Distortion-independent Pairwise Underwater Image Perceptual Quality Comparison

Miao Yang, *Member, IEEE,* Zhuoran Xie, Jinnai Dong, Hantao Liu, *Senior Member, IEEE,* Haiwen Wang, and Mengjiao Shen

***Abstract*—Ranking underwater images according to their quality is a key indicator comparing the performance of different methodologies and therefore is critical in the field of instrumentation and measurement. Perceiving differences in the quality of underwater images is a challenging task that has received relatively less attention, primarily due to mixed distortion, diversified image content, and the absence of high-quality reference images. Thus, based on the pairwise underwater image quality voting data, we develop a novel ternary classification transformer to predict a quality comparison of underwater images without reference. This is the first attempt to model the quality discrimination of an image pair. The proposed model combines the perception of convolutional neural networks and Transformer encoder to explore local quality features and visual perceptual connections between different patch tokens. Experimental results reveal that the proposed pairwise underwater image quality comparison (PUIQC) scheme predicts noticeable quality differences correlating well with subjective perception. The quantification of complex distortions in underwater images compared to other learning-based methods is a compelling feature of this technology. It delivers competitive results in ranking the different enhancement outputs. In addition, we reveal the self-attention of local quality features within the two images and capture their responsive contribution to the quality decision, which explains the underlying subjective quality-sensitive mechanism during image quality comparison.**

***Index Terms*—Underwater image quality evaluation, Learning to rank, Image pair, Quality comparison without reference, Transformer.**

I. Introduction

VISION, as an essential sensing technique and measuring tool, involves numerous information for understanding subsea environments [1]–[3]. Due to the unstructured and hazardous underwater environment, underwater images are widely-used for exploring, recognizing, and monitoring the underwater world, playing an irreplaceable and significant role in the underwater application system. Ranking the quality of underwater images has important instructive implications for various tasks, including image screening, underwater image enhancement, comparison of restoration results, and underwater imaging system design [4]–[10]. Although numerous algorithms have been developed to predict the quality of natural images, automatically measuring the quality of underwater images in a way that is consistent with human perception is challenging because no referenced image can be associated, and the levels of mixed distortion in underwater images can not be grouped. Underwater images are typically low-quality, so it is hard to illuminate the perceived quality discrimination with specific image quality evaluation (IQE) values. Moreover, a lack of effective quality comparison methods for different enhancement methods conforms to subjective perception.

The blind image quality assessment (BIQA) methods estimate image quality without accessing information about the reference image. As discussed in [11], the numerical quality score for each image, the mean opinion score (MOS), is highly challenging and noise-prone. Besides, due to the limited number of underwater images in the existing IQA database, i.e., TID2013 [12], LIVE database [13], CSIQ database [14] and KonIQ-10k database [15], the degradation in the real underwater images can not be predicted accurately. Although deep learning models have great potential in image quality evaluation due to their excellent performance in image classification and recognition fields [16]–[19], compared with the large training data included in the existing image recognition datasets, the variety and number of underwater images [20], [21] are far from sufficient for training an authentic deep model.

Training a model using image pairs is an alternative data augmentation process in deep learning based image quality evaluation [22]–[25]. However, most of the pairwise samples are generated by grouping a reference image and its corresponding distorted versions [11], [24], [25], or by associating a threshold value of some full-reference IQA metrics on the image pair. In both cases, the reference image is necessary [22], [23]. Indeed, modeling the perception of quality differences between images with different content or the quality uncertainty in an image pair is rarely considered by the existing pairwise IQE methods. Besides, the perception of image quality has an inevitable relationship with visual attention. However, the knowledge about the implications of underwater image quality judgment affected by the distortion and attention areas is significantly limited. It is worth noting that how the visual attention changing by the image quality appears prominently when observers view two images simultaneously.

This paper is modeling the quality comparison of underwater images by combining CNN and transformer encoder on the preference label underwater image quality database (PLUIQD) [26], entitled the pairwise underwater image quality

Miao Yang is with the School of Electronic Engineering, Jiangsu Ocean University, Lianyungang 222005, China, and also with the Jiangsu Institute of Marine Resources Development, Lianyungang 222005, China, e-mail: lemonmiao@gmail.com

Zhuoran Xie, Jinnai Dong, Haiwen Wang, and Mengjiao Shen are with the School of Electronic Engineering, Jiangsu Ocean University, Lianyungang 222005, China

Hantao Liu is with the School of Computer Science and Informatics, Cardiff University, Cardiff, CF24 3AA, UK

comparison (PUIQC) model. This paper's contributions are as follows. (i) Unlike previous quality comparison models that depend on the referred image, the developed scheme is reference image-independent. (ii) Conducting a critical step towards treating the quality ranking of underwater images as an accumulation of the ternary classification results, where the image pairs are not necessarily assigned a visually distinguishable label. (iii) The model endeavors to capture the cross-local perceptual quality gap between two images by combining a pair of CNN and Transformer encoder. Additionally, it examines the role of attention mechanisms when analyzing two underwater images with varying content; (iv) We empirically demonstrate that segmenting images into patches and assigning a uniform quality score cannot represent the quality difference perceived from image pairs and deteriorates the model's performance. Experiments on the underwater image quality databases demonstrate that the suggested PUIQC model can accurately reveal the noticeable and unnoticeable quality differences between underwater imagery. Indeed, the accumulated preferred score provides an indicator to rank the performance of different enhancement methods. Moreover, when distinguishing the quality difference between underwater images, the image attention areas are analyzed to provide insights into optimally exploiting visual attention in underwater image quality research, which is largely unexplored.

The remainder of the paper is organized as follows. Section II reviews and summarizes the current underwater image quality evaluation, the deep learning-based BIQA, and the pairwise ranking methods. Section III introduces the framework of the proposed PUIQC network and introduces the underwater preference database explored. Sections IV and V conduct the experiments, comparative analysis against state-of-the-art models, and discuss the results. Finally, Section VI concludes this work.

II. Related work

This section briefly reviews the related BIQA algorithms, underwater image quality assessment (UIQA) methods, and deep learning-based ranking methods applied in image quality evaluation.

A. No-reference image quality assessment

Since no pristine image can be obtained in a water environment, this section discusses the BIQA metrics. Existing BIQA methods hypothesize that distortions in natural images can be recognized as at least one distortion simulated in the existing IQE datasets, such as the TID2013, LIVE, and KADID-10k [27] databases. Such BIQA methods are grouped into natural scene statistics (NSS) related methods [28], i.e., blind/referenceless image spatial quality evaluator (BRISQUE) [29], the blind image integrity notator using DCT statistics (BLIINDS [30], BLIINDS-II [31]), the integrated local NIQE (IL-NIQE) [32], and the unsupervised blind image quality evaluation via statistical measurements of structure, naturalness, and perception (SNP-NIQE) [33], and learning based methods. Without NSS knowledge, Peng *et al.* proposed a codebook representation for no-reference image assessment (CORNIA) [34] and revealed that features could be learned directly from the original image [35]. Although knowledge-driven image quality evaluation achieves appealing results for evaluating the labeled natural images, optimizing these methods when the data changes is challenging. However, deep learning schemes offer a potentially powerful framework for data-driven models [36] and have been rapidly developed to construct end-to-end BIQA solutions.

One of the challenges in applying the CNN to BIQA is the lack of sufficient training data. To overcome that, several solutions have been proposed for deep neural network (DNN) based BIQA methods, such as rotating, cropping, and mapping the images [37]. These methods are characterized by working with image patches [38], and the reference image [16], [39], or the distortion recognition is included [17], [40]. Although these methods achieve comparable performance on the existing IQE databases, relying on the people's opinion-based score to determine the whole image quality and utilizing the quality labels of the image patches is illogical. Besides, the distortion discrimination procedure reduces the models' generalization to measure the image quality with unknown distortions. Moreover, the quality comparison of two images is a complex psychological and visual physiological interaction, which cannot be represented by exclusively comparing the image blocks.

It is proved that the self-attentional mechanism can learn global and local features and express adaptive kernel weights and dynamic receptive fields similar to the deformable convolution. Image quality assessment is also a task closely related to the mutual characteristics of long-distance spatial image blocks. Besides, the great success in natural language processing extends the application of Transformer in vision-based quality evaluation, which has become a new research direction. For instance, You *et al.* [41] proposed applying the Transformer in image quality (TRIQ) by inputting the features of the last layer of ResNet50 into a shallow Transformer encoder utilizing an adaptive position embedding. Zhu *et al.* [42] employed salience region as a query and combined it with a Transformer-based encoder to conduct quality prediction. Ke *et al.* [43] introduced a multi-scale image quality Transformer (MUSIQ) which divides the input images of various resolutions into blocks, utilized as the Transformer input. Alireza *et al.* [44] aggregated features from different CNN layers into a Transformer to extract global and local features of the input images. However, the existing methods focus on the attention mechanism for a single image.

B. Underwater image quality evaluation

The absorption and scattering dominated by the water's inherent optical properties (IOPs) disturb and degrade underwater imaging when light propagates through the water. The complex physical and chemical properties of seawater result in the interaction of the forward scattering (the randomly deviated light while moving from an object to the camera) and the backscattering (the fraction of the light reflected by the water toward the camera before it reaches the objects in the scene). The wavelength depended on absorption attenuation,

the various concentration of plankton, and the color-dissolved organic substances that induce the non-uniform color casting. Beyond these features, waves, swirls, and silt produce irregular blurring in underwater images. Even worse, the artificial lighting presents the light sparkle at the center of the image and aggravates the scattering caused by the suspended particles. Underwater images are therefore dominated by a mixed distortion governed by low contrast, non-uniform illumination, blurring, non-uniform color casting, and various noise sources. Fig.1 illustrates some examples of underwater imagery. Given that it is challenging to fully and systematically explain the imaging mechanism of underwater environments and the distortion types existing in an image, current BIQA methods for natural images underperform in predicting the quality of underwater images.

(a) Whitish (b) Bluish (c) Greenish (d) Yellowish

Fig. 1. Examples of underwater images.

To evaluate the performance of the enhancement or restoration algorithms on various underwater images, the objective and subjective UIQA methods are explored extensively. For example, Karen *et al.* [20] proposed an underwater image quality measure (UIQM) method, where the training dataset contains 30 randomly selected underwater images captured with different devices and under different depths. In this work, the MOS values of the tested underwater images are gathered from 10 image processing experts. Jiang *et al.* [45] developed an effective No-reference (NR) Underwater Image Quality metric (NUIQ) to evaluate the visual quality of enhanced underwater images automatically. Moreover, in our previous work, we presented an underwater color image quality evaluation (UCIQE) method for fuzzy and low-contrast underwater monitoring images [21] and a multi-topic underwater image quality assessment (MUIQE) for a specific distortion topic [46]. However, the various UIQA value gaps cannot easily be used as an indicator to compare the perceptual underwater image quality. To highlight the inefficiency of the current methods, Table I reports the results of some common BIQA and UIQA methods when applied to the images presented in Fig.1. The best quality value for each method is highlighted in bold, and a better image quality corresponds to a lower IL-NIQE/SNP-NIQE value. Since NUIQ sorts different enhancement methods based on the same underwater image, the quality of Fig.1(c) is negative. The results infer that evaluating the underwater image quality using common BIQA methods is unsatisfactory and inconsistent with the subjective assessments. For example, some BIQA metrics score a higher value on Fig.1(b) due to the high contrast caused by the nonuniform lighting and a higher score on Fig.1(c) which is a blurred image. The outputs of different enhancement methods on underwater images are perceptual similar in quality, and subjectively hard to vote. Therefore, individually predicting the image quality score cannot illustrate an accurate quality comparison for underwater images.

C. Pairwise comparison

The pairwise ranking was initially used to estimate the preference order of different systems [49]–[52], algorithms and processing parameters. Learning from rank has recently been applied in IQA methods. Indeed, Gao *et al.* [22] exploited the preference image pairs to train a BIQA model, explored the multiple kernel learning algorithm lasso (MKLGL) to measure the similarity of different images, and mapped the paired image (NSS) features to a binary preference label. By using synthetically generated distortions, for which the relative image quality is known, Liu *et al.* [24] trained a Siamese network to rank images based on their quality and proposed the RankIQA. This model was the first to demonstrate the capability of image ranking to image quality discrimination. Specifically, the authors fine-tuned a VGG-16 network on available IQE databases and output the related quality sorting. In this method, the weight coefficients are shared by the two sub-networks. Isogawa *et al.* [53] proposed an IQA method for inpainting image repair evaluation to select the best from multiple results. For this scheme, the training data are generated by adding different levels of distortions to the inpainting images, and the features of the two images are linearly fitted to the binary pairwise preference based on the RankingSVM [49].

Ma *et al.* [23] proposed an opinions-unaware BIQA model, exploiting the quality-discriminated image pairs (dipIQ) based on RankNet, where the quality of discriminable image pairs (Dips) is measured using FRIQA metrics. Additionally, two parallel streams with shared weights are applied to process two images individually, where each network produces one quality prediction. The softmax of the predicted quality difference is the final binary classification output. Prashnani *et al.* [11] suggested a pairwise-learning framework to predict the preference probability of the reference and the distorted images (PieAPP, perceptual image-error assessment through pairwise preference). Images in groups of three are fed into an error estimation module, where the probability of preferring image A over image B with respect to the reference image R is estimated by the error differences between the two distorted images from the reference. To confront the cross-dataset quality evaluation, Zhang *et al.* [48] introduced the ResNet-34 backboned unified no-reference image quality and uncertainty evaluator (UNIQUE) to fit the MOS and the standard deviation values provided in the LIVE, CSIQ and KADID-10K datasets by a combined loss. The 270,000 training pairs are forwarded to the weight-shared Siamese network. Li *et al.* [54] developed a ranked prediction involved method for video quality evaluation, where the rank error of a sequence of images in the video, rather than a pair of images, is applied as the first assessor.

D. Inspiration

Evidently, these rank methods are all natural distortion database dependent, which generates image pairs employing

TABLE I
MEASUREMENTS ON UNDERWATER IMAGE QUALITY.

Image/Metric	DIIVINE [47]	BRISQUE [29]	BLIINDS-II [31]	CORNIA [34]	IL-NIQE [32]	SNP-NIQE [33]	DipIQ [23]	UCIQE [21]	UIQM [20]	KonCept512 [15]	UNIQUE [48]	NUIQ [45]
Fig.1(a)	21.36	8.51	14.00	44.88	**24.76**	**5.87**	**−2.30**	0.54	1.61	**0.72**	**1.04**	**6.87**
Fig.1(b)	**88.58**	61.51	**39.00**	85.77	76.39	12.19	-4.52	**0.62**	1.03	0.60	0.96	0.13
Fig.1(c)	66.26	**63.03**	38.00	**89.88**	62.54	13.68	-16.83	0.36	1.24	0.35	0.99	-0.40
Fig.1(d)	25.72	15.62	21.00	53.42	38.35	6.91	-8.06	0.60	**1.73**	0.68	1.01	1.73

distorted images and the associated reference image, suggesting that the two images are of the same content. Unlike the Siamese modules, we extract features of both images using a paired CNN without a shared weight to accommodate the different quality gaps for irrelevant content. The degree of reliance on global and fine-grained information varies when comparing two images of different quality. This motivates us to apply the Transformer on this quality comparison task, utilizing patch tokens of two images as input. Besides, we model the unnoticeable quality difference in an image pair, which endows the network with the ability to simulate the imperceptible quality difference. By ranking all possible image pairs in the dataset, the accumulated preference labels (APLs) can be obtained. The APLs collecting offers another reasonable way to rank different sources of images. Additionally, observing two images with different content is a natural stimulus rather than being forced to learn where to look for visual artifacts as eye-tracking [55]. The response inversed deduction illustrates the contribution of image areas to the quality judgment, guaranteeing the reliability of the visual attention analysis.

III. METHODOLOGY

A. Preference label underwater image quality dataset

In our previous work, we designed an underwater image quality evaluation voting procedure [26] and collected a dedicated preference label underwater image quality database (PLUIQD). The PLUIQD database comprises 1000 underwater images, and we collected the parts' labels of the 1000 images (300 images) through full pairwise voting and the others by dichotomy insertion [26]. The images in PLUIQD have different content and involve a different degree of mixed low contrast, non-uniform color degradation and illumination, and blurring distortions. The underwater color images have a size of 512×512. For further details on the PLUIQD database, the reader is referred to [26].

B. Underwater image pair training dataset

From the 1000 underwater images in PLUIQD, 1000 × (1000-1)/2 = 499,500 possible image pairs can be generated, which is a considerable number compared with the existing image quality databases. We randomly select 800 images to pair for training, of which 10% is set as the validation dataset. The remaining 200 images are paired and comprise the testing dataset. We categorize the pairs into three classes and labeled them as the preference $\{+1,-1\}$, $\{-1,+1\}$ and $\{0,0\}$, respectively. Among the three labels, $\{+1,-1\}$ represents that the quality score of the above/left image is higher than the image below/right, and $\{-1,+1\}$ is the opposite label. The $\{0,0\}$ label represents a class of underwater image pairs whose relative quality is hard for viewers to distinguish. To ensure the samples' deterministic quality difference [22] and balance the samples in the class labeled with $\{0,0\}$, we augment the image pairs with $\{0,0\}$ label and restrict the image pairs with $\{+1,-1\}$ and $\{-1,+1\}$ labels in the dataset, using the following procedure.

For convenience, we map the APL values of all images in the dataset to the centesimal scores, noted as APL-C values. Let the original training set of $\{0,0\}$ produced by subjective voting be S_{00}, and the maximum and the mean difference of the APL-C scores between the image pairs in S_{00} be $\delta_{S_{00max}}$ and $\delta_{S_{00mean}}$, respectively. To suppress the abnormal noise in the subjective evaluation process [23], we calculate the average IL-NIQE [52] difference of the image pairs in S_{00} and denote it as $\delta_{S_{ILmean}}$. For a given underwater image pair $p_{i,j}$, suppose $\delta_{S_{ij}}$ is the difference on the APL-C score and $\delta_{IL_{ij}}$ is the difference on the IL-NIQE score. We label the corresponding image pair $l_{i,j}$ by enforcing the following constraints:

$$l_{i,j}=\begin{cases}\{+1,-1\} & \delta_{S_{ij}}\geq\delta_{S_{00max}},\delta_{IL_{ij}}\leq-\left|\delta_{S_{ILmean}}\right|\\ \{-1,+1\} & \delta_{S_{ij}}\leq-\delta_{S_{00max}},\delta_{IL_{ij}}\geq\left|\delta_{S_{ILmean}}\right|\\ \{0,0\} & \left|\delta_{S_{ij}}\right|\leq\delta_{S_{00max}},\left|\delta_{IL_{ij}}\right|\leq\left|\delta_{S_{ILmean}}\right|\end{cases} \tag{1}$$

Fig.2 illustrates the three classification distributions of $\delta_{S_{ij}}$. In total, 41,359 image pairs with $\{+1,-1\}$ labels, 42,186 image pairs with $\{-1,+1\}$ labels, and 30,289 image pairs with $\{0,0\}$ labels are generated in training dataset. Moreover, 4565 image pairs with $\{+1,-1\}$ label, 4643 image pairs with $\{-1,+1\}$ labels, and 3339 image pairs with $\{0,0\}$ labels are included in the validation dataset, and 4,888 pairs with $\{+1,-1\}$ labels, 4,930 pairs with $\{-1,+1\}$ labels, 1,578 pairs with $\{0,0\}$ labels constructed in the testing dataset. Fig.3 illustrates samples of the image pairs. The paired images for training are available at https://github.com/JOU-UIP/PUIQC.

C. Overview of the proposed method

The proposed PUIQC architecture is presented in Fig.4(a), highlighting that a CNN pair captures the basic visual elements and a Transformer encoder obtains the correlation of these basic visual elements and their effects on each other's quality perception.

The inputs with shape [b, 3, 512, 512] are input into the CNN-pair, comprising base and stem networks. The developed model exploits the pre-trained inceptionresnetv2 as the base network to extract perceptually meaningful features. Two feature maps of size [b, 192, 61, 61] are forwarded to different stem networks instead of pooling to down-sample the feature maps and obtain the difference perception of the two images.

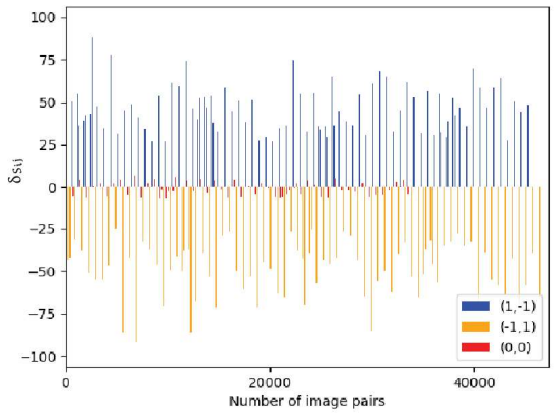
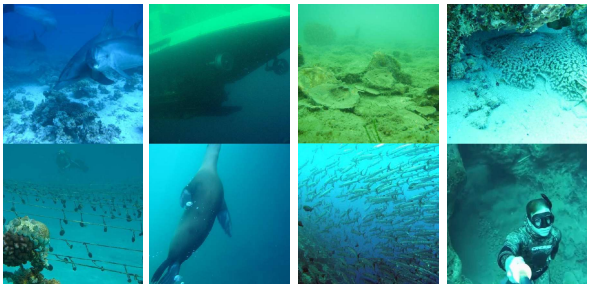


Fig. 2. The distribution of $\delta_{S_{ij}}$ in the three categorizations.

(a) The image above is noticeably better than the one below in quality.

(b) The image above is noticeably worse than the one below in quality.



(c) The image above is hard to distinguish with the one below in quality.

Fig. 3. Samples of the underwater image pairs in the three categorizations.

Subsequently, the merged feature maps are flattened into patch tokens with shapes [b, 128, 50]. Similar to ViT [56] and BERT [57], a learnable extra class token is appended before patch tokens, and then the learnable position embeddings are added. Therefore, the combination of patch tokens, class tokens, and position embedding is input into the Transformer encoder. Moreover, we consider six encoder layers, each comprising a multi-head attention and a feed-forward network, as illustrated in Fig.4(b). Layer normalization and residual connection are performed in each of sub-layers.

The output of the Transformer encoder is the final representation of the quality-aware aggregated information that is input into an MLP head. The latter consists of 4 fully connected layers and 3 dropout layers to predict the class label of the underwater image pairs.

D. Training

For a given underwater image pair $p_{i,j}$ and the corresponding preference label $l_{i,j}$, we utilize the cross-entropy loss to learn the visual perception of the quality difference constrained by the image view position to the preference label. Moreover, we adopt the L2 regularization and the Adam stochastic gradient, and the Leaky ReLU (LReLU) activation function is applied as the derivative is always non-zero, preventing the gradient disappearance. The batch size is set to 16, and we adopt a learning rate scheduler with warm-up and cosine decay. The learning rate increases to 5e-5 after three epochs, and decreases according to a cosine function in the last six epochs. The model ends training after 9 epochs. The convergence curve is illustrated in Fig.5. Finally, the PUIQC model involves 82M parameters, less than the 276M parameters of the Samiment network (two VGG 16 backbones), and is more complex than the 3-FC-layers RankNet adopted in the dipIQ [23].

E. Accumulated image quality label score

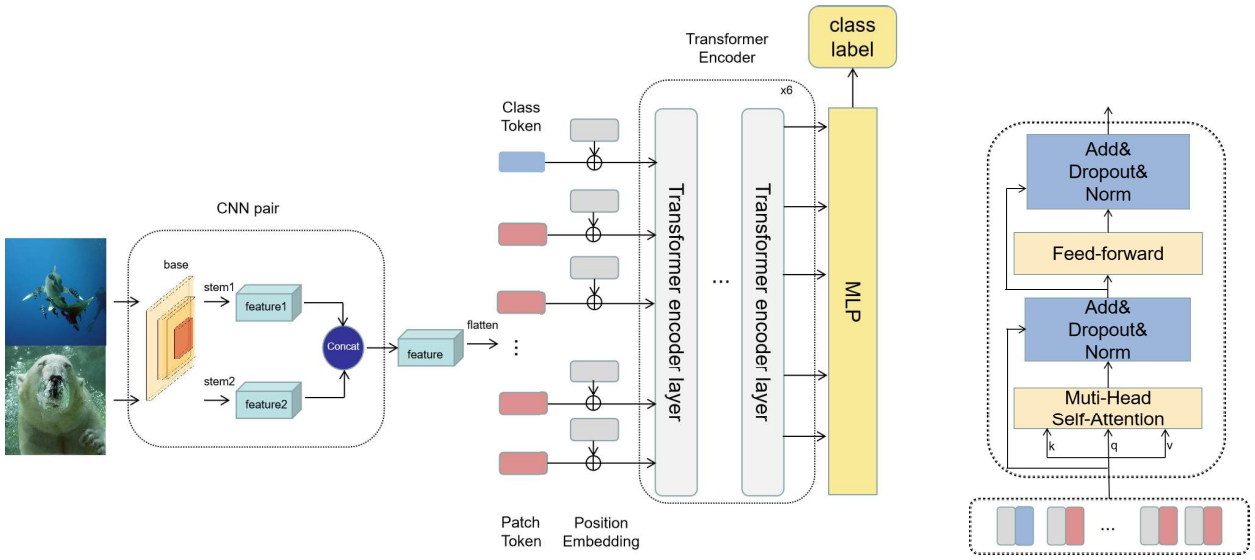
Suppose the pairwise image dataset to be predicted is P. By obtaining the predicted preference label $l_{i,j}$ for all image pairs $p_{i,j}$ in P, the APL score S for image i can be computed as:

$$S_i = \sum_j l_{i,j} \quad i \neq j. \tag{2}$$

IV. Experiments

Initially, we present the databases and the settings used to conduct the comparison experiments. Moreover, we compare the accuracy against the state-of-the-art BIQA, ranking algorithms and the UIQA metrics to verify the proposed method's performance for underwater image quality comparison. The performance of the proposed model is also validated in underwater image enhancement application. Furthermore, an ablation experiment is performed to prove the optimization network architecture of the PUIQC model.

The proposed model is challenged against the most frequently used methods in the previous BIQA studies, i.e., BRISQUE [29], BLIINDS - II [31], CORNIA [34] and SNP-NIQE [33] algorithms, and the new no-reference CNN-based image quality method KonCept512 [15], CNN-based ranking methods DipIQ [23], rankIQA [24] and UNIQUE [48]. We also compare the proposed PUIQC method with the two UIQA algorithms [20], [21]. For a fair and comprehensive comparison with the other methods, we adopt the experimental settings suggested in their original works and re-trained them on the PLUIQD with APL-C values as labels.



(a) PUIQC (b) Transformer encoder layer

Fig. 4. Architecture of PUIQC and Transformer encoder layer.

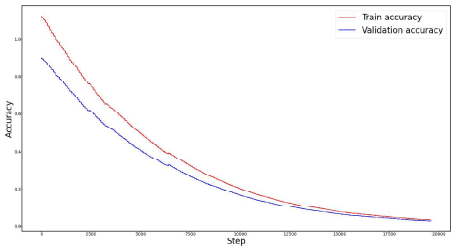


Fig. 5. The loss curve of the proposed model.

A. Accuracy experiment on image pairs

The predicted quality scores of the 200 underwater images in the testing dataset output from other methods are fit to the true APL-C values using the 5-parameter mapping function [29] before image pairing, denoted as $q_i, 1 \leq i \leq 200$. To ensure the true labels are included in the subsets constructed for other BIQA methods, the predicted labels $l_{i,j}$ for each image pair in the three testing subsets are determined by:

$$l_{ij} = \begin{cases} \{+1, -1\} & \delta q_{ij} \geq \delta_{S_{00max}} \\ \{-1, +1\} & \delta q_{ij} \leq -\delta_{S_{00max}} \\ \{0, 0\} & |\delta q_{ij}| \leq \delta_{S_{00mean}} \end{cases} \quad (3)$$

where $\delta_{q_{ij}} = q_i - q_j$ is the difference of predicted quality for a given pair $p_{i,j}$. It is worth noting that the difference gap for the discriminated image pairs has been enlarged compared to Formula (1).

Fig.6 illustrates the confusion matrix of eleven metrics for three categories of underwater image pairs, highlighting that most BIQA methods designed for natural images underperform in predicting the quality difference between underwater image pairs, although the quality gap for the discriminable image pairs has been enlarged. The accuracy obtained by the UNIQUE is higher than the other natural BIQA methods benefiting from training on the cross-data set. Furthermore, predicting significant quality differences is easier than predicting underwater images with similar qualities for the two UIQA methods (Figs.6(g) and (h)). Opposing current methods, the proposed PUIQC has better accuracy than the other BIQA and UIQA methods regardless if the image pairs have distinct quality differences or similar quality. Particularly, the proposed method distinguishes the noticeable quality difference for underwater image pairs affording better correctness. Additionally, the accuracy on the $\{0, 0\}$ category is lower than on the other two categories, possibly due to the uncertainty of the image quality difference being related to the observing environment and the psycho-physiology fluctuations of the observers' subjective perception and experience. Nevertheless, the accuracy for $\{0, 0\}$ labels reaches 84.10 %, and reaches 97.11 %, 96.91 % for $\{+1, -1\}$ and $\{-1, +1\}$ labels, respectively.

Three groups of instances are shown in Fig.7 and the comparison results produced using the competitor methods are listed in Table II. The wrong outputs are marked in red. Table II shows that the underwater images with similar quality are hard to compare, and most statistical-based metrics in the BIQA methods fail to identify the perceptual quality difference. By using the proposed PUIQC model, the underwater image pairs with different quality are classified correctly and correlated well with visual perception, based on which a reliable quality ranking can be established.

B. Testing on tank image sequences with gradual distortion

We also conduct tests on the OUC underwater database [55] to verify the ranking for images with the same distortion type but under different levels. The OUC database contains 64 experimental images, divided into four groups according to their contents. We examine $U, U = 4$ groups of underwater images, each containing 20 underwater images acquired at the same position and angle but with increased water turbidity

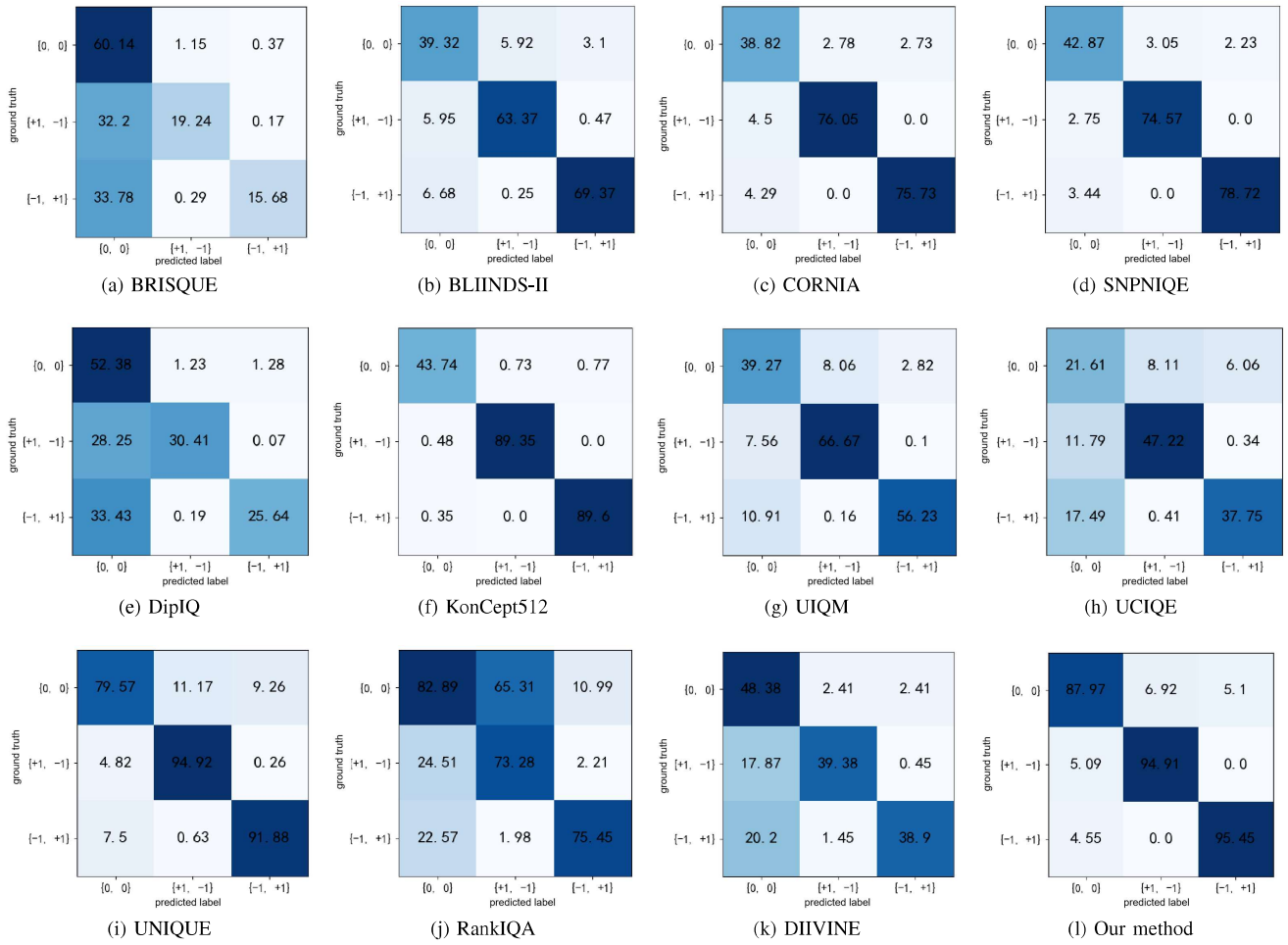
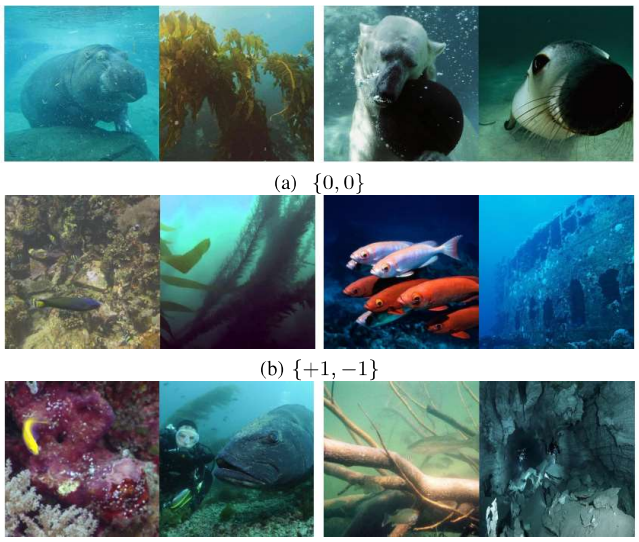


Fig. 6. The confusion matrices for the eleven methods.



(a) $\{0, 0\}$

(b) $\{+1, -1\}$

(c) $\{-1, +1\}$

Fig. 7. Image pairs with true labels.

simulated by adding milk. Therefore, 20×19/2 image pairs for each group are examined. The quality of the images in each group decreases monotonically. The list-wise ranking consistency test (L-test) [23] is applied to inspect the consistency of a BIQA method under the sequential test images differing only in their distortion levels.

$$L_u = \frac{1}{J}\sum_{i=1}^{J} SRCC(I_i, U_i) \tag{4}$$

where I_i indicates the images included in the i-th group, and U_i is the corresponding APLs computed according to the model predictions. Obviously, J in formula (4) is 4. The L-test is listed in Table III, and the data illustrates the comparable performance of the proposed PUIQC method in gradual distortion distinction. The L_u value of the proposed PUIQC is smaller than the UCIQE because the content remains unchanged, and the UCIQE is trained for fuzzy and low-contrast underwater images. Some images with equal labels are illustrated in Fig.8, revealing that the quality difference lies in the fact that these images are difficult to identify visually.

To further validate the comparative performance of the proposed method in gradual distortion distinction, we conducted tests on TankImage-I database [58]. TankImage-I is a dataset of underwater sequence images collected in a tank, where the

TABLE II
LABELS FOR THE SAMPLES OF IMAGE PAIR.

Method	Fig.7(a)(left)	Fig.7(a)(right)	Fig.7(b)(left)	Fig.7(b)(right)	Fig.7(c)(left)	Fig.7(c)(right)
BRISQUE	$\{0,0\}$	$\{+1,-1\}$	$\{+1,-1\}$	$\{-1,+1\}$	$\{0,0\}$	$\{0,0\}$
BLIINDS-II	$\{+1,-1\}$	$\{+1,-1\}$	$\{0,0\}$	$\{0,0\}$	$\{0,0\}$	$\{0,0\}$
CORNIA	$\{+1,-1\}$	$\{0,0\}$	$\{+1,-1\}$	$\{0,0\}$	$\{-1,+1\}$	$\{-1,+1\}$
SNP-NIQE	$\{+1,-1\}$	$\{+1,-1\}$	$\{+1,-1\}$	$\{0,0\}$	$\{-1,+1\}$	$\{0,0\}$
DipIQ	$\{+1,-1\}$	$\{0,0\}$	$\{-1,+1\}$	$\{+1,-1\}$	$\{-1,+1\}$	$\{0,0\}$
RankIQA	$\{+1,-1\}$	$\{+1,-1\}$	$\{0,0\}$	$\{0,0\}$	$\{-1,+1\}$	$\{-1,+1\}$
KonCept512	$\{+1,-1\}$	$\{+1,-1\}$	$\{+1,-1\}$	$\{+1,-1\}$	$\{-1,+1\}$	$\{0,0\}$
UNIQUE	$\{+1,-1\}$	$\{0,0\}$	$\{+1,-1\}$	$\{+1,-1\}$	$\{-1,+1\}$	$\{0,0\}$
UIQM	$\{-1,+1\}$	$\{+1,-1\}$	$\{0,0\}$	$\{0,0\}$	$\{0,0\}$	$\{0,0\}$
UCIQE	$\{-1,+1\}$	$\{+1,-1\}$	$\{-1,+1\}$	$\{+1,-1\}$	$\{0,0\}$	$\{0,0\}$
OURs	$\{0,0\}$	$\{0,0\}$	$\{+1,-1\}$	$\{+1,-1\}$	$\{-1,+1\}$	$\{-1,+1\}$

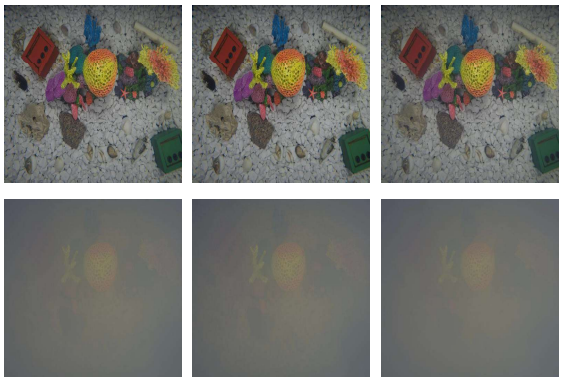


Fig. 8. Samples of indistinguishable image quality.

TABLE III
L-TEST ON TANK IMAGE SEQUENCES WITH GRADUAL DISTORTION.

Method	RankIQA	DipIQ	UNIQUE	UCIQE	UIQM	OURs
L-test	0.5243	0.9819	0.4775	**1.000**	0.6722	**0.9760**

camera distance gradually varies. The sharpness and color distortion of the images worsen as the distance from the camera increases. We selected six groups for testing, including SFR board and ColorChecker card targets with three different water transparency levels (325 cm clear, 182 cm medium turbid, and 85 cm turbid). The transparency of the water was altered by adding aluminum hydroxide. Each group was tested with $8\times(8\text{-}1)/2 = 28$ (or $7\times(7\text{-}1)/2 = 21$) image pairs, and the APLs was obtained by formula (2). The sorting results of the tank sequence images of ColorChecker card based on the PUIQC under a water transparency of 325 cm (clear) are shown in Fig.9. The results demonstrate that sorting results are consistent with the level of distortion. Due to space limitations, we provide all six group testing results in the supplementary materials.

C. Ablation experiment

In this experiment, we check whether the input size and structure of the PUIQC model are optimal. The experimental setup involved a GTX1070 GPU and an Intel i7-6700 CPU at 4.00 GHz. The corresponding experimental results are reported in Table IV.

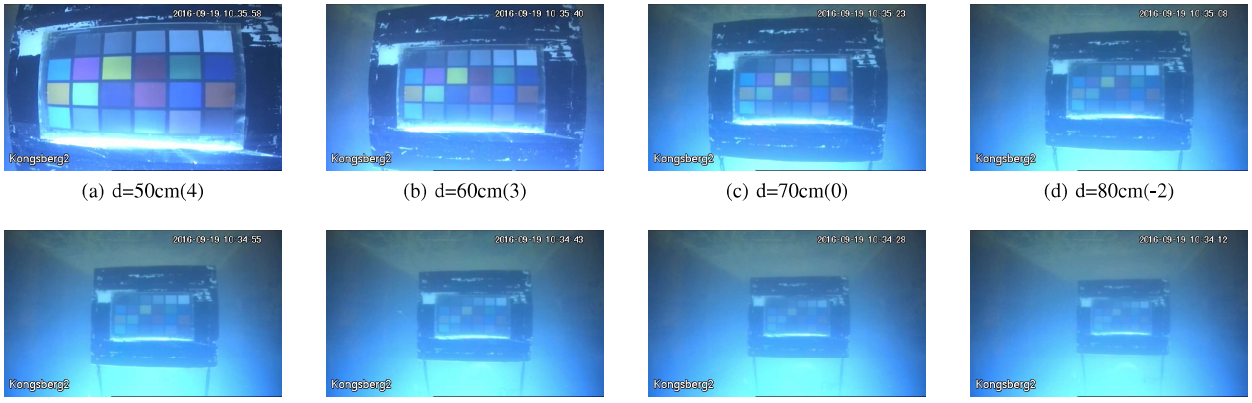
1) Size of input image: As demonstrated in Table IV, the performance of the proposed model is optimal when the input image preserves its original size, as assigning the label of the two images to each patch pair is unsuitable for comparing an image pair. When viewing two images simultaneously, the overall information from both images is involved in voting. Indeed, comparing the image blocks cannot replace the quality comparison of the two images.

TABLE IV
ACCURACY COMPARISON FOR DIFFERENT PARAMETERS AND ARCHITECTURES OF THE PROPOSED MODEL ON UNDERWATER DATABASE.

	Changed conditions	Test accuracy
Model 1	Input: 128×128	0.84
Model 2	Using image patches as input	0.57
Model 3	Using 2 encoder layers	0.91
Model 4	Using 4 encoder layers	0.93
Model 5	Pre-trained on TID2013	0.92
Model 6	Pre-trained on CSIQ	0.92
Model 7	Without pre-trained of inceptionresnetV2 on imagenet	0.90
Model 8	Merging two feature maps on dimension of channel	0.93
Model 9	Our model:Input: 512×512	**0.94**

2) Effect of the Transformer encoder layers: To demonstrate the suitability of the number of encoder layers in the Transformer encoder, we test 2 and 4 encoder layers, respectively. The results presented as Model 3 and Model 4 in Table IV illustrate that the proposed model performs the best for this quality comparison task.

3) Impact of pretraining: Many DNN-based IQA models have their backbones pre-trained with VGG-16 [24] or the FRIQA metric. Hence, we check the performance of the PUIQC model for underwater image quality comparison by pre-training it on the TID2013 or CSIQ database. As reported in Table IV, the classification accuracy of Models 5 and 6 is undesirable because the distortion type and the influencing factors in the underwater image degradation are complicated. Pre-training the model on the TID2013 or the CSIQ database forces the model to focus on a single distortion under different levels in natural images while perceiving the distortions in underwater image pairs is a different learning process. We also investigate the performance when the inceptionresnetV2 is not pre-trained on ImageNet. In this case, the final accuracy rate is 90%, highlighting that the feature extraction capabilities of



(a) d=50cm(4) (b) d=60cm(3) (c) d=70cm(0) (d) d=80cm(-2)

(e) d=90cm(-2) (f) d=100cm(-2) (g) d=110cm(-3) (h) d=120cm(-3)

Fig. 9. The sorting results of the tank sequence images taken from various distance (d) under water with transparency of 325 cm by the proposed PUIQC.

(a) HDGAN(30) (b) LiEnh(29) (c) FuEnh(28) (d) YEnh(23)

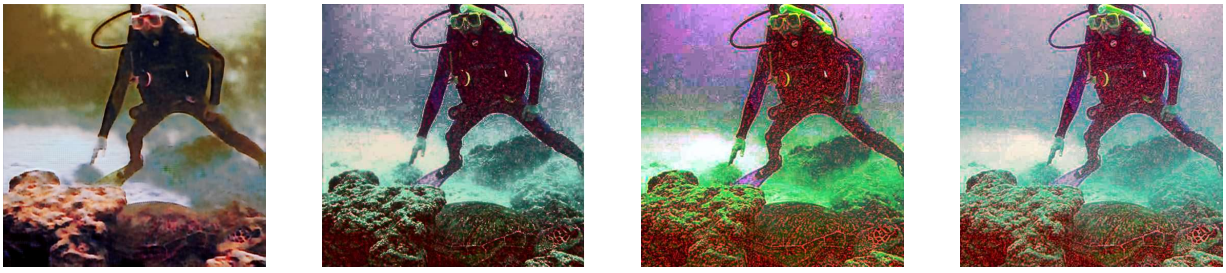
(e) GEnh(19) (f) DeepSESR(19) (g) Original(11) (h) FUNIEGAN(9)

Fig. 10. Ranking (order) of the results produced by different enhancement methods on the underwater image I.

the inceptionresnetV2 learned from the ImageNet benefit the proposed PUIQC model.

D. Ranking for underwater image enhancement results

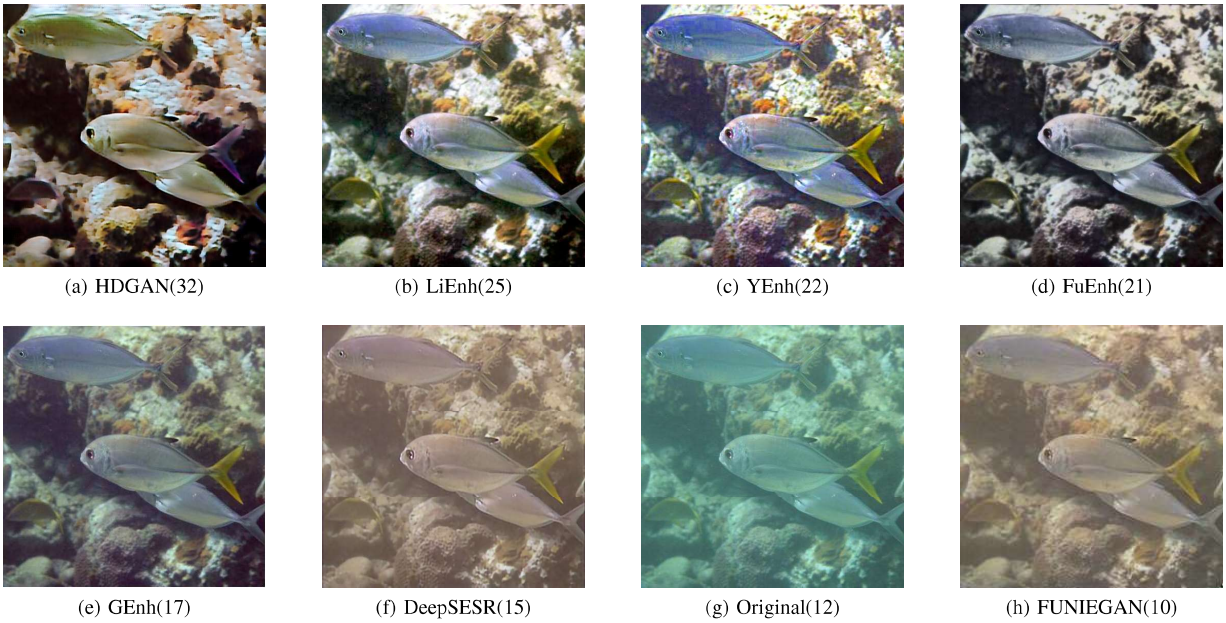
Objectively evaluating an enhancement or restoration algorithm's performance for underwater images is always challenging. Given that the UIQA metrics such as UCIQE [21] and UIQM [20] can be biased for over-enhanced images, a user study (subjective experiment) is recently applied to evaluate the restoration or enhancement results among all the competitor methods [59], [60]. However, collecting the opinion of several observers by a 10-grade or 6-grade single stimulus voting is not a rigorous procedure, and inaccurate scoring due to repeatedly viewing the same image content is likely. Competitively, the image quality ranking of the enhanced results provides an effective solution. To validate the ranking performance of the proposed method on results from different underwater enhancement methods, we process four images representing the bluish, greenish, whitish, and yellowish underwater images, respectively, using the methods proposed by Galdran *et al.* [61] (abbreviated as GEnh), Fu *et al.* [62] (FuEnh), Li *et al.* [63](LiEnh), Peng *et al.* [64] and Yang *et al.* [1] (YEnh), deep learning methods including DeepSESR [65], UWCNN [66], FUNIE-GAN [67] and HybridDetectionGAN [68] (abbreviated as HDGAN). The outputs of underwater image enhancements are same in content, and different in perceptual quality. Therefore, we concatenate the image feature maps of the CNN pairs along the channel dimension, and the shape of the merged image quality patch tokens is [b, 256, 25]. In this way, the Transformer encoder module focuses on the same positions of the two images. We characterize this model as Model 8 in Table IV. We compare all 780 (N(N-1)/2, N=10×4) image pairs using the alternative PUIQC model. The APLs of per image are obtained by



(a) HDGAN(30) (b) FuEnh(27) (c) LiEnh(25) (d) YEnh(23)

(e) DeepSESR(17) (f) GEnh(16) (g) Original(12) (h) FUNIEGAN(7)

Fig. 11. Ranking (order) of the results produced by different enhancement methods on the underwater image II.

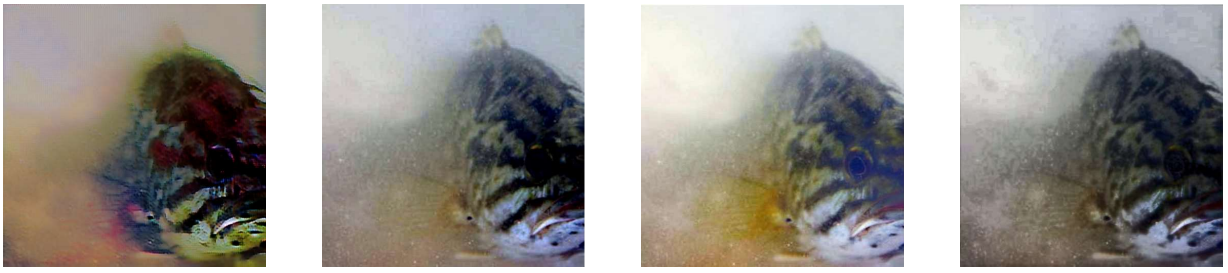


(a) HDGAN(32) (b) LiEnh(25) (c) YEnh(22) (d) FuEnh(21)

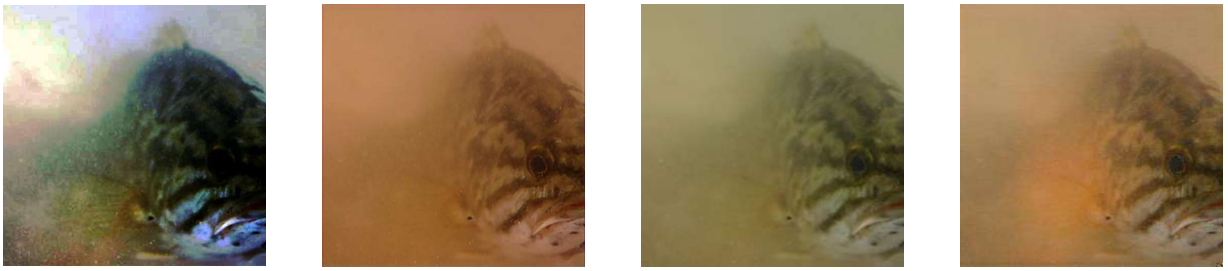
(e) GEnh(17) (f) DeepSESR(15) (g) Original(12) (h) FUNIEGAN(10)

Fig. 12. Ranking (order) of the results produced by different enhancement methods on the underwater image III.

applying Formula (2), according to which the quality ranking of the enhanced results is predicted. The maximum label score for these 40 images is +34, and the minimum label score is -35. The ranking of eight out of ten enhancement results are compared in Table V, and the ranking results for the four image groups using the proposed model 8 are presented in Figs.10 to 13. By comparing the label values obtained from each image, we obtain the quality ordering among the outputs of various methods of the same image and the quality comparisons between different images. From Table V and Figs.10 to 13, we conclude that the improved contrast and colorfulness attain a better quality rank. The pictures presented in Figs.10(a)-(d) have a better visual quality, and the image in Fig.10(h) has the worst quality due to the opposite color casting. However, the image presented in Fig.10(h) achieves the first ranking position under the UIQM metric, as listed in Table V. A similar situation occurs in Figs.12 and 13, where the result obtained by the FUNIEGAN method has a better quality due to its high UIQM score, obviously inconsistent with the fact. By counting all the labels listed in Figs.10 to 13, the enhanced whitish image in Fig.12(a) is the best, and the whole group achieves a higher average quality ranking, i.e.,



(a) HDGAN(12) (b) YEnh(7) (c) GEnh(6) (d) FuEnh(5)



(e) LiEnh(4) (f) DeepSESR(3) (g) Original(2) (h) FUNIEGAN(1)

Fig. 13. Ranking (order) of the results produced by different enhancement methods on the underwater image IV.

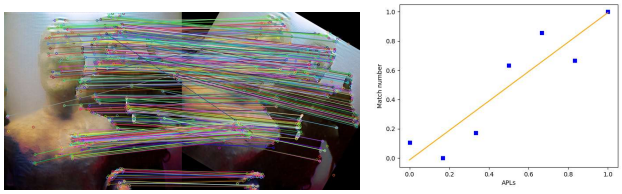
TABLE V
RANKING OF THE ENHANCEMENT RESULTS.

Quality metric	UCIQE	UIQM	PUIQC
Ranking of group1	LiEnh>YEnh>HDGAN >FuEnh >GEnh >FUNIEGAN>DeepSESR >Original	FUNIEGAN>Original >LiEnh >HDGAN >YEnh>DeepSESR>FuEnh >GEnh	HDGAN>LiEnh>FuEnh>YEnh >GEnh=DeepSESR >Original>FUNIEGAN
Ranking of group2	LiEnh >YEnh>HDGAN>FuEnh >GEnh >FUNIEGAN >Original = DeepSESR	Original=DeepSESR >LiEnh >YEnh >FUNIEGAN>FuEnh>GEnh >HDGAN	HDGAN>FuEnh>LiEnh>YEnh >DeepSESR >GEnh >Original >FUNIEGAN
Ranking of group3	LiEnh>HDGAN>YEnh >FuEnh >GEnh >FUNIEGAN >DeepSESR >Original	FUNIEGAN >FuEnh>DeepSESR>LiEnh >Original >YEnh>GEnh >HDGAN	HDGAN>LiEnh>YEnh>FuEnh >GEnh>DeepSESR>Original>FUNIEGAN
Ranking of group4	LiEnh>YEnh>FuEnh >GEnh >HDGAN>FUNIEGAN >DeepSESR>Original	FUNIEGAN >LiEnh>GEnh >Original >HDGAN >YEnh>DeepSESR>FuEnh	HDGAN >YEnh>GEnh>FuEnh >LiEnh >DeepSESR>Original>FUNIEGAN

the images taken in shallow coastal waters have the potential for quality improvement. Moreover, the UCIQE designed for a contrast and color variance measure, is more desirable than the UIQM when comparing underwater image quality. Due to the limited space, we present the results of the four groups of underwater images sorted by the UIQM and UCIQE in the supplementary materials. Outputting a judgment label based on the quality comparison between two images from a deep learning model rather than the statistical computation reduces the inaccuracy presented in the BIQA methods. Due to the undefined gap during pairing, some image pairs attain equal scores, e.g., the pictures in Figs.10(e) and (f). An exception in Fig.11 is the fake colorfulness in Figs.11 (a) and (c), receiving many votes because most underwater images win quality comparisons due to better global contrast. Generally, the sorting results are illustrated in Figs.10 to 13 highlight that the PUIQC model provides a more effective ranking for the quality comparison than other UIQA methods when comparing the enhanced underwater images. This is a good indicator of the relative quality difference between the two images.

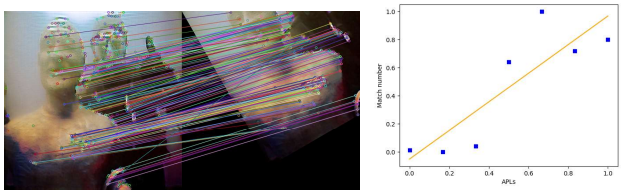
E. Applications

Underwater image enhancement techniques can improve the performance of vision-based tasks such as local keypoint matching, edge detection and saliency detection. We used the Scale Invariant Feature Transform (SIFT) operator to compare the number of effective local keypoint matches between the images in Group I and their counterpart rotated by 30 degrees, e.g., Fig.14 (a). The results indicate that as the APLs increases, the number of matched local keypoints also increases. They have a correlation coefficient of 0.85, indicating a strong correlation, as shown in Fig.14 (b). To investigate the impact of occlusion on the relationship between the number of matches and the APLs of image, we conducted experiments by rotating the images 330 degrees, which resulted in the occlusion of the diver shown in Fig.15. Based on the experimental results, we also fitted the number of matches and the APLs of image in Fig.15 (b). Despite the occlusion of the diver, the correlation coefficient remains above 0.8 at 0.82. This indicates a strong correlation between them, regardless of the presence of occlusion. The edge detection results of Group I underwater images are shown in Fig.16 , which is ordered by a cumulative score that is the same as in Fig.10. The results indicate that the higher-ranked images are detected with more edge features. Moreover, as presented in Fig.17, the saliency detection results of the underwater images of Group II are arranged in the same permutation order as Fig.11. The results demonstrate that more saliency information is detected on the images in the first row



(a) Example of local keypoint matching (b) Correlation between the number of matches and the APLs.

Fig. 14. Example of local keypoint matching.



(a) Example of occluded local keypoint matching (b) Correlation between the number of occluded matches and the APLs.

Fig. 15. Example of occluded local keypoint matching.

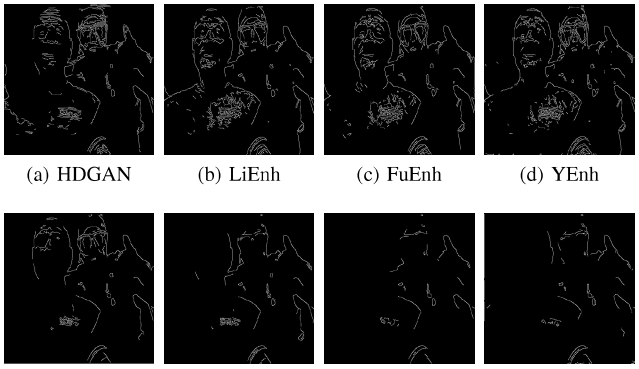
than in the second row.

There are several common scenarios in underwater environments: overlapping, occlusion, and underwater background interference. The object detection results for these scenarios are shown in Fig.18. By using the proposed PUIQC model, images with different enhancement methods are compared based on pairs, and the images are arranged from left to right in descending order of cumulative scores. From the analysis of the last two cases in Fig.18, it can be seen that because the image quality of Fig.18(a) is higher than that of Fig.18(b-d), the model can easily distinguish the occluded holothurian and the originally blurred starfish. In addition, the underwater environment scenes are complex and changeable. Some background objects are similar in shape and texture to the target object. While the image quality is improved, the characteristics of scallops are more similar to those of stones, which is easy to cause missed detection, as shown in the first case. Overall, the improvement of image quality is beneficial for human experts' annotation work and also helps improve the accuracy of object detection. Therefore, the established ranking order of the underwater images obtained by the proposed PUIQC model has a guiding significance for subsequent tasks.

V. Discussion

A. Influence of the threshold of image-pair generation on the model accuracy

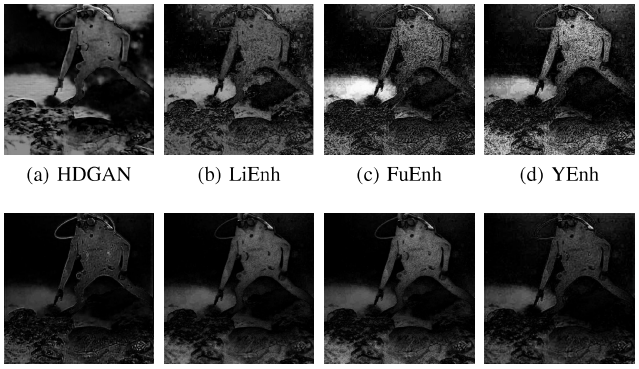
Subjective experiments show that the consistency of the quality difference judgment between the observers fluctuates when two images are too similar. Different observers may give controversial results. Gao *et al.* [22] constituted the image pairs with completely distinguishable image quality differences (when the absolute image quality difference computed



(a) HDGAN (b) LiEnh (c) FuEnh (d) YEnh

(e) GEnh (f) DeepSESR (g) Original (h) FUNIEGAN

Fig. 16. Edge detection results of the underwater images shown in Fig.10.

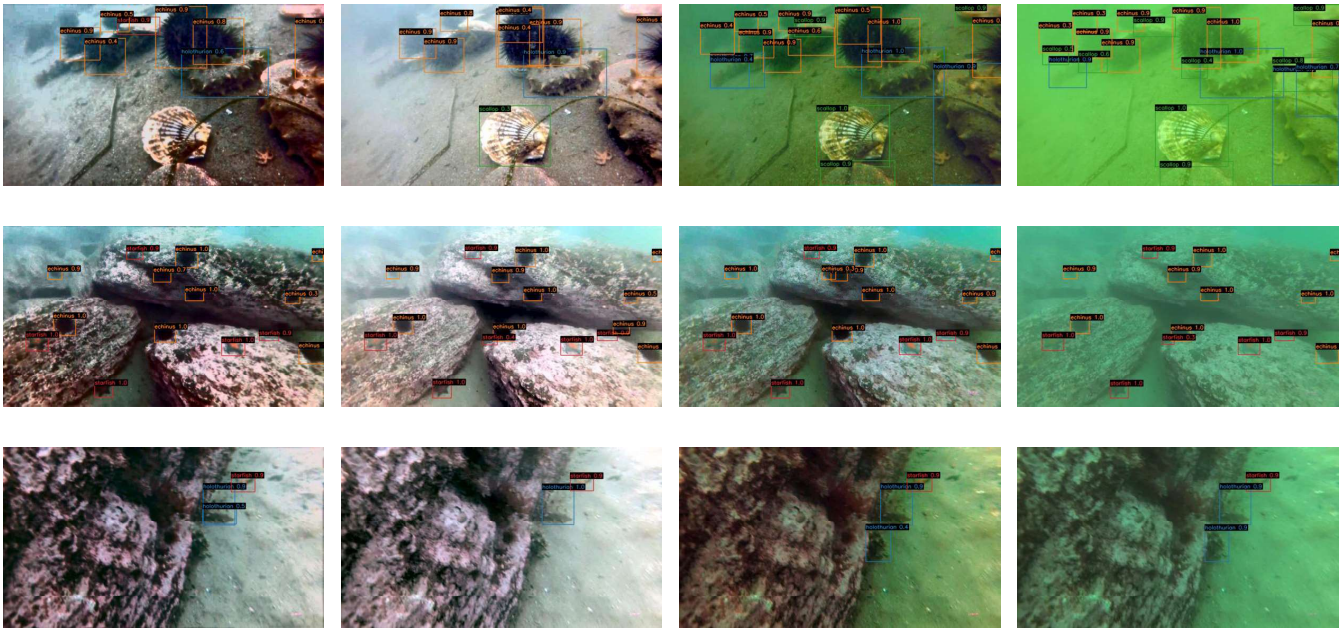


(a) HDGAN (b) LiEnh (c) FuEnh (d) YEnh

(e) GEnh (f) DeepSESR (g) Original (h) FUNIEGAN

Fig. 17. Salient detection results of the underwater image shown in Fig.11.

by the FRIQA metrics exceeds 20). In contrast, we augment the three categories by using Formula (1), and thus the APL-C differences in the $\{-1,+1\}$ and $\{+1,-1\}$ datasets are widely distributed (Fig.2), increasing the prediction difficulty. Nevertheless, the model predicts the quality difference between the image pairs with fairly good results. To illustrate how the generating principle of the image pairs influence accuracy, we investigate alternative solutions to generate image pairs. Specifically, we sort the image in the corpus by the APLs and construct image pairs with a quality order interval of 200 pieces. For instance, the first image in the sorted image sequence is paired with the image whose quality is located behind the 200-th. Additionally, the image pairs with $\delta_{S_{i,j}}$ less than $\delta_{S_{00mean}}$ constitute the $\{0,0\}$ dataset. Hence, the $\{-1,+1\}$ and $\{+1,-1\}$ categories have an evident quality difference, and the image pairs in the $\{0,0\}$ dataset have a higher certainty of unnoticeable image quality. The proposed PUIQC model predicts the three types of tags with 100 % testing accuracy, indicating that our method can accurately predict the subjective judgment of images with significant quality differences. On the contrary, we shrink the thresholds for the image pairs with labels $\{+1,-1\}$and $\{-1,+1\}$ to evaluate a more dedicated judgment of the quality difference between the two images. The constraints for constructing the image pairs for the three categories are:



(a) FuEnh (b) YEnh (c) GEnh (d) Original

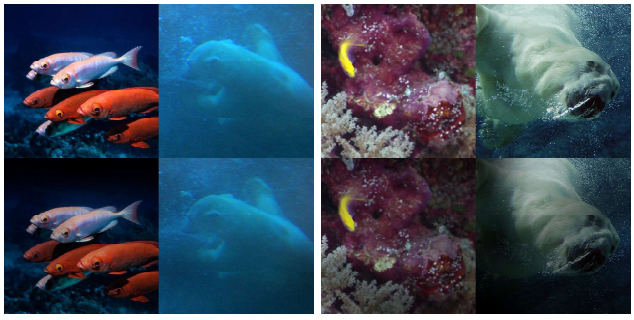
Fig. 18. Example images of object detection results by various enhancement methods.

$$l_{i,j} = \begin{cases} \{+1,-1\} & \delta_{S_{ij}} \geq \delta_{S_{00max}} \\ \{-1,+1\} & \delta_{S_{ij}} \leq -\delta_{S_{00max}} \\ \{0,0\} & \left|\delta_{S_{ij}}\right| \leq \delta_{S_{00mean}}, \\ & \left|\delta_{S_{IL_{ij}}}\right| \leq \left|\delta_{S_{ILmean}}\right| \end{cases} \quad (5)$$

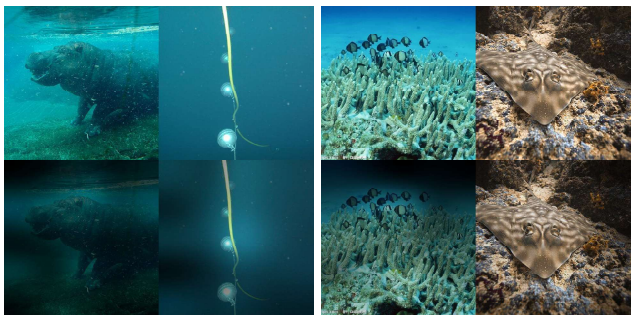
In this case, the identifiable quality difference gap between the two images is reduced considerably. Due to the inconspicuous quality difference, the classification accuracy of the model decreases to 83%, 87% and 89% for $\{0,0\}$, $\{+1,-1\}$ and $\{-1,+1\}$ three categories, respectively.

B. Quality comparison attention analysis

Since a self-attention mechanism is employed in the Transformer encoder, we visualize the attention weights used to present the contribution of each region to the quality difference judgments (Fig.19). The attention weight of the last encoder layer is used as a mask, and is applied to the input images. Brighter regions represent more important regions to the final classification, revealing that the visual attention areas are different for image pairs with different patterns of quality difference. For image pairs with obvious quality gaps, higher-quality images contribute more to the quality comparison. For example, attention in higher-quality images in Fig.19(a) distributes mainly near the object, while attention in lower-quality images distributes dispersively. However, the detail areas in both images with similar quality attract keen attention more equally. The quality comparison relies more on the edge strength and details comparison when voting the visual quality difference for the images with similar quality (Fig.19 (b)). Future work will disclose more attention mechanisms when comparing two images with different content.



(a) Two image pairs with obvious quality difference.



(b) Two image pairs with indivisible quality difference.

Fig. 19. Regional contributions to quality difference judgement for image pairs (upper: original image pairs; lower:image pairs with the masked attention distribution).

VI. Conclusion

This work presents a novel image-pair-based CNN & Transformer model to compare the quality of underwater images similar to human eye perception. The proposed PUIQC method is the first trial of learning the image quality difference

perception as a ternary classification problem. It explores a new way of ranking images from various sources and does not require a reference image. We illustrate the performance of the PUIQC framework through several experiments and the application of the enhancement algorithms. The results indicate that extracting the global information by leveraging the Transformer encoder module is a feasible visual perception mechanism for developing a learning-based BIQA method. Our findings suggest that the visual importance weights of different areas that vary with the quality interval between image pairs should be considered in UIQA methods construction. Future work will explore the full-pairwise ranking-based BIQA for the cross-datasets to realize quality comparison of images acquired in different environments.

ACKNOWLEDGMENTS

This work was supported in part by NSFC under Grant 62271236, Natural Science Foundation of Jiangsu province under Grant BK20191469, Jiangsu Natural Resources Development Special Fund (Marine Science and Technology Innovation) under Grant JSZRHYKJ202116, Graduate Research and Practice Innovation Program under Grant KYCX2021-053 and KYCX22-3395, College Students Innovation and Entrepreneurship Training Program of Jiangsu Provincial under Grant SZ202111641651001.

REFERENCES

[1] M. Yang, A. Sowmya, Z. Wei, and B. Zheng, "Offshore underwater image restoration using reflection-decomposition-based transmission map estimation," *IEEE Journal of Oceanic Engineering*, vol. 45, no. 2, pp. 521–533, 2020.

[2] M. Yang, K. Hu, Y. Du, Z. Wei, Z. Sheng, and J. Hu, "Underwater image enhancement based on conditional generative adversarial network," *Signal Processing: Image Communication*, vol. 81, p. 115723, 2020.

[3] M. Yang, J. Hu, C. Li, G. Rohde, Y. Du, and K. Hu, "An in-depth survey of underwater image enhancement and restoration," *IEEE Access*, vol. 7, pp. 123 638–123 657, 2019.

[4] F. Russo, "Automatic enhancement of noisy images using objective evaluation of image quality," *IEEE Transactions on Instrumentation and Measurement*, vol. 54, no. 4, pp. 1600–1606, 2005.

[5] P. Zhao, X. Chen, V. Chung, and H. Li, "Delfiqe—a low-complexity deep learning-based light field image quality evaluator," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.

[6] A. De Angelis, A. Moschitta, F. Russo, and P. Carbone, "A vector approach for image quality assessment and some metrological considerations," *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 1, pp. 14–25, 2009.

[7] L. Shen, B. Zhao, Z. Pan, B. Peng, S. Kwong, and J. Lei, "Channel recombination and projection network for blind image quality measurement," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–12, 2022.

[8] G. Yue, C. Hou, T. Zhou, and X. Zhang, "Effective and efficient blind quality evaluator for contrast distorted images," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 8, pp. 2733–2741, 2019.

[9] Y. Liu, X. Yin, Y. Wang, Z. Yin, and Z. Zheng, "Hvs-based perception-driven no-reference omnidirectional image quality assessment," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–11, 2023.

[10] X. Yang, F. Li, L. Li, K. Gu, and H. Liu, "Study of natural scene categories in measurement of perceived image quality," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–12, 2022.

[11] E. Prashnani, H. Cai, Y. Mostofi, and P. Sen, "Pieapp: Perceptual image-error assessment through pairwise preference," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1808–1817.

[12] 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)*, 2013, pp. 106–111.

[13] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "Live image quality assessment database release 2," [Online]. Available: http://www.image-net.org/.

[14] L. D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging.*, vol. 19, no. 1, p. 011006, 2010.

[15] 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.

[16] J. Kim and S. Lee, "Fully deep blind image quality predictor," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 1, pp. 206–220, 2017.

[17] K. Ma, W. Liu, K. Zhang, Z. Duanmu, Z. Wang, and W. Zuo, "End-to-end blind image quality assessment using deep neural networks," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1202–1213, 2018.

[18] G. Yue, C. Hou, T. Zhou, and X. Zhang, "Effective and efficient blind quality evaluator for contrast distorted images," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 8, pp. 2733–2741, 2019.

[19] Y. Liu, H. Yu, B. Huang, G. Yue, and B. Song, "Blind omnidirectional image quality assessment based on structure and natural features," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.

[20] 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, 2016.

[21] M. Yang and A. Sowmya, "An underwater color image quality evaluation metric," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 6062–6071, 2015.

[22] F. Gao, D. Tao, X. Gao, and X. Li, "Learning to rank for blind image quality assessment," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 10, pp. 2275–2290, 2015.

[23] K. Ma, W. Liu, T. Liu, Z. Wang, and D. Tao, "dipiq: Blind image quality assessment by learning-to-rank discriminable image pairs," *IEEE Transactions on Image Processing*, vol. 26, no. 8, pp. 3951–3964, 2017.

[24] X. Liu, J. Van De Weijer, and A. D. Bagdanov, "Rankiqa: Learning from rankings for no-reference image quality assessment," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 1040–1049.

[25] L. Ma, L. Xu, Y. Zhang, Y. Yan, and K. N. Ngan, "No-reference retargeted image quality assessment based on pairwise rank learning," *IEEE Transactions on Multimedia*, vol. 18, no. 11, pp. 2228–2237, 2016.

[26] M. Yang, G. Yin, Y. Du, and Z. Wei, "Pair comparison based progressive subjective quality ranking for underwater images," *Signal Processing: Image Communication*, vol. 99, p. 116444, 2021.

[27] H. Lin, V. Hosu, and D. Saupe, "Kadid-10k: A large-scale artificially distorted iqa database," *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1–3, 2019.

[28] Q. Wu, H. Li, F. Meng, K. N. Ngan, B. Luo, C. Huang, and B. Zeng, "Blind image quality assessment based on multichannel feature fusion and label transfer," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 3, pp. 425–440, 2016.

[29] 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, 2012.

[30] M. A. Saad, A. C. Bovik, and C. Charrier, "A dct statistics-based blind image quality index," *IEEE Signal Processing Letters*, vol. 17, no. 6, pp. 583–586, 2010.

[31] ——, "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, 2012.

[32] 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, 2015.

[33] Y. Liu, K. Gu, Y. Zhang, X. Li, G. Zhai, D. Zhao, and W. Gao, "Unsupervised blind image quality evaluation via statistical measurements of structure, naturalness, and perception," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 4, pp. 929–943, 2020.

[34] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 1098–1105.

[35] Q. Jiang, F. Shao, W. Lin, K. Gu, G. Jiang, and H. Sun, "Optimizing multistage discriminative dictionaries for blind image quality assessment," *IEEE Transactions on Multimedia*, vol. 20, no. 8, pp. 2035–2048, 2018.
[36] Z.-J. Zha, D. Liu, H. Zhang, Y. Zhang, and F. Wu, "Context-aware visual policy network for fine-grained image captioning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2019.
[37] Y. Li, L.-M. Po, L. Feng, and F. Yuan, "No-reference image quality assessment with deep convolutional neural networks," in *2016 IEEE International Conference on Digital Signal Processing (DSP)*, 2016, pp. 685–689.
[38] B. Yan, B. Bare, and W. Tan, "Naturalness-aware deep no-reference image quality assessment," *IEEE Transactions on Multimedia*, vol. 21, no. 10, pp. 2603–2615, 2019.
[39] J. Kim and S. Lee, "Deep learning of human visual sensitivity in image quality assessment framework," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1969–1977.
[40] 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, 2020.
[41] J. You and J. Korhonen, "Transformer for image quality assessment," pp. 1389–1393, 2021.
[42] M. Zhu, G. Hou, X. Chen, J. Xie, H. Lu, and J. Che, "Saliency-guided transformer network combined with local embedding for no-reference image quality assessment," pp. 1953–1962, 2021.
[43] J. Ke, Q. Wang, Y. Wang, P. Milanfar, and F. Yang, "Musiq: Multi-scale image quality transformer," pp. 5148–5157, 2021.
[44] S. A. Golestaneh, S. Dadsetan, and K. M. Kitani, "No-reference image quality assessment via transformers, relative ranking, and self-consistency," pp. 1220–1230, 2022.
[45] 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*, 2022.
[46] M. Yang, G. Yin, Y. Du, and H. Wang, "Multitopic underwater image quality assessment with visual attention factors," *Journal of Electronic Imaging*, vol. 31, no. 2, p. 023020, 2022.
[47] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3350–3364, 2011.
[48] 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.
[49] T. Joachims, "Optimizing search engines using clickthrough data," in *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2002, pp. 133–142.
[50] C. Burges, T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton, and G. Hullender, "Learning to rank using gradient descent," in *Proceedings of the 22nd international conference on Machine learning*, 2005, pp. 89–96.
[51] M.-F. Tsai, T.-Y. Liu, T. Qin, H.-H. Chen, and W.-Y. Ma, "Frank: A ranking method with fidelity loss," in *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, 2007, pp. 383–390.
[52] Y. Freund, R. Iyer, R. E. Schapire, and Y. Singer, "An efficient boosting algorithm for combining preferences," *Journal of machine learning research*, vol. 4, no. Nov, pp. 933–969, 2003.
[53] M. Isogawa, D. Mikami, K. Takahashi, and H. Kimata, "Image quality assessment for inpainted images via learning to rank," *Multimedia Tools and Applications*, vol. 78, no. 2, pp. 1399–1418, 2019.
[54] D. Li, T. Jiang, and M. Jiang, "Unified quality assessment of in-the-wild videos with mixed datasets training," *International Journal of Computer Vision*, vol. 129, pp. 1238–1257, 2021.
[55] M. Jian, Q. Qi, J. Dong, Y. Yin, W. Zhang, and K.-M. Lam, "The ouc-vision large-scale underwater image database," in *2017 IEEE International Conference on Multimedia and Expo (ICME)*, 2017, pp. 1297–1302.
[56] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.
[57] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.
[58] M. Yang, G. Yin, H. Wang, J. Dong, Z. Xie, and B. Zheng, "A underwater sequence image dataset for sharpness and color analysis," *Sensors*, vol. 22, no. 9, p. 3550, 2022.
[59] H. Li and P. Zhuang, "Dewaternet: A fusion adversarial real underwater image enhancement network," *Signal Processing: Image Communication*, vol. 95, p. 116248, 2021.
[60] C. Li, S. Tang, and H. Wu, "Simple estimation of red channel's transmittance and balanced color correction for underwater image enhancement," in *2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, 2020, pp. 1132–1136.
[61] A. Galdran, D. Pardo, A. Picón, and A. Alvarez-Gila, "Automatic red-channel underwater image restoration," *Journal of Visual Communication and Image Representation*, vol. 26, pp. 132–145, 2015.
[62] X. Fu, P. Zhuang, Y. Huang, Y. Liao, X.-P. Zhang, and X. Ding, "A retinex-based enhancing approach for single underwater image," in *2014 IEEE International Conference on Image Processing (ICIP)*, 2014, pp. 4572–4576.
[63] C. Li, J. Guo, C. Guo, R. Cong, and J. Gong, "A hybrid method for underwater image correction," *Pattern Recognition Letters*, vol. 94, pp. 62–67, 2017.
[64] Y.-T. Peng and P. C. Cosman, "Underwater image restoration based on image blurriness and light absorption," *IEEE Transactions on Image Processing*, vol. 26, no. 4, pp. 1579–1594, 2017.
[65] M. J. Islam, P. Luo, and J. Sattar, "Simultaneous enhancement and super-resolution of underwater imagery for improved visual perception," *arXiv preprint arXiv:2002.01155*, 2020.
[66] C. Li, S. Anwar, and F. Porikli, "Underwater scene prior inspired deep underwater image and video enhancement," *Pattern Recognition*, vol. 98, p. 107038, 2020.
[67] M. J. Islam, Y. Xia, and J. Sattar, "Fast underwater image enhancement for improved visual perception," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 3227–3234, 2020.
[68] L. Chen, Z. Jiang, L. Tong, Z. Liu, A. Zhao, Q. Zhang, J. Dong, and H. Zhou, "Perceptual underwater image enhancement with deep learning and physical priors," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 8, pp. 3078–3092, 2021.

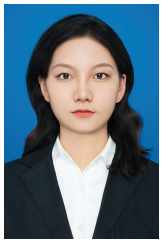
Miao.Yang received the B.S. and M.S. degrees in Electronic Engineering from Lanzhou University, Gansu Province, China in 2004 and the Ph.D. degree in Information Science and Engineering from Ocean University of China, Qingdao, in 2009. From 2010 to 2013, she was a post-doctoral fellow with Internet of Things Engineering Department, Jiangnan University, China. Since 2009, she has been a Professor with the Electronic Engineering Department, Jiangsu Ocean University. She is currently cooperating with Qingdao National Lab. of Marine Science and Technology in underwater image understanding. Her research interests include underwater vision, image processing, computer vision and 3D reconstruction.

Zhuoran. Xie is a postgraduate student, she was born in Pingdingshan City, Henan Province, China, in 1995. She received the B.S. degrees in Information and Computation Science from the North China University of Water Resources and Electric Power, Henan Province, China in 2018. Her research interests include underwater object detection and machine learning.

Jinnai. Dong is a postgraduate student, He was born in Xing Tai City, Hebei Province, China, in 1998. He received the B.E. degree in electrical engineering and the automatization specialty from Hebei University, Hebei Province, China in 2019. His research interests include underwater object detection.

Hantao. Liu received the Ph.D. degree from the Delft University of Technology, Delft, The Netherlands, in 2011. He is currently an Assistant Professor with the School of Computer Science and Informatics, Cardiff University, Cardiff, U.U.K. His research interests include visual media quality assessment, visual attention modeling and applications, visual scene understanding, and medical image perception. He is currently serving as the Chair of the Interest Group on Quality of Experience for Multimedia Communications at the IEEE MMTC and an Associate Editor of the IEEE Transactions on human-machine systems.

Haiwen. Wang is a postgraduate student, he was born in Wei Hai, Shandong Province, China, in 1996. His research interests include image processing and machine learning.



Mengjiao. Shen is a postgraduate student. She was born in Shao Xing City, Zhejiang Province, China, in 2000. She received the B.E. degree in automation from the School of Information Engineering of Hangzhou University of Electronic Science and Technology, ZheJiang Province, China in 2022. Her research interests include image quality assessment.