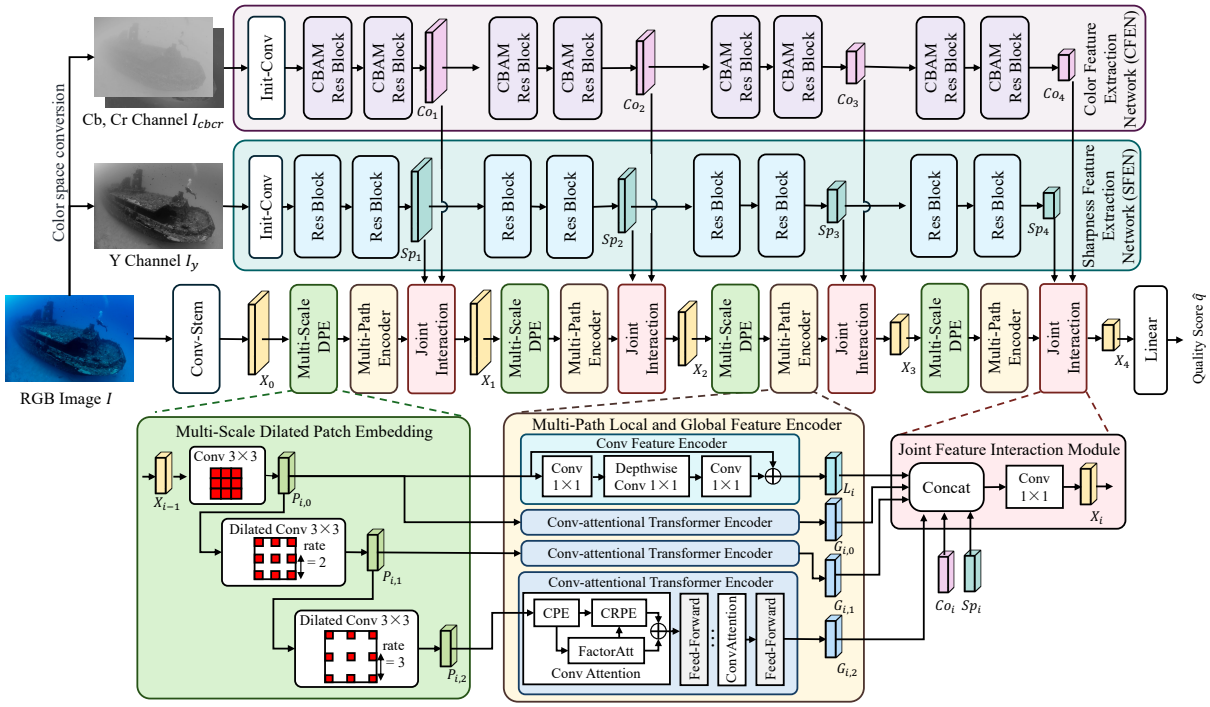

Fig. 7. The architecture of the proposed PAUQA-Net.

convolutional layer. The convolutional stem block produces the feature map $X_0 \in \mathbb{R}^{C_0 \times H/4 \times W/4}$, where $C_0 = 64$. After passing through the initial block, the feature map is further processed by four multi-scale dilated convolutional patch embedding modules, multi-path local and global feature encoder blocks, and a joint feature interaction module. This results in feature maps represented as $X_i \in \mathbb{R}^{C_i \times H_i \times W_i}$ at stage i, where $(i \in \{1, 2, 3, 4\})$.

1) Multi-Scale Dilated Convolutional Patch Embedding Module: Image quality is influenced by both local details and global composition [17], which aligns with the multi-scale processing mechanism in human visual perception. Convolutional patch embedding is a straightforward method for multi-scale feature extraction [42], while dilated convolution effectively expands the receptive field without increasing computational burden [43]. Based on this consideration, we employ dilated convolutions with various sampling rates to extract overlapping patches, thereby enhancing the representation of cross-scale and cross-space image quality information.

In stage i, the input feature or token map X_{i-1} from the previous stage is processed by the DCPE $F_{k,r}(\cdot)$. The DCPE consists of a dilated convolution layer with a $k \times k$ kernel and a sampling rate r, followed by batch normalization and Hardswish activation functions [44]. Mathematically, this process is represented as:

$$P_i = F_{k,r}(X_{i-1}), \tag{2}$$

where $X_{i-1} \in \mathbb{R}^{C_{i-1} \times H_{i-1} \times W_{i-1}}$ represents the input token map with resolution $H_{i-1} \times W_{i-1}$ and C_{i-1} channels. $P_i \in \mathbb{R}^{C_i \times H_i \times W_i}$ denotes the output token map with a resolution of $H_i \times W_i$ and C_i channels. This output is then flattened to size $H_i W_i \times C_i$ and normalized using layer normalization before being passed to the transformer feature encoders. The stride of the dilated convolution is set to 2 to reduce spatial resolution, and 1 otherwise.

To generate multi-scale token maps, multiple DCPEs with different sampling rates are applied in series at each stage. This involves stacking consecutive convolution operations, each maintaining the same number of channels, to progressively expand the receptive field layer by layer. This series connection structure enables the integration of features from different scales in a non-linear fashion, enhancing the response to quality-related features. In stage i, a total number of j DCPE operations are performed, resulting in the set of token maps $\mathbb{P}_i$, expressed as:

$$\mathbb{P}_i = \{P_{i,0}, P_{i,1}, ..., P_{i,j-1}, \}, \tag{3}$$

in which the input to the $0-th$ DCPE is X_0, while subsequent DCPEs take the output of the pervious DCPE as their input. This can be mathematically represented as:

$$\begin{cases} P_{i,0} = F_{k_0,r_0}(X_{i-1}) \\ P_{i,n} = F_{k_n,r_n}(P_{i,n-1}), n \in [1, j-1] \end{cases}, \tag{4}$$

where $k_n, k_1, ..., k_{j-1}$ and $r_0, r_1, ..., r_{j-1}$ denote the kernel sizes and sampling rates of each DCPE, respectively. These token maps are then individually processed by multi-scale convolutional and transformer encoders. Specifically, in stage 1, two DCPEs are applied, each with a 3×3 kernel and sampling rates of 1 and 2 to minimize computational demands. In stages 2 to 4, three DCPEs are utilized, each with the same 3×3 kernel size and sampling rates ranging from 1, 2 to 3.

2) Multi-path Local and Global Feature Encoder: When assessing image quality, it is essential to follow the local-global information fusion mechanism of human perception

and consider both global context and local relationships. Transformers leverage self-attention mechanisms to prioritize different regions of an image, effectively capturing global dependencies. In contrast, CNNs excel at extracting hierarchical features and focusing on local spatial correlations, making them efficient at capturing low-level image contexts.

Based on this consideration, we utilize convolutional encoders and Transformer encoders to extract local and global features from image token maps, respectively, following the approach in [42]. These encoders process the image token map in parallel, forming a multi-path structure. Specifically, in stage i, the local feature $L_i \in \mathbb{R}^{C_i \times H_i \times W_i}$ is extracted using a depthwise convolution feature encoder. This encoder consists of a 1×1 convolution layer, a 3×3 depthwise convolution, and another 1×1 convolution layer. A residual connection is integrated within this module for feature prediction. The input to this encoder is the token map $P_{i,0}$.

To extract global features, an efficient conv-attentional transformer module is employed. For input image tokens $P_{i,j} \in \mathbb{R}^{C_i \times N_i}$ in stage i, where N_i and C_i denote the number of tokens and the embedding dimension, respectively, this module utilizes the following processes to extract features:

(a) The convolutional position encoding integrates positional information into the Transformer model. In this process, a convolutional operation $Conv(\cdot)$ is first performed on the input token map $P_{i,j}$ to generate spatial position encodings Cpe_i, representing the spatial context information of the token. Subsequently, we combine the input token map $P_{i,j}$ with its position encoding using elementwise addition operation, forming the Transformer input $P'_{i,j} \in \mathbb{R}^{C_i \times N_i}$:

$$P'_{i,j} = P_{i,j} + Cpe_{i,j} = P_{i,j} + Conv(P_{i,j}). \tag{5}$$

(b) The factorized attention mechanism processes the self-attention operation:

$$FactorAtt(P'_{i,j}) = \frac{Q_{i,j}}{\sqrt{C_{i,j}}}(softmax(K_{i,j})^\top V_{i,j}). \tag{6}$$

where $Q_i, j, K_i, j, V_i, j \in \mathbb{R}^{C_i \times N_i}$ represent the linearly projected queries, keys, and values. N_i is the number of tokens and C_i is the embedding dimension. (c) The convolutional relative position encoding is used to generate a relative attention map. In this process, a 2-D depthwise convolution $DepthwiseConv(W, V_{i,j})$ with a window size of $M \times M$ and kernel weights W is first applied to reshaped image tokens, extracting feature maps $Crpe_{i,j}$ tthat encapsulate spatial information. Afterward, we obtain the relative attention map $EV_{i,j} \in \mathbb{R}^{C_i \times N_i}$ using the Hadamard product $\circ$:

$$EV_{i,j} = Q_{i,j} \circ Crpe_{i,j} = Q_{i,j} \circ DepthwiseConv(W, V_{i,j}). \tag{7}$$

Following the design in [**?**], we use global average pooling (GAP) instead of a class token in the attention operation. Thus, the entire conv-attention operation can be represented as:

$$ConvAtt(X'_{i,j}) = \frac{Q_{i,j}}{\sqrt{C_{i,j}}}(softmax(K_{i,j})^\top V_{i,j}) + EV_{i,j}. \tag{8}$$

The Feed-Forward operation is then applied to obtain the output global features $G_{i,j} \in \mathbb{R}^{C_i \times N_i}$ of the encoder.

For the token map extracted by the multi-scale DCPE, we employ two parallel Transformer encoders to extract two distinct sets of global features in stage 1, while three Transformer encoders extract three sets of features in stages 2 to 4.

3) Joint Feature Interaction Module: To comprehensively evaluate the quality of images, a feature interaction layer that integrates various types of image features is designed. This layer synchronizes the local feature extracted by the convolutional encoder with the global features derived from the Transformer encoders. Additionally, it incorporates the sharpness and color features extracted from the image's luminance and chrominance channels.

With the local feature $L_i \in \mathbb{R}^{C_i \times H_i \times W_i}$ and the 2D-reshaped global feature $G_{i,j} \in \mathbb{R}^{C_i \times H_i \times W_i}$ in stage i, the concatenation operation is used to aggregate them with the sharpness features $Sp_i \in \mathbb{R}^{C_i \times H_i \times W_i}$ and color features $Co_i \in \mathbb{R}^{C_i \times H_i \times W_i}$,

$$A_i = Concat([L_i, G_{i,0}, G_{i,1}, ..., G_{i,j}, Sp_i, Co_i]), \tag{9}$$

where j is the number of global features, $A_i \in \mathbb{R}^{(j+3)C_i \times H_i \times W_i}$ is the aggregated feature. After that, we apply a joint interaction function $J(\cdot)$ to produce the final feature map $X_i \in \mathbb{R}^{C_{i+1} \times H_i \times W_i}$ with the channel dimension C_{i+1} for the next stage. The function $J(\cdot)$ is constructed using a 1×1 convolution.

C. Color and Sharpness Feature Extraction

The luminance and chrominance channels in underwater images carry crucial information related to the human vision perception-consistent quality degradation. For instance, effects such as haze and local structural losses are typically visible in the luminance channel, while color distortions, such as color bias and over-enhancement, can be identified through the chrominance channel [13]. To facilitate this analysis, underwater images are converted from the RGB color space to the YCbCr color space, enabling separate processing of the luminance and chrominance components. Two distinct convolutional networks [45], SFEN and CFEN, are then used to extract multi-scale color and sharpness features from these channels.

1) Sharpness Feature Extraction Network: To assess the degradation of clarity and detail, primarily affected by phenomena such as light absorption, scattering, and backscatter, we extract sharpness features from the luminance channel (Y channel) in YCbCr space. For the input $I_y \in \mathbb{R}^{1 \times H \times W}$ with a resolution of $H \times W$, it is initially processed through two convolution layers followed by a max-pooling operation, resulting in initial feature maps $Sp_0 \in \mathbb{R}^{C_0 \times H/4 \times W/4}$.Subsequently, these feature maps undergo four down-sampling operations and convolution blocks, enabling the extraction of sharpness features. In each stage i,, the sharpness feature maps are represented as $Sp_i \in \mathbb{R}^{C_i \times H_i \times W_i}$.

Specifically, the convolutional blocks designed for extracting sharpness features are based on the ResBlock architecture [46] [47]. The sharpness features of each stage are extracted by 2 ResBlocks connected in series. The extracted sharpness image features are then utilized as inputs for the multi-patch joint feature interaction modules.

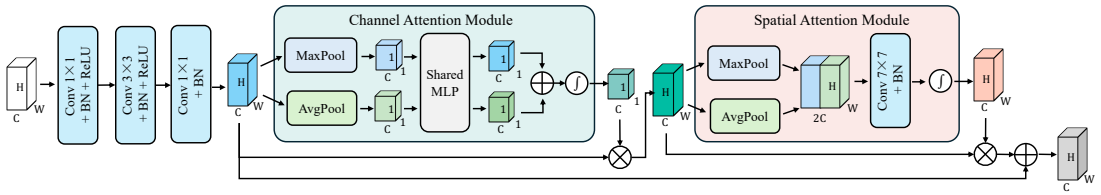

Fig. 8. Structure of the ResBlock integrated with CBAM in the proposed color feature extraction network.

2) Color Feature Extraction Network: Throughout the underwater image, different wavelengths are progressively absorbed at varying depths, resulting in color absorption and color casts. In localized areas, factors such as artificial lighting, light reflection and refraction, and shading can cause local color inhomogeneity. Relevant color characteristics can be extracted from the chrominance channels (Cb and Cr channels) in YCbCr space. Based on this consideration, we employ the visual attention mechanism to enhance the extraction of color features. The channel attention mechanism (CAM) facilitates the feature representation process with cross-channel processing, crucial for revealing the relationship across color channels and the human visual perception of color loss. The spatial attention mechanism (SAM) selectively focuses on important spatial regions in images, while disregarding less relevant areas. This guides the network to concentrate specifically on regions where color loss is more influential than in others.

We develop the visual attention mechanism for color feature extraction based on the convolutional block attention module (CBAM) [48] and ResBlock. The structure of CBAM is shown in Figure 8. CBAM computes the attention map sequentially in both the channel and spatial dimensions and integrates these attentions with the input features through element-wise multiplication. Taking the chrominance channels $I_{cbcr} \in \mathbb{R}^{2\times H\times W}$ as input, the feature map $Co_0 \in \mathbb{R}^{C_0\times H/4\times W/4}$ is obtained through two initial convolution layers and a max-pooling operation. Subsequent extraction of color features $Co_i \in \mathbb{R}^{C_i\times H_i\times W_i}$ occurs through a ResBlock integrated with CBAM and down-sampling operations. The color features of each stage are extracted by 2 ResBlocks integrated with CBAM and connected in series.

D. Image Quality Prediction

Finally, to predict the quality score, a fully connected layer (FC) [49] [50] is used to map the extracted features X_4 to a specific image quality score. The calculation of the image quality score for each image is conducted as follows:

$$\hat{q} = FC(X_4), \tag{10}$$

where $\hat{q}$ refers to the predicted image quality score.

To train the entire PAUQA-Net network, we adopt a dataset $\mathcal{D} = \{I^n, q^n\}_{n=1}^N$, where q^n denotes the corresponding subjective quality score of image I^n, and N represents the number of images in $\mathcal{D}$. We minimize the $\mathcal{L}_1$ loss, which is defined as:

$$\mathcal{L}_1 = \frac{1}{N}\sum_{i=1}^{N}|\hat{q}^n - q^n|, \tag{11}$$

where $\hat{q}^n$ and q^n correspond to the predicted and true image quality scores of image I^n, respectively. The workflow of the training stage of the proposed PAUQA-Net is summarized in Algorithm 1.

Algorithm 1 Training Stage of PAUQA-Net Model.

Input: UIQA training set $\mathcal{D} = \{I^n, q^n\}_{n=1}^N$.
Output: Predict image quality score $\hat{\mathcal{Q}} = \{q^n\}_{n=1}^N$.

```
Initialize all the parameters of the PAUQA model;
for iteration = 1, 2, ... do
    Sample a batch of k images from D;
    for batch = 1, 2, ... do
        Convert {I^n}_{n=1}^k to YCbCr {I^n_{ycbcr}}_{n=1}^k.
        // Color Feature Extraction.
        for stage i from 1 to 4 do
            Extract feature map {Co_i^n}_{n=1}^k using {I^n_{cbcr}}_{n=1}^k;
        end for
        // Sharpness Feature Extraction.
        for stage i from 1 to 4 do
            Extract feature map {Sp_i^n}_{n=1}^k using {I_y^n}_{n=1}^k;
        end for
        // Multi-Scale image Feature Extraction.
        for stage i from 1 to 4 do
            Extract global and local feature map
            {L_i^n, G_{i,0}^n, G_{i,1}^n, ..., G_{i,j}^n}_{n=1}^k using {I^n}_{n=1}^k;
            Concat aggregated feature map {A_i^n}_{n=1}^k;
            Predict feature map {X_i^n}_{n=1}^k;
        end for
        // Image Quality Prediction.
        Output {q̂^n}_{n=1}^k using FC and {X_4^n}_{n=1}^k.
        Update model parameters by using L_1 with {q^n}_{n=1}^k;
    end for
end for
```

V. Experimental Setup

In this section, the experimental setup is described, including the implementation details, datasets, baselines, and evaluation metrics.

A. Datasets

The proposed LUIQD dataset was randomly divided in an 80:20 ratio into training and testing datasets. Specifically, the training set comprises 5,135 raw images along with their corresponding 46,215 enhanced images, while the testing set includes 1,282 raw images and their corresponding 11,538 enhanced images. All images were resized to a uniform dimension of 384×384 pixels to serve as inputs for the

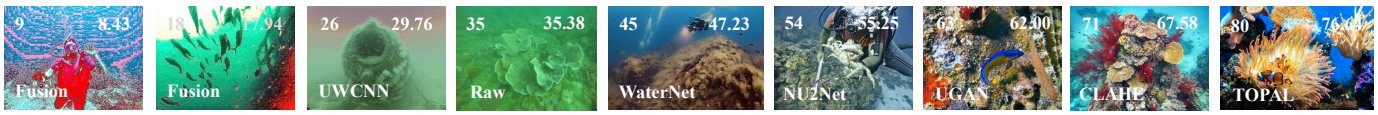

Fig. 9. Underwater images of varying quality levels, annotated with their corresponding perceptual quality scores and PAUQA-Net prediction results. The UIE algorithm applied is indicated in the bottom-left corner of each image. The scores in the upper-left corner represent subjective quality scores, while those in the upper-right corner display the PAUQA-Net prediction results.

network. To validate the quality assessment capability for real underwater images, we also tested various IQA methods using all the raw underwater images in the LUIQD dataset. This specific subset of data is referred to as LUIQD-Raw.

In addition to the proposed LUIQD dataset, we evaluated the performance of the PAUQA-Net network using two additional datasets: the UWIQA dataset [19] and the SAUD dataset [16]. The UWIQA dataset comprises 890 real-world underwater images, each annotated with a corresponding MOS quality score. The SAUD dataset is designed for assessing the quality of underwater image enhancement algorithms and includes 100 raw underwater images along with their 1,000 enhanced versions, scored using the B-T score from the DS-PC methodology. Both the training and testing sets for these datasets were randomly divided in an 80:20 ratio.

To further validate the effectiveness and robustness of the proposed PAUQA-Net, we also performed comparisons on two in-air image quality assessment datasets: the TID2013 dataset [51] and the EVA dataset [52]. Among them, the TID2013 dataset is a technical quality assessment dataset that includes 25 natural images and 3,000 distorted images, representing different distortion types and levels. The EVA dataset is an aesthetic quality assessment dataset containing 5,101 images. For the EVA dataset, we randomly sampled 80% of the images for training and 20% for testing. For the TID2013 dataset, we used the training-testing set partitioning strategy from the original paper.

B. Baselines

We evaluated the performance of the proposed PAUQA-Net network by comparing it with other NR-IQA methods, including traditional unsupervised general-purpose IQA methods such as BRSIQUE [53], BIQI [54], BLIINDS [55], NIQE [56], and ILNIQE [57], as well as deep learning-based IQA methods, including CNNIQA [58], DBCNN [59], WaDIQaM-NR [60], UNIQUE [61], HyperIQA [62], NIMA [63], and MUSIQ [17]. We also compared with the TQ4AQ [64] method, which disentangles aesthetic quality from technical quality. For methods specifically designed for UIQA, we evaluated CCF [12], UIQM [9], UICQE [8], FDUM [19], URanker [7], and ATUIQP [18]. The first four methods are traditional feature-based unsupervised methods, while the latter two are supervised approaches.

C. Evaluation Metrics

To assess the effectiveness of the PAUQA-Net method, we used four widely recognized statistical metrics: SROCC, Kendall's Rank Correlation Coefficient (KROCC), Pearson's Linear Correlation Coefficient (PLCC), and the Root Mean Square Error (RMSE) [65]. A superior IQA method typically exhibits higher absolute values for SROCC, KROCC, and PLCC, along with a lower RMSE.

For learning-based IQA approaches, the networks are re-trained from scratch on the training set before calculating validation results on the test set. These results are then used to compute evaluation metrics, which are compared to the ground-truth quality scores. For objective IQA methods, these metrics are calculated by correlating the objective scores with the ground-truth quality scores.

Following the guidelines of the Video Quality Experts Group (VQEG), it is recommended that, before calculating PLCC and RMSE, the outputs are mapped to subjective quality scores using a five-parameter logistic function [61], defined as:

$$m(s) = \theta_1(\frac{1}{2} - \frac{1}{1+exp(\theta_2 \cdot (s-\theta_3))}) + \theta_4 \cdot s + \theta_5, \quad (12)$$

where $m(s)$ denotes the mapped quality score, s represents to the outputs of IQA methods, and $\theta_1 \sim \theta_5$ are the mapping parameters.

D. Implementation Details

The proposed PAUQA-Net network was implemented using the PyTorch framework and deployed on an NVIDIA GeForce RTX 3090 graphics card. We used the Adam optimizer, setting the initial learning rate to 1×10^{-5} and applying a weight decay of 1×10^{-4}. The training phase was conducted over 100 epochs with a batch size of 16.

VI. Experimental Results

In this section, we present the experimental results obtained using both underwater and in-air image quality assessment datasets.

A. Quality Prediction Performance Comparison on LUIQD

First, we conducted a comprehensive evaluation of the proposed PAUQA-Net network by comparing it with other IQA methods using both the LUIQD and LUIQD-Raw datasets. Table VI-A presents the results of this quantitative comparison.

The results reveals that unsupervised IQA methods initially designed for natural images like NIQE [56], LINIQE [57], and BIQI [54], struggle to accurately evaluate the quality of underwater images, due to the unique degradation characteristics of underwater environments are rarely encountered in natural images. Furthermore, unsupervised methods tailored specifically for underwater images, like UIQM [9], UICQE [8], CCF [12], and FDUM [19]achieve only moderate accuracy.

TABLE III
EXPERIMENTAL RESULTS ON THE LUIQD AND LUIQD-RAW DATASETS, EXPRESSED IN TERMS OF SROCC, KROCC, PLCC, AND RMSE VALUES. THE TOP TWO RESULTS IN EACH COLUMN ARE HIGHLIGHTED IN BOLD AND UNDERLINED, RESPECTIVELY.

Method	LUIQD				LUIQD-Raw			
	SROCC↑	KROCC↑	PLCC↑	RMSE↓	SROCC↑	KROCC↑	PLCC↑	RMSE↓
BRSIQUE [53]	0.2953	0.2009	0.3165	11.2570	0.2961	0.2033	0.2669	9.2843
BIQI [54]	0.1482	0.1007	0.1555	11.7684	0.1988	0.1360	0.2519	9.3233
BLIINDS [55]	0.1689	0.1195	0.1879	11.6416	0.2235	0.1564	0.2199	9.3977
NIQE [56]	0.3924	0.2690	0.3844	10.9842	0.4123	0.2854	0.4056	8.8055
ILNIQE [57]	0.5393	0.3745	0.5178	10.1252	0.4709	0.3258	0.4495	8.6055
CNNIQA [58]	0.6829	0.5013	0.6555	8.9439	0.6574	0.4780	0.6448	7.4600
DBCNN [59]	0.5903	0.4203	0.5712	10.6214	0.3295	0.2290	0.2977	10.5367
WaDIQaM-NR [60]	0.8001	0.6122	0.7882	7.3066	0.7126	0.5241	0.6911	6.9925
UNIQUE [61]	0.8562	0.6748	0.8009	7.6902	0.8262	0.6431	0.7378	7.2372
NIMA [63]	0.8393	0.6558	0.8285	6.7801	0.8171	0.6333	0.8124	5.6839
HyperIQA [62]	0.8328	0.6479	0.8234	7.1803	<u>0.8313</u>	<u>0.6476</u>	<u>0.8209</u>	5.9164
MUSIQ [17]	0.8513	0.6711	0.8272	6.9047	0.8299	0.6455	0.8083	5.7750
TQ4AQ [64]	<u>0.8609</u>	<u>0.6811</u>	<u>0.8551</u>	<u>6.1450</u>	0.8162	0.6316	0.8092	6.0635
CCF [12]	0.4186	0.2930	0.4174	10.7631	0.4597	0.3245	0.4718	8.4940
UIQM [9]	0.5245	0.3659	0.4377	10.7336	0.5141	0.3569	0.5135	8.2661
UICQE [8]	0.3418	0.2377	0.3573	11.0668	0.4502	0.3167	0.4701	8.5029
FDUM [19]	0.5368	0.3796	0.5209	10.1135	0.5991	0.4303	0.5998	7.7086
URanker [7]	0.8489	0.6687	0.8422	6.4045	0.8252	0.6416	0.8199	**5.5382**
ATUIQP [18]	0.8394	0.6558	0.8314	6.6149	0.8070	0.6199	0.8049	5.7470
PAUQA-Net (Ours)	**0.8745**	**0.6991**	**0.8635**	**6.0703**	**0.8413**	**0.6605**	**0.8351**	<u>5.5532</u>

This may be due to their dependence on a limited set of manually designed features that do not fully capture the complexity of human perception — particularly the content and context of the image itself — and are less adaptable across diverse underwater scenarios.

Compared to unsupervised methods, data-driven IQA methods align more consistently with human perception. Deep learning-based IQA methods designed for natural images, such as MUSIQ [17], HyperIQA [62], TQ4AQ [64], and NIMA [63], effectively leverage the correlations between image features and human perception scores. The UIQA networks, such as URanker [7] and ATUIQP [18], meticulously consider factors that impair underwater image quality, enhancing their simulation of human visual perception. However, these methods are often not stable enough and cannot achieve accurate quality estimation of both enhanced images and raw underwater images.

In comparison to other prevalent IQA methods, the proposed PAUQA-Net exhibits superior performance, achieving the highest SROCC, KROCC, and PLCC values, thereby indicating enhanced consistency with the human perception of underwater images. Specifically, the SROCC for PAUQA-Net is 0.8745, representing a 0.0136 gain over the second-best method, TQ4AQ. Similarly, PAUQA-Net also records high KROCC and PLCC scores of 0.6991 and 0.8635, respectively, underscoring its effectiveness. Additionally, it scores the lowest RMSE value at 6.0703. This highlights PAUQA-Net's enhanced ability to accurately reflect human perception of image quality to underwater enhanced images, representing a notable improvement over other IQA and UIQA methods. Additionally, PAUQA-Net demonstrates exceptional capability in assessing the quality of raw underwater images. Specifically, on the LUIQD-Raw subset, PAUQA-Net achieves the optimal value, with an SROCC value of 0.8413, a KROCC value of 0.6605, a PLCC value of 0.8351, and an RMSE value of 5.5532. These metrics illustrate that PAUQA-Net is a robust tool for not only validating the effectiveness of enhanced images but also for accurately assessing the quality of raw underwater images, aligning closely with human perception.

Further, we selected nine raw and enhanced underwater images of varying quality levels from the LUIQD dataset. The perceptual quality scores and the prediction results of PAUQA-Net are shown in Figure 9. It is evident that images with lower perceptual scores exhibit significant color casts or blurring. Conversely, images with higher scores display distinct details, enhanced sharpness, and vibrant colors. PAUQA-Net leverages an efficient image feature extraction network that focuses on color and sharpness, allowing it to closely align prediction results with subjective assessments.

Additionally, we present scatter plots comparing human-perceived quality scores with those predicted by different UIQA and IQA methods, as shown in Figure 10. Analysis of these scatter plots reveals that the proposed PAUQA-Net method demonstrates the highest correlation with human-perceived scores. This is evidenced by the scatter points of PAUQA-Net being more tightly clustered around the human perception scores, represented by the red line. In contrast, scatter plots for other IQA methods, particularly the unsupervised ones, show a more dispersed distribution of predicted results. This indicates a weaker correlation with human perception.

B. Quality Prediction Performance Comparison on UWIQA and SAUD

To further verify the effectiveness of the proposed PAUQA-Net network, we conducted comparative experiments using two additional underwater image quality datasets: the UWIQA dataset [19] and the SAUD dataset [16]. The comparison results on the UWIQA dataset, as displayed in Table IV, demonstrate that the proposed PAUQA-Net method surpasses other IQA methods in terms of quality assessment of the raw underwater images. Specifically, when compared to UIQA and IQA methods designed for natural images, PAUQA-Net has

TABLE IV
EXPERIMENT RESULTS ON THE UWIQA [19] AND SAUD [16] DATASETS, EXPRESSED AS SROCC, KROCC, PLCC, AND RMSE VALUES. THE TOP TWO RESULTS IN EACH COLUMN ARE EMPHASIZED IN BOLD AND UNDERLINED, RESPECTIVELY.

Method	UWIQA				SAUD			
	SROCC↑	KROCC↑	PLCC↑	RMSE↓	SROCC↑	KROCC↑	PLCC↑	RMSE↓
BRSIQUE [53]	0.5330	0.3981	0.5467	0.1278	0.3411	0.2330	0.3335	0.1998
BIQI [54]	0.4706	0.3434	0.4799	0.1340	0.2652	0.1824	0.2954	0.2024
BLIINDS [55]	0.3285	0.2488	0.2515	0.1478	0.1740	0.1203	0.2074	0.2073
NIQE [56]	0.5449	0.4213	0.5728	0.1252	0.1443	0.0935	0.1885	0.2081
ILNIQE [57]	0.6536	0.5028	0.6388	0.1175	0.2961	0.2032	0.3424	0.1991
CNNIQA [58]	0.7259	0.5768	0.7071	0.1096	0.1945	0.1306	0.3581	0.1985
DBCNN [59]	0.6981	0.5489	0.6789	0.1375	0.3757	0.2595	0.3622	0.2082
WaDIQaM-NR [60]	0.3357	0.2485	0.3480	0.3786	0.2396	0.1974	0.2536	0.2181
UNIQUE [61]	<u>0.7711</u>	<u>0.6142</u>	0.7396	0.1092	0.6082	0.6232	0.8220	**0.1408**
NIMA [63]	0.6688	0.5180	0.6520	0.1260	0.4734	0.3332	0.5298	0.1832
HyperIQA [62]	0.7255	0.5678	0.7036	0.1395	<u>0.6287</u>	<u>0.4587</u>	<u>0.6621</u>	0.1607
MUSIQ [17]	0.7375	0.5883	0.7222	0.1091	0.6177	0.4352	0.6221	0.1895
TA4AQ [64]	0.6710	0.5349	0.5334	0.1441	0.6976	0.5038	0.6770	0.1616
CCF [12]	0.4659	0.3563	0.4924	0.1329	0.2278	0.1557	0.2797	0.2034
UIQM [9]	0.4763	0.3579	0.4617	0.1354	0.2742	0.1878	0.3937	0.1968
UICQE [8]	0.6068	0.4613	0.6139	0.1205	0.3103	0.2143	0.2773	0.2036
FDUM [19]	0.7262	0.5761	<u>0.7399</u>	<u>0.1027</u>	0.1946	0.1311	0.2371	0.2058
URanker [7]	0.5449	0.4240	0.5565	0.1361	0.5881	0.4418	0.6006	0.1817
ATUIQP [18]	0.5269	0.4051	0.5362	0.1360	0.6181	<u>0.4587</u>	0.6157	0.1707
PAUQA-Net (Ours)	**0.8006**	**0.6558**	**0.7907**	**0.1008**	**0.7866**	**0.5983**	**0.7826**	<u>0.1507</u>

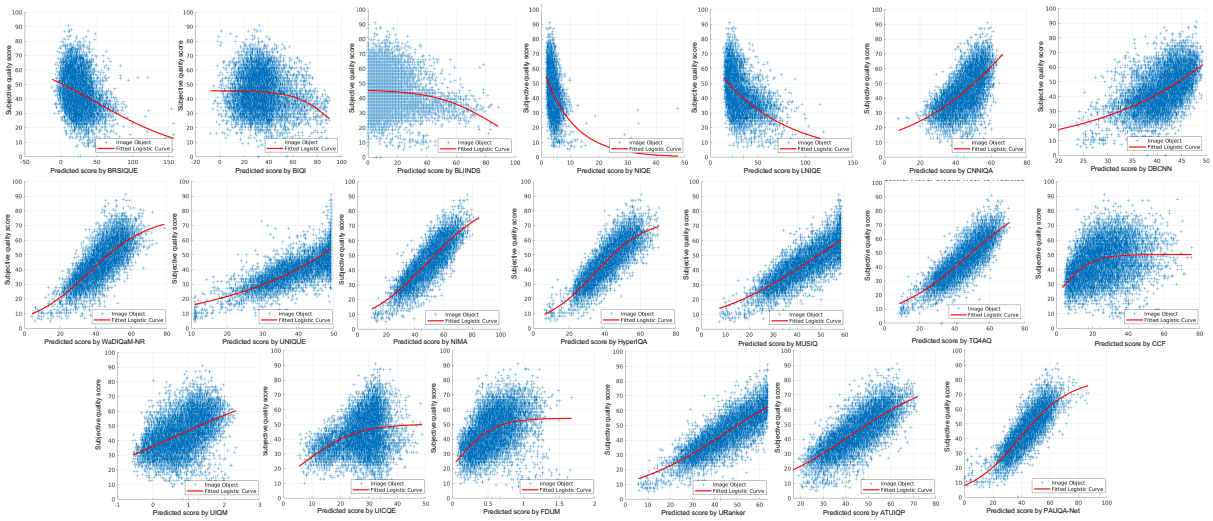

Fig. 10. Scatter plots of the predicted objective scores against the subjective quality scores of different UIQA and IQA methods on the proposed LUIQD validation set. The red line is the curve fitted using the logistic function.

been observed to enhance the SROCC from 0.7711 to 0.8006, the KROCC from 0.6142 to 0.6558, the PLCC from 0.7396 to 0.7907, and the RMSE from 0.1092 to 0.1008, thereby exceeding the performance of UNICQE. These results affirm the effectiveness of the proposed PAUQA-Net method in the specialized domain of UIQA for raw underwater images.

The comparison results for the SAUD dataset are presented in Table IV. These experimental outcomes highlight the excellent generalizability of the proposed PAUQA-Net method within the field of UIQA for enhanced images, achieving the highest scores in SROCC, KROCC, and PLCC. Specifically, the SROCC for PAUQA-Net reaches 0.7866, the KROCC reaches 0.5983, the PLCC reaches 0.7826, and the RMSE reaches 0.1507. These results underscore the robustness and validity of the PAUQA-Net method for assessing the quality of enhanced underwater images and demonstrate its adaptability across various data acquisition techniques.

In summary, experimental evaluations across various underwater image datasets have shown that the proposed PAUQA-Net method outperforms existing IQA and UIQA methods in terms of prediction accuracy and is more closely aligned with human perception models. This demonstrates its effectiveness in providing reliable and human-like assessments of image quality, making it a better choice for applications in underwater image analysis.

C. Cross Dataset Validation

To validate the generalization capabilities of the proposed PAUQA-Net network and mitigate the risk of overfitting, we applied the deep-learning-based IQA and UIQA model, initially trained on the LUIQD dataset, to the UWIQA dataset. Table V presents the results of this cross-dataset validation.

TABLE V
CROSS-DATASET EXPERIMENT RESULTS. EACH MODEL IS TRAINED ON THE LUIQD DATASET AND VALIDATED ON THE UWIQA VALIDATION SET. THE TOP TWO RESULTS IN EACH COLUMN ARE EMPHASIZED IN BOLD AND UNDERLINED, RESPECTIVELY.

Method	SROCC↑	KROCC↑	PLCC↑	RMSE↓
CNNIQA [58]	0.5374	0.4061	0.5548	0.1270
DBCNN [59]	<u>0.6650</u>	0.5167	0.6695	0.1134
WaDIQaM-NR [60]	0.5127	0.3892	0.5300	0.1295
UNIQUE [61]	0.6642	**0.5264**	0.6690	0.1135
NIMA [63]	0.6505	0.5100	<u>0.6799</u>	<u>0.1120</u>
HyperIQA [62]	0.6581	0.5220	0.6718	0.1131
MUSIQ [17]	0.6347	0.4996	0.6578	0.1150
URanker [7]	0.5823	0.4531	0.6038	0.1217
ATUIQP [18]	0.5205	0.4001	0.5631	0.1262
PAUQA-Net (Ours)	**0.6658**	<u>0.5252</u>	**0.6892**	**0.1106**

The experimental results indicate that the PAUQA-Net model, when trained on the LUIQD dataset, achieves the best results on the UWIQA validation set compared to the results produced by the baselines. Specifically, PAUQA-Net reaches optimal performance metrics with a SROCC value of 0.6709, KROCC value of 0.5346, PLCC value of 0.6860, and an RMSE value of 0.1111. This performance highlights the PAUQA-Net model's strong generalization capabilities and robustness, showcasing its superior performance on an unseen dataset. These attributes render PAUQA-Net highly effective for practical UIQA tasks, enhancing its usability in real-world applications.

D. Quality Prediction Performance Comparison on TID2013 and EVA

To evaluate the effectiveness of the proposed PAUQA-Net in addressing the general IQA challenge, we compared it against typical deep-learning-based IQA and UIQA methods using two in-air IQA datasets: the EVA [52] dataset and the TID2013 [51] dataset. The comparison results are presented in Table VI. Specifically, PAUQA-Net achieved SROCC, KROCC, PLCC, and RMSE values of 0.4545, 0.3128, 0.4583, and 1.1665, respectively, on the EVA dataset. On the TID2013 dataset, the values were 0.5158, 0.3737, 0.5544, and 1.7365, respectively. These values are better than, or at least comparable to, those achieved by the baseline methods designed specifically for in-air natural images. This demonstrates the ability of the proposed PAUQA-Net to effectively assess both the technical and aesthetic quality of images through an efficient convolutional attention vision transformer, combined with color and sharpness feature extraction. The results also indicate that PAUQA-Net shows greater robustness and adaptability than other UIQA networks.

E. Ablation Study

For the proposed PAUQA-Net network, we conducted ablation studies to validate the effectiveness of its key modules. For simplicity, we only report results on the LUIQD dataset.

1) Effect of the Color and Sharpness Feature Extraction Network: Firstly, we design experiments to test the effectiveness of the color and sharpness feature extraction networks. Utilizing the conv-attentional vision Transformer network as the backbone, we separately analyzed the contributions of the CFEN and the SFEN to the overall prediction performance.

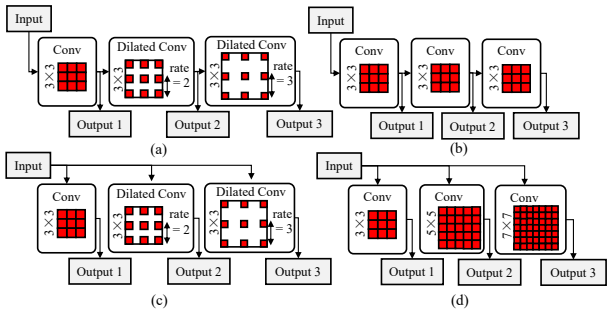

Fig. 11. Network structures of different convolutional patch embedding modules used for comparison.

The experimental results, detailed in Table VII, reveal that the combination of CFEN with the Transformer backbone achieves superior prediction performance compared to a network structure that incorporates only SFEN, likely due to the pronounced impact of color distortion in underwater images. However, when both CFEN and SFEN are integrated into the PAUQA-Net network, the performance peaks, suggesting that combining these two modules optimally addresses the complexities of underwater image assessment. This integration effectively captures both color and sharpness, crucial elements that influence the perceptual quality of underwater images, thus facilitating a more accurate and holistic approach to the assessment of underwater image quality.

2) Effect of the Multi-scale Dilated Convolutional Patch Embedding: In the proposed multi-scale joint prediction transformer, we employed a multi-scale dilated convolutional patch embedding operation to effectively represent cross-scale and cross-space quality information, as depicted in Figure 11(a). To assess the effectiveness of this configuration, we conducted experiments with variations in the structure of the patch embedding operation.

To verify the effectiveness of improving the receptive field, we tested a series arrangement of three convolution layers, each featuring a kernel size of 3, as shown in Figure 11(b). This setup aimed to explore the benefits of enlarging the receptive field without changing the kernel size or adding dilation. To compare the serial and parallel structures, we also experimented with one convolution layer and two dilated convolution layers arranged in parallel, maintaining the same kernel size of 3×3 and dilation rates of 2 and 3, as depicted in Figure 11(c). Finally, we tested the effectiveness of using large-kernel convolutions by arranging three convolution layers in parallel with kernel sizes of 3, 5, and 7, respectively, as illustrated in Figure 11(d).

The experimental results, detailed in Table VIII, demonstrated that the series convolution structure with dilated convolutions provides a nonlinear feature representation and allows for an asymptotic expansion of the receptive field without increasing computational resources. This structure was particularly effective in capturing the quality loss features of underwater images, demonstrating optimal performance.

TABLE VI
EXPERIMENT RESULTS ON THE EVA [52] AND TID2013 [51] DATASETS ARE EXPRESSED AS SROCC, KROCC, PLCC, AND RMSE VALUES. THE TOP TWO RESULTS IN EACH COLUMN ARE EMPHASIZED IN BOLD AND UNDERLINED RESPECTIVELY.

Method	EVA				TID2013			
	SROCC↑	KROCC↑	PLCC↑	RMSE↓	SROCC↑	KROCC↑	PLCC↑	RMSE↓
UNIQUE [61]	0.3097	0.2100	0.2114	1.0681	<u>0.4611</u>	<u>0.3170</u>	0.5498	1.2712
NIMA [63]	0.3720	0.2532	0.3561	0.9950	0.2832	0.1932	0.2818	1.2843
HyperIQA [62]	0.4131	0.2839	0.3900	0.9881	0.3922	0.2669	0.3823	<u>1.2139</u>
MUSIQ [17]	<u>0.4353</u>	<u>0.2990</u>	<u>0.4344</u>	**0.9448**	0.4247	0.3029	**0.5617**	**1.1331**
TQ4AQ [64]	0.4185	0.2855	<u>0.4344</u>	1.0438	0.4246	0.2947	0.4146	1.4338
URanker [7]	0.3497	0.2362	0.3484	<u>0.9705</u>	0.1668	0.1181	0.0333	1.2993
ATUIQP [18]	0.2756	0.1869	0.2720	1.0185	0.3784	0.2608	0.4388	2.0697
PAUQA-Net (Ours)	**0.4545**	**0.3128**	**0.4583**	1.1665	**0.5158**	**0.3737**	<u>0.5544</u>	1.7365

TABLE VII
ABLATION EXPERIMENT RESULTS RELATED TO THE CFEN AND SFEN IN THE PROPOSED PAUQA-NET. THE BEST RESULTS IN EACH COLUMN ARE EMPHASIZED IN BOLD.

Network	SROCC↑	KROCC↑	PLCC↑	RMSE↓
Backbone	0.8545	0.6734	0.8375	6.8488
CFEN+Backbone	0.8649	0.6868	0.8577	6.1947
SFEN+Backbone	0.8606	0.6819	0.8465	6.3831
PAUQA-Net	**0.8745**	**0.6991**	**0.8635**	**6.0703**

TABLE VIII
ABLATION EXPERIMENT RESULTS RELATED TO THE DCPE AND OTHER CONVOLUTION-BASED PATCH EMBEDDING METHODS IN THE PROPOSED PAUQA-NET. THE BEST RESULTS ARE EMPHASIZED IN BOLD.

Patch Embedding Method	SROCC↑	KROCC↑	PLCC↑	RMSE↓
Multi-Scale DCPE	**0.8745**	**0.6991**	**0.8635**	**6.0703**
p=[3,3,3], series	0.8732	0.6969	0.8580	6.4771
p=[3,3,3], r=[1,2,3], parallel	0.8702	0.6941	0.8613	6.1425
p=[3,5,7], parallel	0.8570	0.6783	0.8499	6.6690

3) Effect of the Convolutional Block Attention Module: Inspired by the human perceptual system and recognizing the significant influence of color on underwater image quality, we incorporated the CBAM to extract color features. To assess the impact of the CBAM, we conducted an ablation study by removing the module from CFEN and evaluating the network's performance. The experimental results, presented in Table IX, show that the inclusion of CBAM enhances the network's accuracy in assessing underwater image quality. This suggests that the attention module in chrominance space is able to capture color-related features that are crucial for accurate image quality evaluation, aligning closely with human perception.

F. Discussion

In order to better evaluate the quality of raw and enhanced underwater images, we organized the LUIQD dataset and trained the PAUQA network to assess image quality in terms of color features, sharpness features, and multi-scale image features. The LUIQD dataset assigns a global, absolute quality score to each underwater image, directly measuring the visual quality of the image. This enables images from different scenes and enhanced by different methods to be directly compared. It can also be used to determine the extent to which a particular enhancement algorithm improves quality across all scenes. This quality score, which does not rely on reference images, avoids the scale drift problem caused by comparative experiments, making it possible for a UIQA method trained on it to become a unified and differentiable optimization target for UIE algorithms. However, more experiments are needed to demonstrate the differences between the proposed global quality scores and ranking-based scores [7]. In future research, we will also expand the number of observers to improve the effectiveness of the dataset.

TABLE IX
ABLATION EXPERIMENT RESULTS RELATED TO THE CBAM IN THE PROPOSED PAUQA-NET. THE BEST RESULTS ARE EMPHASIZED IN BOLD.

Network	SROCC↑	KROCC↑	PLCC↑	RMSE↓
w/o CBAM	0.8726	0.6969	0.8587	6.1357
w CBAM (PAUQA-Net)	**0.8745**	**0.6991**	**0.8635**	**6.0703**

We evaluated the proposed PAUQA-Net on various underwater and natural image quality assessment datasets. The experimental results show that PAUQA-Net outperforms other algorithms on the UIQA task, and outperforms or is at least comparable to other IQA methods on technical quality and aesthetic quality assessment tasks for in-air images. This fully demonstrates the robustness of the proposed network. However, considering that UIE methods may introduce unnatural aberrations and color deviations that reduce aesthetic quality, or excessively increase contrast and cause blurring of local areas, thereby affecting the technical quality of images, there are still certain limitations when evaluating quality based only on depth features. Further research is needed to develop content-independent networks that can separate aesthetic features and technical quality features in images [66], in order to better assess the quality of enhanced underwater images.

VII. CONCLUSION

To accurately evaluate the quality of underwater images and gauge the effectiveness of various UIE algorithms, a robust UIQA algorithm is essential. In this paper, we introduce PAUQA-Net, a novel perception-aware UIQA network that uniquely integrates multi-path transformers to encode multi-scale image features within the UIQA framework, enabling precise assessment of underwater image quality. To enhance the extraction of quality-relevant features, we developed a multi-scale Dilated Convolutional Patch Embedding (DCPE) module that progressively enlarges the receptive field by extracting multi-scale image tokens. Given the unique characteristics of underwater images, we employ convolutional

neural networks to extract features from the chrominance and luminance domains, forming a comprehensive network based on joint feature extraction and interaction.

To further enhance the proposed PAUQA-Net, we compiled a comprehensive underwater image quality evaluation dataset, named LUIQD. This dataset includes a total of 64,180 raw and enhanced underwater images, covering a diverse array of underwater imaging conditions, targets, and scenes. A subjective user study was designed to annotate the LUIQD dataset and validate the effectiveness of the perceptual quality scores. Utilizing both the LUIQD and other IQA datasets, the experimental results demonstrate PAUQA-Net's superiority over other state-of-the-art IQA and UIQA models in accurately predicting image quality. This robust dataset supports the network's effectiveness in handling diverse underwater imaging scenarios, reinforcing its applicability and performance in real-world conditions.

REFERENCES

[1] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao, "An Underwater Image Enhancement Benchmark Dataset and Beyond," *IEEE Transactions on Image Processing*, vol. 29, pp. 4376–4389, 2020.

[2] R. Liu, X. Fan, M. Zhu, M. Hou, and Z. Luo, "Real-World Underwater Enhancement: Challenges, Benchmarks, and Solutions Under Natural Light," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 12, pp. 4861–4875, Dec. 2020.

[3] L. Cao, L. Shen, M. Yu, Z. Wang, and C. Shen, "Prior-Guided Dual-Reference Contrastive Learning for Underwater Object Detection," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–1, 2025.

[4] Z. Cheng, G. Fan, J. Zhou, M. Gan, and C. L. P. Chen, "FDCE-Net: Underwater Image Enhancement With Embedding Frequency and Dual Color Encoder," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 35, no. 2, pp. 1728–1744, Feb. 2025.

[5] H. Wang, S. Sun, and P. Ren, "Underwater Color Disparities: Cues for Enhancing Underwater Images Toward Natural Color Consistencies," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 34, no. 2, pp. 738–753, Feb. 2024.

[6] C. Ancuti, C. O. Ancuti, T. Haber, and P. Bekaert, "Enhancing underwater images and videos by fusion," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2012, pp. 81–88.

[7] C. Guo, R. Wu, X. Jin, L. Han, Z. Chai, W. Zhang, and C. Li, "Underwater Ranker: Learn Which Is Better and How to Be Better," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37. AAAI Press, Nov. 2022, pp. 702–709.

[8] M. Yang and A. Sowmya, "An Underwater Color Image Quality Evaluation Metric," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 6062–6071, Dec. 2015.

[9] K. Panetta, C. Gao, and S. Agaian, "Human-Visual-System-Inspired Underwater Image Quality Measures," *IEEE Journal of Oceanic Engineering*, vol. 41, no. 3, pp. 541–551, Jul. 2016.

[10] Y. Liu, K. Gu, J. Cao, S. Wang, G. Zhai, J. Dong, and S. Kwong, "UIQI: A Comprehensive Quality Evaluation Index for Underwater Images," *IEEE Transactions on Multimedia*, pp. 1–15, 2023.

[11] J. Zhu, L. Shen, Z. Wang, and Y. Yu, "Underwater Image Quality Assessment Using Feature Disentanglement and Dynamic Content-Distortion Guidance," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–1, 2025.

[12] Y. Wang, N. Li, Z. Li, Z. Gu, H. Zheng, B. Zheng, and M. Sun, "An imaging-inspired no-reference underwater color image quality assessment metric," *Computers & Electrical Engineering*, vol. 70, pp. 904–913, Aug. 2018.

[13] 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*, vol. 33, no. 3, pp. 1123–1139, Mar. 2023.

[14] Q. Jiang, X. Yi, L. Ouyang, J. Zhou, and Z. Wang, "Toward Dimension-Enriched Underwater Image Quality Assessment," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 35, no. 2, pp. 1385–1398, Feb. 2025.

[15] A. A. Laghari, V. V Estrela, and S. Yin, "How to Collect and Interpret Medical Pictures Captured in Highly Challenging Environments that Range from Nanoscale to Hyperspectral Imaging," *Current Medical Imaging*, vol. 20, p. e281222212228, 2024.

[16] 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*, vol. 32, no. 9, pp. 5959–5974, Sep. 2022.

[17] J. Ke, Q. Wang, Y. Wang, P. Milanfar, and F. Yang, "MUSIQ: Multi-scale Image Quality Transformer," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE Computer Society, Oct. 2021, pp. 5128–5137.

[18] 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*, pp. 1–14, 2024.

[19] 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*, vol. 94, p. 116218, May 2021.

[20] 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*, vol. 31, pp. 5456–5468, 2022.

[21] G. Hou, Y. Li, H. Yang, K. Li, and Z. Pan, "UID2021: An Underwater Image Dataset for Evaluation of No-Reference Quality Assessment Metrics," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 19, no. 4, pp. 151:1–151:24, Feb. 2023.

[22] G. Hou, S. Zhang, T. Lu, Y. Li, Z. Pan, and B. Huang, "No-reference quality assessment for underwater images," *Computers and Electrical Engineering*, vol. 118, p. 109293, Aug. 2024.

[23] 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*, vol. 72, pp. 1–15, 2023.

[24] Z. Wang, L. Shen, Z. Wang, Y. Lin, and J. Chen, "Prior-Based Underwater Enhanced Image Quality Assessment Network," *IEEE Journal of Oceanic Engineering*, vol. 49, no. 2, pp. 592–605, Apr. 2024.

[25] 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*, vol. 49, no. 2, pp. 637–648, Apr. 2024.

[26] Q. Jiang, Y. Gu, Z. Wu, C. Li, H. Xiong, F. Shao, and Z. Wang, "Deep Underwater Image Quality Assessment With Explicit Degradation Awareness Embedding," *IEEE Transactions on Image Processing*, vol. 34, pp. 1297–1310, 2025.

[27] P. L. Drews, E. R. Nascimento, S. S. Botelho, and M. F. Montenegro Campos, "Underwater Depth Estimation and Image Restoration Based on Single Images," *IEEE Computer Graphics and Applications*, vol. 36, no. 2, pp. 24–35, Mar. 2016.

[28] A. M. Reza, "Realization of the Contrast Limited Adaptive Histogram Equalization (CLAHE) for Real-Time Image Enhancement," *The Journal of VLSI Signal Processing-Systems for Signal, Image, and Video Technology*, vol. 38, no. 1, pp. 35–44, Aug. 2004.

[29] C. Li, S. Anwar, and F. Porikli, "Underwater scene prior inspired deep underwater image and video enhancement," *Pattern Recognition*, vol. 98, p. 107038, Feb. 2020.

[30] 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*, vol. 30, pp. 4985–5000, 2021.

[31] C. Fabbri, M. J. Islam, and J. Sattar, "Enhancing Underwater Imagery Using Generative Adversarial Networks," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, May 2018, pp. 7159–7165.

[32] Z. Jiang, Z. Li, S. Yang, X. Fan, and R. Liu, "Target Oriented Perceptual Adversarial Fusion Network for Underwater Image Enhancement," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 10, pp. 6584–6598, Oct. 2022.

[33] B. Series, "Methodology for the subjective assessment of the quality of television pictures," *Recommendation ITU-R BT*, vol. 500, no. 13, 2012.

[34] 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*, vol. 29, pp. 4041–4056, 2020.

[35] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C. C. Jay Kuo, "Image database TID2013: Peculiarities, results and perspectives," *Signal Processing: Image Communication*, vol. 30, pp. 57–77, Jan. 2015.

[36] A. A. Laghari, V. V. Estrela, H. Li, Y. Shoulin, A. A. Khan, M. S. Anwar, A. Wahab, and K. Bouraqia, "Quality of experience assessment

in virtual/augmented reality serious games for healthcare: A systematic literature review," *Technology and Disability*, vol. 36, no. 1-2, pp. 17–28, Feb. 2024.
[37] M. E. Wolak, D. J. Fairbairn, and Y. R. Paulsen, "Guidelines for estimating repeatability," *Methods in Ecology and Evolution*, vol. 3, no. 1, pp. 129–137, 2012.
[38] N. Gisev, J. S. Bell, and T. F. Chen, "Interrater agreement and interrater reliability: Key concepts, approaches, and applications," *Research in Social and Administrative Pharmacy*, vol. 9, no. 3, pp. 330–338, 2013.
[39] C. Spearman, "The proof and measurement of association between two things," *The American journal of psychology*, vol. 100, no. 3/4, pp. 441–471, 1987.
[40] S. N. Pattanaik, J. A. Ferwerda, M. D. Fairchild, and D. P. Greenberg, "A multiscale model of adaptation and spatial vision for realistic image display," in *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '98. New York, NY, USA: Association for Computing Machinery, Jul. 1998, pp. 287–298.
[41] K. Nayar, J. Franchak, K. Adolph, and L. Kiorpes, "From local to global processing: The development of illusory contour perception," *Journal of Experimental Child Psychology*, vol. 131, pp. 38–55, Mar. 2015.
[42] Y. Lee, J. Kim, J. Willette, and S. J. Hwang, "MPViT: Multi-Path Vision Transformer for Dense Prediction," in *Proceedings - 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022*. IEEE Computer Society, 2022, pp. 7277–7286.
[43] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking Atrous Convolution for Semantic Image Segmentation," Dec. 2017.
[44] R. Avenash and P. Viswanath, "Semantic segmentation of satellite images using a modified CNN with hard-swish activation function." in *Visigrapp (4: Visapp)*, 2019, pp. 413–420.
[45] U. Saeed, K. Kumar, M. A. Khuhro, A. A. Laghari, A. A. Shaikh, and A. Rai, "DeepLeukNet—A CNN based microscopy adaptation model for acute lymphoblastic leukemia classification," *Multimedia Tools and Applications*, vol. 83, no. 7, pp. 21 019–21 043, Feb. 2024.
[46] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
[47] A. A. Laghari, Y. Sun, M. Alhussein, K. Aurangzeb, M. S. Anwar, and M. Rashid, "Deep residual-dense network based on bidirectional recurrent neural network for atrial fibrillation detection," *Scientific Reports*, vol. 13, no. 1, p. 15109, Sep. 2023.
[48] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," in *Proceedings of the European Conference on Computer Vision (ECCV)*. Springer-Verlag, 2018, pp. 3–19.
[49] S. Das, A. Adhikary, A. A. Laghari, and S. Mitra, "Eldo-care: EEG with Kinect sensor based telehealthcare for the disabled and the elderly," *Neuroscience Informatics*, vol. 3, no. 2, p. 100130, Jun. 2023.
[50] S. H. S. Basha, S. R. Dubey, V. Pulabaigari, and S. Mukherjee, "Impact of Fully Connected Layers on Performance of Convolutional Neural Networks for Image Classification," *Neurocomputing*, vol. 378, pp. 112–119, Feb. 2020.
[51] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, L. Jin, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. J. Kuo, "Color image database TID2013: Peculiarities and preliminary results," in *European Workshop on Visual Information Processing (EUVIP)*, Jun. 2013, pp. 106–111.
[52] C. Kang, G. Valenzise, and F. Dufaux, "EVA: An Explainable Visual Aesthetics Dataset," in *Joint Workshop on Aesthetic and Technical Quality Assessment of Multimedia and Media Analytics for Societal Trends (ATQAM/MAST'20), ACM Multimedia*. Seattle, United States: ACM, Oct. 2020, pp. 5–13.
[53] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-Reference Image Quality Assessment in the Spatial Domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
[54] A. K. Moorthy and A. C. Bovik, "A Two-Step Framework for Constructing Blind Image Quality Indices," *IEEE Signal Processing Letters*, vol. 17, no. 5, pp. 513–516, May 2010.
[55] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind Image Quality Assessment: A Natural Scene Statistics Approach in the DCT Domain," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.
[56] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "Completely Blind" Image Quality Analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, Mar. 2013.
[57] L. Zhang, L. Zhang, and A. C. Bovik, "A Feature-Enriched Completely Blind Image Quality Evaluator," *IEEE Transactions on Image Processing*, vol. 24, no. 8, pp. 2579–2591, Aug. 2015.
[58] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional Neural Networks for No-Reference Image Quality Assessment," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2014, pp. 1733–1740.
[59] 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*, vol. 30, no. 1, pp. 36–47, Jan. 2020.
[60] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, and W. Samek, "Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment," *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 206–219, Jan. 2018.
[61] 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*, vol. 30, pp. 3474–3486, 2021.
[62] S. Su, Q. Yan, Y. Zhu, C. Zhang, X. Ge, J. Sun, and Y. Zhang, "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*, 2020, pp. 3667–3676.
[63] H. Talebi and P. Milanfar, "NIMA: Neural Image Assessment," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 3998–4011, Aug. 2018.
[64] X. Sheng, L. Li, P. Chen, J. Wu, L. Xu, Y. Yang, and Y. Li, "Technical Quality-Assisted Image Aesthetics Quality Assessment," in *Pattern Recognition and Computer Vision*, Q. Liu, H. Wang, Z. Ma, W. Zheng, H. Zha, X. Chen, L. Wang, and R. Ji, Eds. Singapore: Springer Nature, 2024, pp. 50–62.
[65] T. V. Q. E. Group, "Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment, Phase II (FR_TV2)," 2003.
[66] H. Wu, E. Zhang, L. Liao, C. Chen, J. Hou, A. Wang, W. Sun, Q. Yan, and W. Lin, "Exploring Video Quality Assessment on User Generated Contents from Aesthetic and Technical Perspectives," in *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE Computer Society, Oct. 2023, pp. 20 087–20 097.

Bosen Lin received his B.E. degree in software engineering from Northwestern Polytechnical University, Xi'an, China, in 2019. and the M.E. degree in control science and technology from National University of Defense Technology, Changsha, China, in 2022. He is currently pursuing the Ph.D. degree with the School of Computer Science and Technology, Ocean University of China, Qingdao, China. His research interests include underwater image quality assessment, enhancement and dehazing.

Junyu Dong received the B.Sc. and M.Sc. degrees from the Department of Applied Mathematics, Ocean University of China, Qingdao, China, in 1993 and 1999, respectively, and the Ph.D. degree in image processing from the Department of Computer Science, Heriot-Watt University, U.K., in 2003. He joined Ocean University of China in 2004. He is currently a Professor and the Dean of the Faculty of Information Science and Engineering, Ocean University of China. His research interests include computer vision, underwater image processing, and machine learning.

Xinghui Dong received the PhD degree from Heriot-Watt University, U.K., in 2014. He worked with the Centre for Imaging Sciences, the University of Manchester, U.K., between 2015 and 2021. Then he jointed Ocean University of China in 2021. He is currently a professor at the Ocean University of China. His research interests include computer vision, defect detection, texture analysis, and visual perception.