

Underwater Ranker: Learn Which Is Better and How to Be Better

Chunle Guo^{1*}, Ruiqi Wu^{1*}, Xin Jin¹, Linghao Han¹, Zhi Chai², Weidong Zhang³, Chongyi Li^{4†}

¹TMCC, CS, Nankai University ²Hisilicon Technologies Co. Ltd.

³School of Information Engineering, Henan Institute of Science and Technology

⁴S-Lab, Nanyang Technological University

guochunle@nankai.edu.cn, {wuruiqi, xjin, lhhan}@mail.nankai.edu.cn

chaizhi2@huawei.com, zwd_wd@163.com, chongyi.li@ntu.edu.sg

Abstract

In this paper, we present a ranking-based underwater image quality assessment (UIQA) method, abbreviated as URanker. The URanker is built on the efficient conv-attentional image Transformer. In terms of underwater images, we specially devise (1) the histogram prior that embeds the color distribution of an underwater image as histogram token to attend global degradation and (2) the dynamic cross-scale correspondence to model local degradation. The final prediction depends on the class tokens from different scales, which comprehensively considers multi-scale dependencies. With the margin ranking loss, our URanker can accurately rank the order of underwater images of the same scene enhanced by different underwater image enhancement (UIE) algorithms according to their visual quality. To achieve that, we also contribute a dataset, URankerSet, containing sufficient results enhanced by different UIE algorithms and the corresponding perceptual rankings, to train our URanker. Apart from the good performance of URanker, we found that a simple U-shape UIE network can obtain promising performance when it is coupled with our pre-trained URanker as additional supervision. In addition, we also propose a normalization tail that can significantly improve the performance of UIE networks. Extensive experiments demonstrate the state-of-the-art performance of our method. The key designs of our method are discussed. Our code and dataset is available at <https://github.com/RQ-Wu/UnderwaterRanker>.

Introduction

Underwater images commonly suffer from quality degradation issues such as color cast, low contrast, blurred details, etc. The wavelength- and distance-dependent light attenuation and scattering are the culprits of these issues (Akkaynak et al. 2017). To improve the visual quality of underwater images, underwater image enhancement (UIE) techniques have been widely studied. With the rapid development of UIE, underwater image quality assessment (UIQA) plays a more and more critical role. A good UIQA approach not only fairly evaluates the visual quality of underwater images, but also further motivates better UIE algorithms.

While existing UIQA approaches show good performance, they still have some limitations. Although full-reference IQA can objectively evaluate image quality, it is almost impossible to simultaneously obtain both underwater images and the ground truths in situ. To bypass this issue, Li et al. (Li et al. 2019) proposed a paired underwater image dataset, named UIEB. UIEB contains real-world underwater images and their reference counterparts. The reference image is obtained by manually selecting the best enhancement among the results enhanced by different UIE algorithms. The advent of UIEB allows one to use full-reference IQA such as PSNR and SSIM to evaluate the quality of underwater images. However, the provided reference images are not always ideal, leading to inaccurate assessment. There are several non-reference UIQA methods. However, they use fixed hand-crafted metrics to evaluate different underwater scenes, which is inappropriate for underwater images with diverse degradations and captured in different scenes. For instance, as shown in Figure 1(a), existing non-reference UIQA methods like UCIQE (Yang and Sowmya 2015) and UIQM (Panetta, Gao, and Agaian 2015) cannot accurately reflect the quality of underwater images.

To address these problems, we propose a ranking-based UIQA method, abbreviated as URanker. Instead of designing the hand-crafted metrics or fitting the visual scores, we attempt to rank the order of different underwater images of the same scene according to their visual quality. The benefit of using the ranker-learning technique (Liu, Van De Weijer, and Bagdanov 2017; Zhang et al. 2019) is to avoid the interference caused by different degradation and scenes, which are especially obvious in underwater images. Besides, in contrast to fitting the visual scores, the relative quality order is easier to be learned for an IQA network. Although there are some ranking-based IQA methods (Liu, Van De Weijer, and Bagdanov 2017; Ou et al. 2021), they either are originally designed for generic images or ignore the unique characteristics of underwater images.

Different from previous methods, our URanker is built on the efficient conv-attentional image Transformer (Xu et al. 2021). Taking the characteristics of underwater images into account, we specially devise two key designs. Considering that the degrees of color casts significantly influence the visual quality of underwater images, we propose a histogram prior that embeds the color distribution of an underwater im-

*These authors contributed equally.

†Corresponding author.

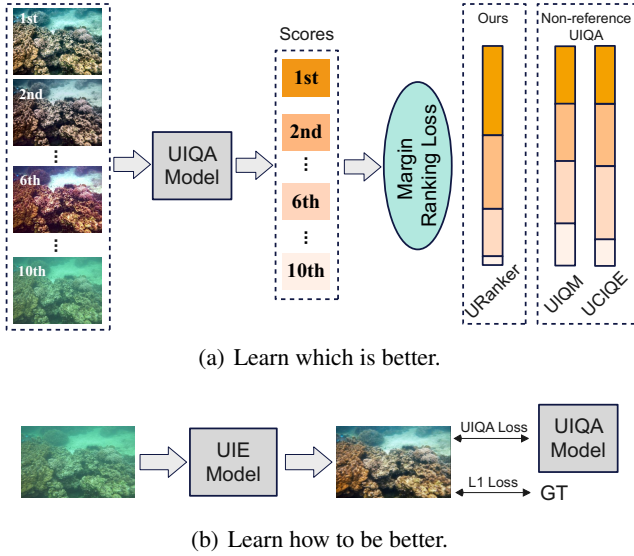


Figure 1: (a) illustrates our URanker that is trained on well-sorted image groups and optimized by margin ranking loss. The right part shows the predicted scores of the proposed URanker and two widely-used non-reference UIQA methods, UCIQE and UIQM. The ratio of the height of the rectangle represents the ratio of the quality score of the corresponding image, and higher, better. Only our URanker produces the scores in the decreasing order (align with the ground truth rankings). (b) shows the overview of a simple UIE model equipped with our pre-trained URanker as additional supervision which helps a UIE model obtain better visual quality.

age as histogram token to attend global degradation. Inspired by the multi-scale features based IQA (Ke et al. 2021), we introduce the dynamic cross-scale correspondence to model local degradation. Besides, we obtain the final prediction by considering multi-scale dependencies according to the class tokens from different scales. To achieve that, we contribute a dataset, URankerSet, containing sufficient results enhanced by different UIE algorithms and the corresponding perceptual rankings. To rank the order of different results, we adopt Bubble Sort with the help of volunteers, which effectively reflects the human visual perception. With the URankerSet, we optimize our URanker using margin ranking loss. Besides the good performance of URanker for UIQA, since all operations in our URanker are differentiable, its feedback can be further used as a loss to optimize UIE networks, as illustrated in Figure 1(b). Further, we propose a simple but effective accessory, normalization tail, for UIE networks, producing significant performance improvements.

Overall, our contributions can be summarized as follows:

1. We present a Transformer network, URanker, for UIQA with novel histogram prior and dynamic cross-scale correspondence.
2. We contribute an underwater image with ranking dataset, URankerSet, that contains underwater images enhanced

by different UIE algorithms and the corresponding visual perceptual rankings.

3. Apart from the good performance in UIQA, our URanker also improves the performance of UIE networks. The performance can be further improved using our proposed normalization tail.

Related Work

Underwater Image Quality Assessment

Due to the unique characteristics of underwater images, the commonly used IQA metrics (Mittal, Moorthy, and Bovik 2012; Talebi and Milanfar 2018; Mittal, Soundararajan, and Bovik 2012) are not suitable. Therefore, several IQA specially designed for underwater images have been proposed. Yang *et al.* (Yang and Sowmya 2015) proposed the UCIQE, which is a linear combination of image chroma, brightness, and saturation. From the perspective of the human visual system, UIQM (Panetta, Gao, and Agaian 2015) was proposed, which quantifies the colorfulness, sharpness, and contrast of the underwater image, respectively. UIQM score is calculated by the weighted sum of these three components. Recently, the Frequency Domain Underwater Metric (FDUM) (Yang et al. 2021) was proposed, which takes the frequency domain features and DCP (He, Sun, and Tang 2010) into consideration. Jiang *et al.* (Jiang et al. 2022) proposed to extract the features of chromatic and luminance and train an SVM for UIQA. These UIQA methods achieve good performance in most cases. Nevertheless, the methods based on hand-craft features (Yang and Sowmya 2015; Panetta, Gao, and Agaian 2015) and statistic regression (Yang et al. 2021; Jiang et al. 2022) do not always hold for diverse underwater images. More recently, Fu *et al.* (Fu et al. 2022) provided a Twice-Mixing framework for UIQA. It generates a set of mid-quality images by mixing up the original images and the corresponding reference version with different degrees. Then, a siamese network is adopted to learn their quality rankings. However, the human visual system is not uniform and the mixing-up manner cannot cover the over-enhanced cases. In our study, we manually sort the images enhanced by different UIE algorithms, which generates more comprehensive results compared to the mixing-up operation. Besides, we specially design a UIQA method for modeling the global and local degradation of underwater images. Moreover, our method models the multi-scale dependencies via class tokens from different scales.

Underwater Image Enhancement

As a key step for improving the visual quality of underwater images, UIE has attracted widespread attention. Existing UIE algorithms can be generally divided into three categories: non-physical model-based, physical model-based methods, and data-driven methods. Non-physical model-based methods (Iqbal et al. 2010; Ancuti et al. 2012; Fu et al. 2017) were proposed to adjust pixel values from the perspective of color balance and contrast. For physical model-based methods, one line is to modify the Dark Channel Prior (DCP) (He, Sun, and Tang 2010) to make it fit the underwater scenes (Drews et al. 2016; Li et al. 2017a; Peng, Cao, and

Cosman 2018). Another line is to solve an underwater image formation model (Galdran et al. 2015; Li et al. 2016; Peng and Cosman 2017). Recently, with the fast development of deep learning, many data-driven methods (Li et al. 2017b; Li, Anwar, and Porikli 2020; Li, Guo, and Guo 2018; Li et al. 2019, 2021) have been proposed. The promising results they achieved show their tremendous potential for UIE. In this study, we demonstrate that a simple UIE network such as U-shape network (Ronneberger, Fischer, and Brox 2015) can achieve state-of-the-art performance when it is equipped with our pre-trained URanker as additional supervision and our proposed normalization tail.

Proposed URankerSet

To train the ranking-based UIQA network, we construct an underwater image with rank dataset, called URankerSet. Following the prototype of Li *et al.* (Li et al. 2019), we use the same UIE algorithms including fusion-based (Ancuti et al. 2012), two-step-based (Fu et al. 2017), retinex-based (Fu et al. 2014), UDCP (Drewns et al. 2016), regression-based (Li et al. 2017a), GDCP (Peng, Cao, and Cosman 2018), Red Channel (Galdran et al. 2015), histogram prior (Li et al. 2016), blurriness-based (Peng and Cosman 2017), DCP (He, Sun, and Tang 2010), MSCNN (Ren et al. 2016), and dive⁺¹ to process 890 raw underwater images provided in Li *et al.*'s work. We discard three of twelve UIE algorithms used in Li *et al.*'s work as these algorithms produce similar outputs with one of the nine selected UIE algorithms. Otherwise, similar outputs may affect the accuracy of visual quality ranking.

Specifically, for each raw underwater image and the nine enhanced results, we invite eleven volunteers with image processing experience to perform pairwise comparisons using the monitors with the same model and settings. We perform Bubble Sort in descending order that is originally designed for arranging a string of numbers in the correct order to rank the visual quality of each set of underwater images (*i.e.*, a raw underwater image and nine results enhanced by nine different UIE algorithms). First, the ten underwater images are randomly arranged from left to right. Then, the volunteers are required to compare two adjacent images at a time in left to right order. In each comparison, the image of the two adjacent images which receive the minority votes is rearranged in descending order from left to right. Then, the volunteers continue to cycle through the entire sequence until it completes a pass without switching any images. In Figure 2, we show some examples of our URankerSet with visual rankings.

Proposed URanker

Recent work has demonstrated that multi-scale features and Transformer structure can boost the performance of IQA (Ke et al. 2021). Inspired by this insight, our URanker is built on the efficient conv-attentional image Transformer (Xu et al. 2021), which adopts the conv-attention mechanism to re-

place the multi-head self-attention mechanism for reducing the computational complexity.

The overview of the proposed URanker is illustrated in Figure 3(a). Firstly, the input image is fed to a series of serial blocks, in which each serial block mainly contains a conv-attentional image Transformer, as shown in Figure 3(b). In each serial block, the input features are first downsampled $2\times$ by a patch embedding operation and are flattened into a group of tokens. These tokens are concatenated with a learnable class token that represents the features from the current scale and a histogram token produced by the proposed Histogram Prior Module (HPM). Then, the conv-attention layer and the feed-forward layer alternately attend all tokens. After self-attention, the image tokens are reshaped to their original shape and used as input of the next serial block. In addition, the multi-scale outputs sequentially produced by the serial blocks are sent to the Dynamic Connection Parallel Block (DCPB). The DCPBs are responsible for integrating cross-scale features in a parallel manner. Besides, the class tokens from different scales of the DCPB are sent to the corresponding Linear layer for predicting the scores of different scales. The final prediction is the average of these scores. In what follows, the key components of the proposed URanker are introduced. More detailed structures and parameters can be found in the supplementary material.

Histogram Prior Module

Global degradation type is essential for assessing the quality of underwater images. Empirically, the color distribution of high-quality images tends to be uniform. Based on this observation, we utilize the color histograms of underwater image as a prior and design a simple yet effective module to embed this prior into the proposed URanker. Figure 4 presents the overview of the proposed HPM. Specifically, for an input underwater image, we first calculate the histogram of each channel $\mathcal{H} \in R^{B \times 3}$, where B is the number of bins and is set to 64. Then $\mathcal{H}$ is flattened into a vector of length $3B$ and fed to a Linear layer to generate the histogram token.

Dynamic Connection Parallel Block

We design the DCPB to introduce the cross-scale correspondence into the URanker. The parallel blocks in the same group communicate with each other through the proposed dynamic connection mechanism, as shown in Figure 5(d). This mechanism can adaptively integrate multi-scale features. Different from the commonly used connection manners such as direct, neighbour, and dense as illustrated in Figure 5(a)-(c), this mechanism avoids the underutilization of multi-scale features and, conversely, prevents the pollution of the original-scale features by overusing other scale features. More discussions about different connection manners are provided in our ablation studies. The details of the DCPB are described as follows.

The features of different scales are first passed through a conv-attention layer independently to keep the channel number consistent. Then, the features from other scales are down-sampled or up-sampled by bilinear interpolation to align the feature size. Given $\mathcal{F}_1$, $\mathcal{F}_2$, and $\mathcal{F}_O$ from different

¹<https://itunes.apple.com/us/app/dive-video-color-correction/id1251506403?mt=8>



Figure 2: Some visual samples of the proposed URankerSet. The correspondence ground truth ranking is marked at the top of the figure. The UIE algorithm for achieving the enhanced result is marked at the top-right corner of each image.

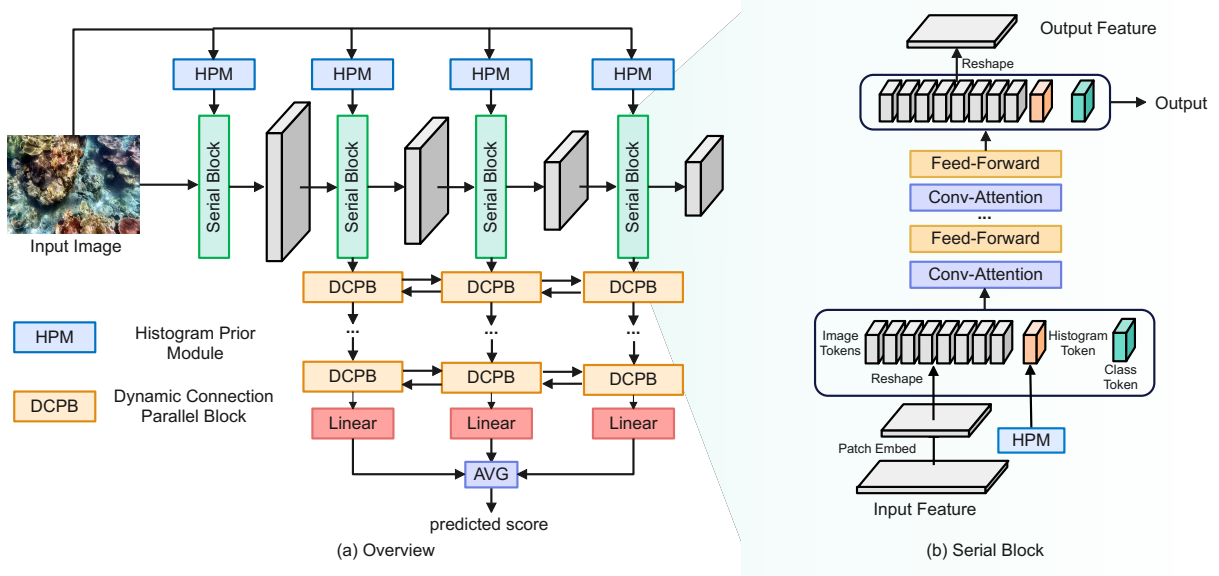


Figure 3: Overview of the proposed URanker. The multi-scale features are generated by serial blocks and the cross-scale features are communicated in a parallel fashion. In each serial block, the tokenized features are combined with the histogram token from the HPM and the class token, and then processed by the conv-attention and feed-forward layer alternately. The final prediction is the average of the scores from different scales.

scales, the dynamic connection mechanism can be written as:

$$\mathcal{F}_{cross} = \mathcal{F}_O + \alpha_1 \mathcal{F}_1 + \alpha_2 \mathcal{F}_2, \quad (1)$$

where $\mathcal{F}_{cross}$ denotes the output cross-scale features, α_1 and α_2 are two learnable parameters to adjust the amplitude. At last, the $\mathcal{F}_{cross}$ is processed by a feed-forward layer and fed to the next DCPB.

Learning to Rank

The ranker learning technique (Liu, Van De Weijer, and Bagdanov 2017; Zhang et al. 2019) is employed to optimize the URanker. It drives the URanker to learn to rank the visual quality of the enhanced results for the same input.

During the training stage, an image pair $\{I_n, I_m\}$ is selected from a set of well-sorted images $I = \{I_1, I_2, \dots, I_N\}$. The URanker predicts the correspondence scores $\{s_n, s_m\}$. Then the scores are fed to margin-ranking loss to constrain the URanker to generate scores that match

the ranking relationship. The margin-ranking loss can be formulated as:

$$\mathcal{L}(s_n, s_m) = \begin{cases} \max(0, (s_m - s_n) + \epsilon), & q_n > q_m \\ \max(0, (s_n - s_m) + \epsilon), & q_n < q_m \end{cases}, \quad (2)$$

where q_n and q_m represent the visual quality of I_n and I_m , respectively. ϵ is a parameter, which prevents the scores from getting too close. ϵ is set to 0.5 in our method.

Proposed UIE Network

The more far-reaching significance of the UIQA approach is to promote the performance of UIE algorithms. Based on this motivation, we investigate whether the proposed UIQA network, URanker, can achieve the merit when it is used as the additional supervision to train a UIE network.

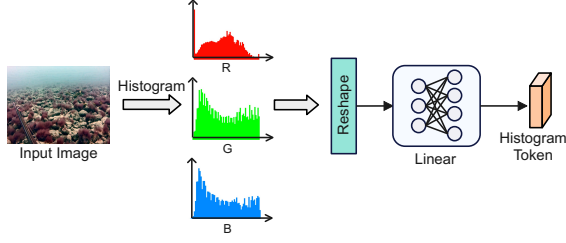


Figure 4: Illustration of our histogram prior module (HPM). The input image is first quantified into histograms, and the results are reshaped into a vector. Then, a Linear layer is followed to generate a histogram token.

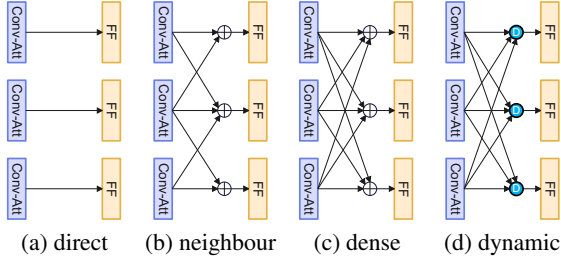


Figure 5: Different connection manners for parallel blocks. (d) illustrates the proposed dynamic connection and (a)-(c) are three variants of it.

Pre-trained URanker Loss

Since our URanker is differentiable, the pre-trained URanker can be utilized to facilitate the training of a UIE network. The pre-trained URanker loss can be expressed as:

$$\mathcal{L}_{URanker}(I_e) = \text{Sigmoid}(-M_{URanker}(I_e)), \quad (3)$$

where I_e is the result enhanced by a UIE network, $M_{URanker}(\cdot)$ is pre-trained URanker. The total loss for training a UIE network is the weighted combination of URanker loss and content loss:

$$\mathcal{L}_{total}(I_e, I_{gt}) = \mathcal{L}_{content}(I_e, I_{gt}) + \lambda \mathcal{L}_{URanker}(I_e), \quad (4)$$

where I_{gt} represents the ground truth and $\mathcal{L}_{content}$ represents the pixel-wise content loss such as L1 loss, L2 loss, or perceptual loss. λ is a trade-off coefficient.

UIE Network

Our URanker loss can be easily plugged into the training phase of any UIE models. As a baseline network, we use a simple Normalization-based U-shape UIE network (NU²Net) as shown in Figure 6. The main body of NU²Net is a stack of Conv-IN-ELU blocks. Besides the basic structure, we found a normalization tail can significantly improve the performance of UIE network.

To be specific, inspired by the stretching operation proposed in a traditional UIE algorithm (Iqbal et al. 2010), we

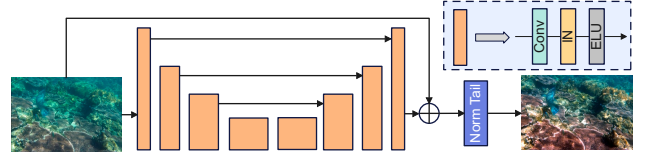


Figure 6: Overall of the proposed NU²Net. It is a residual network consisting of a stack of Conv-IN-ELU blocks. The normalization tail is followed at the end of the network.

formulate it as a normalization tail to handle the overflow values, *i.e.* outside of $[0, 1]$. When the overflow values exist in channel $c \in \{r, g, b\}$, it can be written as:

$$\hat{I}_e^c = \frac{I_e^c - \min(I_e^c)}{\max(I_e^c) - \min(I_e^c)}. \quad (5)$$

Please note that the normalization tail only works for the overflow values. More discussions are provided in the ablation studies.

Experiments

Experimental Setting

Implementations. We train our URanker for 100 epochs with the Adam optimizer with default parameters ($\beta_1 = 0.9, \beta_2 = 0.999$) and the fixed learning rate 1.0×10^{-5} . For data augmentation, the input images are randomly flipped with a probability of 0.5 in both vertical and horizontal directions. The proposed NU²Net is trained for 250 epochs with a batch size of 16. Adam optimizer with an initial learning rate of 0.001 is adopted. The learning rate is adjusted by the cosine annealing strategy. All inputs are cropped into a size of 256×256 and the same data augmentation as training URanker is employed. All experiments are implemented by PyTorch on an NVIDIA Quadro RTX 8000 GPU.

Datasets & Metrics. Following the experimental settings of (Li et al. 2021), we randomly select 800 image groups in URankerSet for training and the rest 90 groups are regarded as the testing set. Such settings are used for UIQE and UIE experiments. To measure the consistency of the IQA results with the manual sorting results, we adopt two rank-wise metrics, Spearman rank-order correlation coefficient (SRCC) and Kendall rank-order correlation coefficient (KRCC), to evaluate the performance of UIQA methods. We discarded the linear correlation metrics generally adopted by IQA methods such as Pearson linear correlation coefficient (PLCC), as our URankerSet only provides the order without the specific scores. For UIE experiments, the PSNR and SSIM are used as evaluation metrics.

Underwater Image Quality Assessment Comparison

We compare URanker with five UIQA methods, including UCIQE (Yang and Sowmya 2015), UIQM (Panetta, Gao, and Agaian 2015), FDUM (Yang et al. 2021), NUIQ (Jiang et al. 2022), Twice-Mixing (Fu et al. 2022), and one recent

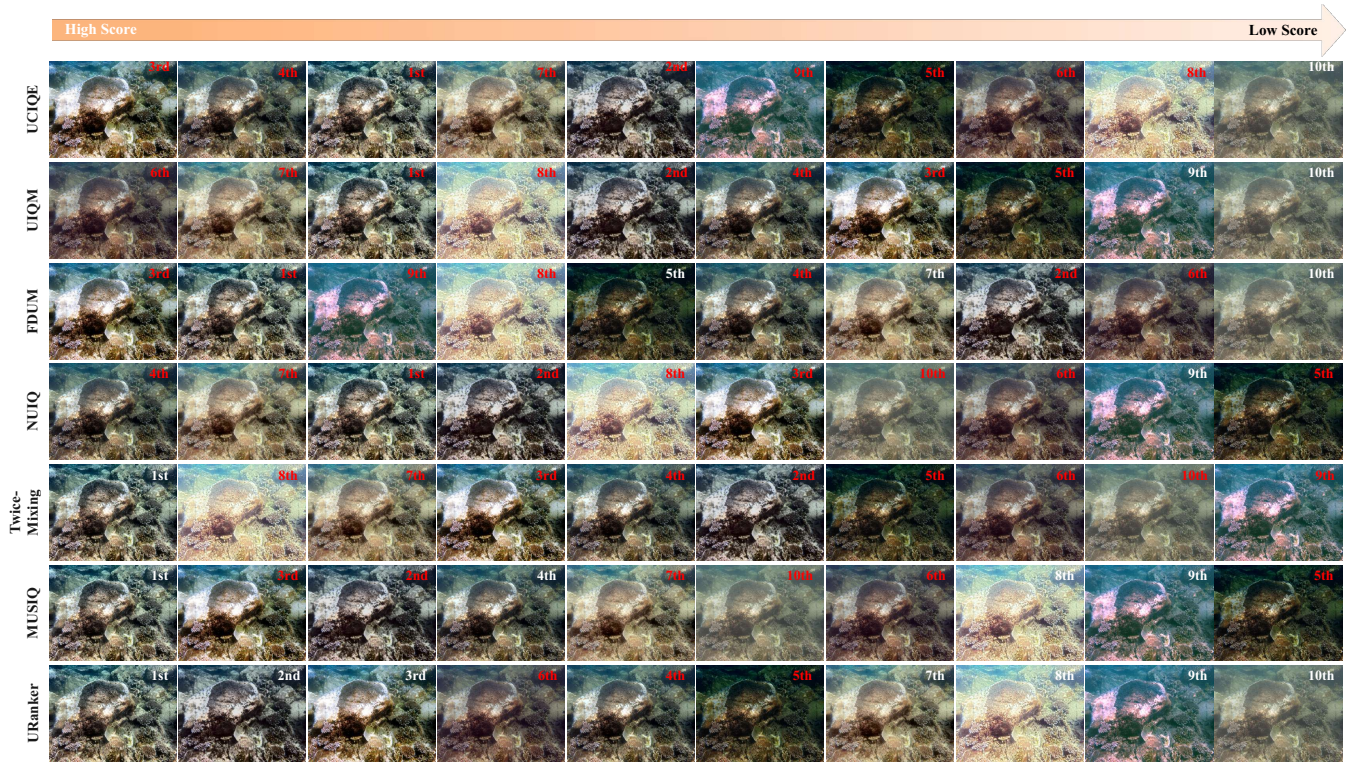


Figure 7: Visual comparison of UIQA. The prediction scores of UIQA methods decrease from left to right. The correspondence ground truth ranking is marked at the top-right corner of each image. Results with incorrect ranking are marked in red.

Methods	SRCC $\uparrow$	KRCC $\uparrow$
UCIQE	0.5039	0.4049
UIQM	0.3705	0.3052
FDUM	0.3104	0.2469
NUIQ	0.5779	0.4346
Twice-Mixing	0.6218	0.4887
MUSIQ	0.8241	0.6928
URanker (Ours)	0.8655	0.7402

Table 1: Quantitative comparison of UIQA methods. **Bold-face** indicates the best result.

Transformer-based generic IQA method, i.e., MUSIQ (Ke et al. 2021). We use the released code of UIQA methods and the PyTorch version of MUSIQ provided by IQA-PyTorch². For a fair comparison, we retrain MUSIQ using the same data as our URanker.

In Figure 7, we compare the prediction results of our URanker with other methods. We also mark the ranking order for a better visual comparison. It can be seen that our URanker achieves the closest results to the ground truth ranking (the failure case is denoted in red) and is more consistent with the human perception as the color bar indicated.

Table 1 shows the quantitative comparisons of our

²<https://github.com/chaofengc/IQA-PyTorch>

Methods	w/o $\mathcal{L}_{URanker}$		w/ $\mathcal{L}_{URanker}$	
	PSNR $\uparrow$	SSIM $\uparrow$	PSNR $\uparrow$	SSIM $\uparrow$
WaterNet	21.814	0.9194	21.926	0.9244
Ucolor	20.433	0.8963	20.686	0.8987
NU ² Net	22.419	0.9227	22.669	0.9246

Table 2: Quantitative improvements of UIE networks by applying our pre-trained URanker as additional supervision.

URanker and other methods. Our URanker achieves state-of-the-art performance. Compared with the second best method, MUSIQ, our method brings 5.02% improvements on the SRCC and 6.84% improvements on the KRCC.

Underwater Image Enhancement Comparison

To further demonstrate the effectiveness of our URanker, we adopt the pre-trained URanker as additional supervision to optimize the NU²Net and two well-known data-driven UIE networks, i.e., WaterNet (Li et al. 2019) and Ucolor (Li et al. 2021). For the content loss, WaterNet and Ucolor use the loss functions in their original works while our NU²Net adopts L1 loss. The trade-off coefficient λ in Eq. (4) is set to 0.025, 0.005, and 0.025 for training these three UIE networks, respectively.

Figure 8 presents a set of visual examples for qualitative comparison. As we can see, the UIE networks with the pre-

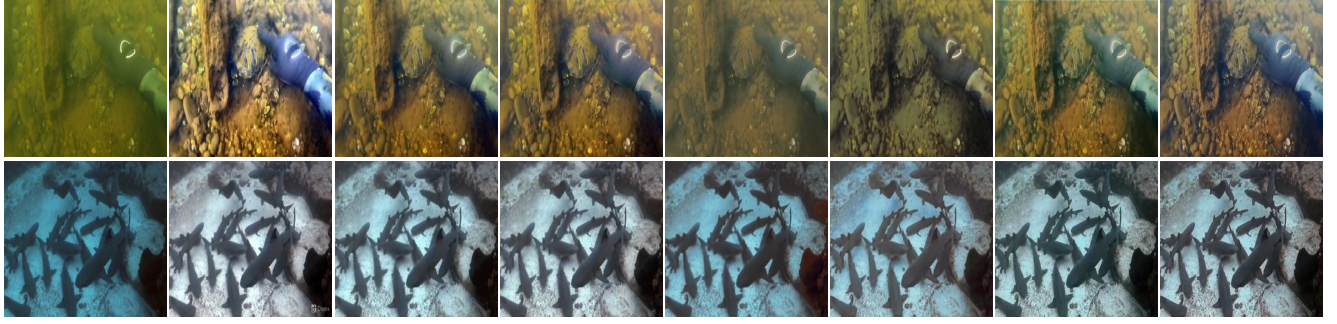


Figure 8: Visual comparison of UIE networks with and without pre-trained URanker as additional supervision. Model+U represents the model is optimized by the pre-trained URanker loss.

Baselines	SRCC $\uparrow$	KRCC $\uparrow$
w/o HPM	0.8585	0.7381
w/ direct	0.8478	0.7219
w/ neighbour	0.8589	0.7363
w/ dense	0.8497	0.7225
full model	0.8655	0.7402

Table 3: Ablation study on the HPM and different connection manners. **Boldface** indicates the best result.

Baselines	PSNR $\uparrow$	SSIM $\uparrow$
w/o normalization tail	18.857	0.8682
w/ Sigmoid	18.249	0.8292
w/ clip	19.963	0.8837
w/ IN+Sigmoid	19.181	0.8797
w/ IN+clip	18.257	0.7559
full model	22.419	0.9227

Table 4: Ablation study on the normalization tail and different operators. **Boldface** indicates the best result.

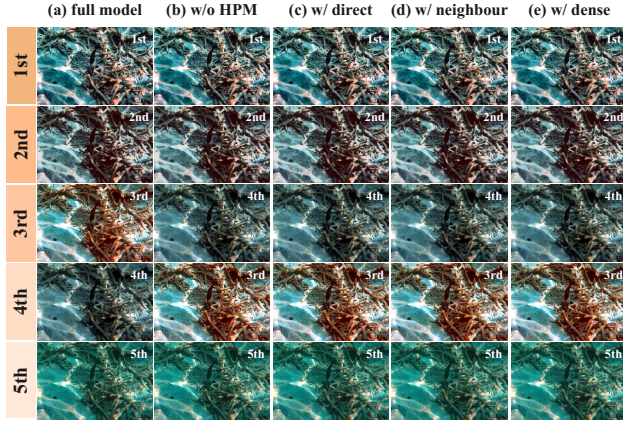


Figure 9: Visual comparison of ablation results on HPM and DCPB. The prediction scores of different ablated models decrease from top to bottom. The ground truth ranking is marked at the top-right corner of each image. Results with incorrect ranking are marked in red.

trained URanker as additional supervision generate results of better visual quality. Besides, our NU²Net is able to generate the most visually pleasing results.

In Table 2, the pre-trained URanker as additional supervision can significantly improve the performance for all three UIE networks. Besides, the proposed NU²Net achieves state-of-the-art performance when compared with the WaterNet and Ucolor.

Ablation Study

We conduct a series of ablation studies to verify the effectiveness of the key components of our designs.

Effectiveness of HPM & DCPB. We remove the histogram prior module (HPM) or replace the dynamic connection parallel block (DCPB) with other connection manners to respectively verify their impacts on UIQA performance.

In Table 3, when the HPM is removed, the performance of the ablated model decreases 0.80% and 0.28% in terms of SRCC and KRCC, respectively. The full model achieves the best performance compared with the other three variants of DCPB. In addition, the model with dense connection performs worse than the neighbour connection, which demonstrates the over-communication of multi-scale features may pollute the original features and further supports the necessity of dynamic connection used in our method. In Figure 9, we show that only the full model can achieve accurate ranking for the five enhanced results. The results further suggest the effectiveness of the proposed HPM and DCPB.

Effectiveness of Normalization Tail. We further analyze the effectiveness of the normalization tail used in our NU²Net. Firstly, the normalization tail is removed as a comparison. Second, we replace it with $[0, 1]$ clip and Sigmoid activation function to handle the outliers. Moreover, since the stretching operation is similar to the Instance Normalization (IN) (Ulyanov, Vedaldi, and Lempitsky 2016), which adjusts the distribution of data by the mean and variance, IN+Sigmoid and IN+clip are also used in ablation studies.

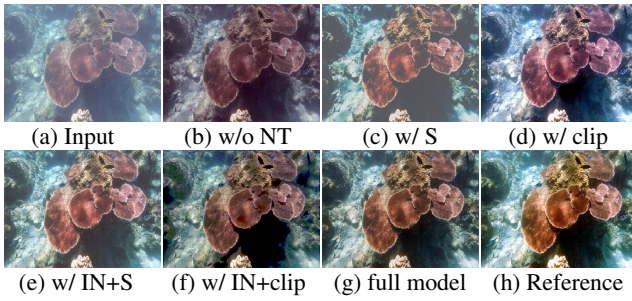


Figure 10: Visual comparison of ablation results on the normalization tail. 'NT' denotes the normalization tail and 'S' the Sigmoid function.

The comparison results are presented in Table 4. The performance can increase 2.456dB and 0.039 at least in terms of PSNR and SSIM by our normalization tail. As Figure 10 shows, the full model produce the perceptually best result compared to others. The results demonstrate this simple operation can boost performance without introducing additional parameters.

Conclusion

In this paper, we propose a new UIQA method to overcome the limitations of previous methods. The success of our method mainly lies in the task-specific designs such as the histogram prior module and dynamic connection parallel block together with a new underwater image with ranking dataset. We also demonstrate the proposed UIQA model can facilitate the performance improvement of UIE networks. Besides, the proposed normalization tail significantly boosts the UIE performance. Both the proposed UIQA model and UIE model outperform the state-of-the-art methods.

Acknowledgements. This work is funded by National Key Research and Development Program of China Grant No.2018AAA0100400, China Postdoctoral Science Foundation (NO.2021M701780). We are also sponsored by CAAI-Huawei MindSpore Open Fund and the code implemented by MindSpore framework is also provided.

References

Akkaynak, D.; Treibitz, T.; Shlesinger, T.; Loya, Y.; Tamir, R.; and Iluz, D. 2017. What is the space of attenuation coefficients in underwater computer vision? In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Ancuti, C.; Ancuti, C. O.; Haber, T.; and Bekaert, P. 2012. Enhancing underwater images and videos by fusion. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Drews, P. L.; Nascimento, E. R.; Botelho, S. S.; and Campos, M. F. M. 2016. Underwater depth estimation and image restoration based on single images. *IEEE Computer Graphics and Applications*.

Fu, X.; Fan, Z.; Ling, M.; Huang, Y.; and Ding, X. 2017. Two-step approach for single underwater image enhancement. In *International Symposium on Intelligent Signal Processing and Communication Systems*.

Fu, X.; Zhuang, P.; Huang, Y.; Liao, Y.; Zhang, X.-P.; and Ding, X. 2014. A retinex-based enhancing approach for single underwater image. In *IEEE International Conference on Image Processing*.

Fu, Z.; Fu, X.; Huang, Y.; and Ding, X. 2022. Twice Mixing: A rank learning based quality assessment approach for underwater image enhancement. *Signal Processing: Image Communication*.

Galdran, A.; Pardo, D.; Picón, A.; and Alvarez-Gila, A. 2015. Automatic red-channel underwater image restoration. *Journal of Visual Communication and Image Representation*.

He, K.; Sun, J.; and Tang, X. 2010. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Iqbal, K.; Odetayo, M.; James, A.; Salam, R. A.; and Talib, A. Z. H. 2010. Enhancing the low quality images using unsupervised colour correction method. In *2010 IEEE International Conference on Systems, Man and Cybernetics*.

Jiang, Q.; Gu, Y.; Li, C.; Cong, R.; and Shao, F. 2022. Underwater Image Enhancement Quality Evaluation: Benchmark dataset and objective metric. *IEEE Transactions on Circuits and Systems for Video Technology*.

Ke, J.; Wang, Q.; Wang, Y.; Milanfar, P.; and Yang, F. 2021. Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*.

Li, C.; Anwar, S.; Hou, J.; Cong, R.; Guo, C.; and Ren, W. 2021. Underwater image enhancement via medium transmission-guided multi-color space embedding. *IEEE Transactions on Image Processing*.

Li, C.; Anwar, S.; and Porikli, F. 2020. Underwater scene prior inspired deep underwater image and video enhancement. *Pattern Recognition*.

Li, C.; Guo, C.; Ren, W.; Cong, R.; Hou, J.; Kwong, S.; and Tao, D. 2019. An underwater image enhancement benchmark dataset and beyond. *IEEE Transactions on Image Processing*.

Li, C.; Guo, J.; and Guo, C. 2018. Emerging from Water: Underwater image color correction based on weakly supervised color transfer. *IEEE Signal Processing Letters*.

Li, C.; Guo, J.; Guo, C.; Cong, R.; and Gong, J. 2017a. A hybrid method for underwater image correction. *Pattern Recognition Letters*.

Li, C.-Y.; Guo, J.-C.; Cong, R.-M.; Pang, Y.-W.; and Wang, B. 2016. Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior. *IEEE Transactions on Image Processing*.

Li, J.; Skinner, K. A.; Eustice, R. M.; and Johnson-Roberson, M. 2017b. WaterGAN: Unsupervised generative network to enable real-time color correction of monocular underwater images. *IEEE Robotics and Automation Letters*.

Liu, X.; Van De Weijer, J.; and Bagdanov, A. D. 2017. RankIQA: Learning from rankings for no-reference image quality assessment. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*.

- Mittal, A.; Moorthy, A. K.; and Bovik, A. C. 2012. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12): 4695–4708.
- Mittal, A.; Soundararajan, R.; and Bovik, A. C. 2012. Making a “completely blind” image quality analyzer. *IEEE Signal Processing Letters*, 20(3): 209–212.
- Ou, F.-Z.; Wang, Y.-G.; Li, J.; Zhu, G.; and Kwong, S. 2021. A Novel Rank Learning Based No-Reference Image Quality Assessment Method. *IEEE Transactions on Multimedia*.
- Panetta, K.; Gao, C.; and Agaian, S. 2015. Human-visual-system-inspired underwater image quality measures. *IEEE Journal of Oceanic Engineering*.
- Peng, Y.-T.; Cao, K.; and Cosman, P. C. 2018. Generalization of the dark channel prior for single image restoration. *IEEE Transactions on Image Processing*.
- Peng, Y.-T.; and Cosman, P. C. 2017. Underwater image restoration based on image blurriness and light absorption. *IEEE Transactions on Image Processing*.
- Ren, W.; Liu, S.; Zhang, H.; Pan, J.; Cao, X.; and Yang, M.-H. 2016. Single image dehazing via multi-scale convolutional neural networks. In *European Conference on Computer Vision*.
- Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-assisted Intervention*, 234–241. Springer.
- Talebi, H.; and Milanfar, P. 2018. NIMA: Neural image assessment. *IEEE Transactions on Image Processing*, 27(8): 3998–4011.
- Ulyanov, D.; Vedaldi, A.; and Lempitsky, V. 2016. Instance Normalization: The Missing Ingredient for Fast Stylization. arXiv:1607.08022.
- Xu, W.; Xu, Y.; Chang, T.; and Tu, Z. 2021. Co-scale conv-attentional image transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*.
- Yang, M.; and Sowmya, A. 2015. An underwater color image quality evaluation metric. *IEEE Transactions on Image Processing*.
- Yang, N.; Zhong, Q.; Li, K.; Cong, R.; Zhao, Y.; and Kwong, S. 2021. A reference-free underwater image quality assessment metric in frequency domain. *Signal Processing: Image Communication*.
- Zhang, W.; Liu, Y.; Dong, C.; and Qiao, Y. 2019. RankSR-GAN: Generative adversarial networks with ranker for image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*.