UVEB: A Large-scale Benchmark and Baseline Towards Real-World Underwater Video Enhancement

Supplementary Material

1. More experiments

1.1. Experiments on underwater image datasets

On benchmark dataset UIEB. UIEB is the benchmark dataset of underwater image enhancement [1]. We select all the data from the UIEB dataset for testing. As shown in Tab. 1, our method achieves better enhancement effects on the UIEB dataset even without training on the UIEB dataset. The performance of UVE-Net is superior to the baseline method Water-Net [1] trained on the UIEB dataset. The results of other methods are based on the official data published by UIEB [1].

On challenging dataset CDUIE. CDUIE [12] is a new publicly available challenging dataset for underwater image enhancement in 2023. The images in the dataset are taken in the Great Lake Superior, USA. These challenging images do not have reference images. We qualitatively compare our method with other recently published methods in the CDUIE [12] dataset. As shown in Fig. 1, our method achieves better color restoration effects on these challenging datasets while the enhancement results of other methods have artifacts or obvious color deviation. Our method trained on the UVEB dataset can achieve better enhancement effects on these challenging data due to the various types of color deviations in the videos of the UVEB dataset (especially green and yellow color deviations).

1.2. Experiments on network performance

Model speed. We evaluate our UVE-Net-s on different devices in Tab. 2. Our model can achieve 6.17 FPS on a computational-constrained embedded development board for 2K videos.

Model performance on video quality metric. Video Quality Metric In Tab. 3, we evaluated the enhancement results of different methods using the DCT-based Video Quality Metric (VQM), and our method still performed well on the VQM metric.

Impact on downstream tasks. Fig. 2 shows that UVE-Net trained on UVEB can increase the number of matching feature points in 3D reconstruction / SLAM tasks.

Comparison with video enhancement method. we reproduce and train the video deblurring method DST-Net of *CVPR2023* on the UVEB dataset with two different settings for comparison in Tab. 4. Our method achieves better results than DST-Net.

1.3. Ablation experiments about UVE-Net.

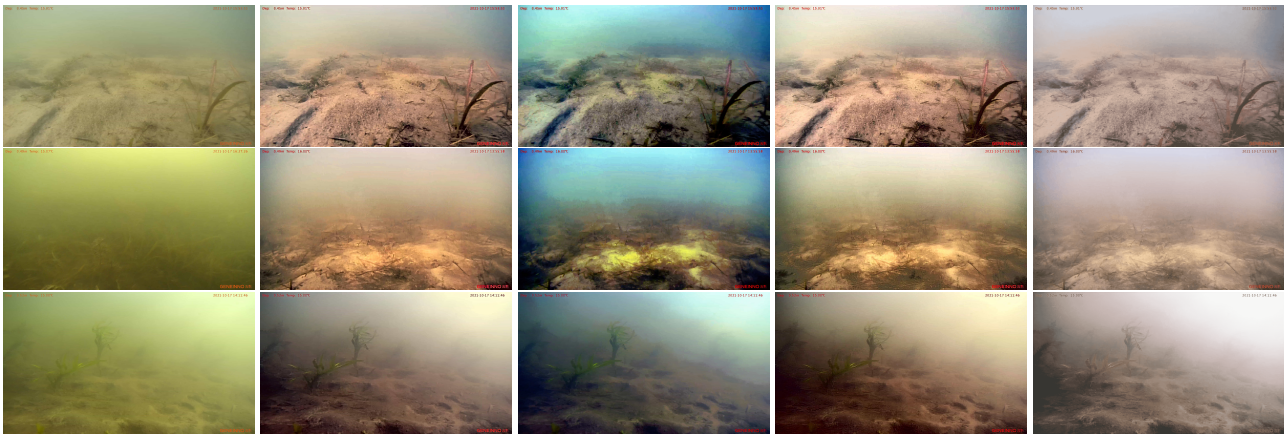
Ablation about the impact of drastic scene changes on model performance. Our method remains effective even when the scene undergoes drastic changes.To verify this, we design an experiment to simulate possible "large changes" by $\times 2$ to $\times 8$ downsampling video frame sequences (not frames). Let V_1 denote a raw video with obvious scene changes. Its $\times 2$ to $\times 8$ downsampling results are represented as V_2, V_4, and V_8. As shown in Tab. 5, the performance is substantially unchanged even if we use V_8 to simulate an ultra-large scene change for inference. The results indicate that our method can adapt to large environmental changes under the same type of water degradation. We also evaluated V_1 without utilizing adjacent frames, which would be a special case ($t = 0$ for inference), and obtained a PSNR of 23.7134 (almost the same as $t = 1$ case).

Ablation about the utilization of UVE-Net on water-degradation characteristics. Based on the similarity of water degradation types in adjacent frames in real-world scenarios, UVE-Net allows adjacent frames and middle frame to follow similar feature enhancement processes. We validated this network design through experiments in Tab. 6. We choose 3 underwater videos with 2 degradation types (*e.g.*, A (blue), B (green), and C (green)) and shuffle their frames in 2 ways. In Tab. 6, the middle frame can still provide practical guidance to adjacent frames in different environments and similar types of water degradation (BCB). The results indicate that UVE-Net can effectively utilize the types of water degradation to guide the adjacent frames.

Ablation on the number of convolutional kernels used for guidance. We further explore the impact of using different numbers of convolutional kernels to guide the enhancement process of VE-Net. In our proposed setting, the input patch size is 512×512, and the number of channels for R and R/6 modules is 24. The number of channels for the R/2 module is 96. The number of convolutional kernels used to convey guidance information is 384. Following the setting, the 32×32 pixels information is converted into a 3×3 convolutional kernel with an information conversion ratio of approximately 100. During the guidance process, the number of network channels for VE-Net changed from 24 to 384 and then to 96. When we use 96 convolutional kernels for guidance, the 64×64 pixels information is converted to a 3×3 convolutional kernel with an information conversion ratio of approximately 400. During the guidance process, the number of network channels for VE-Net changed from 24 to 96. When we use 1536 convolutional

Table 1. Quantitative comparisons of enhanced video quality on UIEB [1] dataset.

Methods	fusion-based [2]	retinex-based [3]	GDCP [4]	histogram prior [5]	blurriness-based [6]	Water CycleGAN [7]	Dense GAN [8]	Water-Net [1]	Ours
PSNR(dB)↑	17.6077	17.0168	12.0929	15.8215	15.3180	15.7508	17.2843	19.1130	19.3813
MSE($\times 10^3$)↓	1.1280	1.2924	4.0160	1.7019	1.9111	1.7298	1.2152	0.7976	0.4059



Raw MLLE [9] USUIR [10] WWPF [11] Ours

Figure 1. Qualitative comparison of enhanced video quality on the CDUIE [12] dataset.

kernels, the information conversion ratio is approximately 25. During the guidance process, the number of network channels for VE-Net changed from 24 to 1536 and then to 96. As shown in Tab. 7, When the number of convolution kernels is 384, UVE-Net achieves the best enhancement results. Although the number of convolution kernels is 96, the PSNR metric will increase by 2.0% (0.53dB), and the MSE metric will decrease by 5.3% (21.4). Overall, the enhancement performance is better when the number of convolution kernels is 384. From Tab. 7, although intuitively reducing the information conversion ratio can improve the video enhancement effect, the drastic changes in the number of network channels also have a negative impact on the network performance. Therefore, when the number of convolution kernels is 1536, the network performance is poor.

1.4. More visual comparisons.

We present more visual comparisons between UVE-Net and other methods in Fig. 4, Fig. 5, and Fig. 6. As shown in Fig. 4, our method has better color restoration effects and we can see the details of underwater plants and soil more clearly. From Fig. 5 and Fig. 6, it can be seen that our method has realistic image colors and textures. We also display the comparison of enhancement effects of different methods on real underwater videos in the provided video materials.

2. More discussions.

The UVEB dataset selects the optimal enhancement results from 20 enhancement methods as the original data's Ground Truth (GT). To some extent, it embodies the advantages of most current methods. Thus, it is unsurprising that our method trained on the UVEB dataset can achieve better enhancement results. We would like to share some issues we meet in this work, as well as discussions on the UVEB dataset and the UVE-Net.

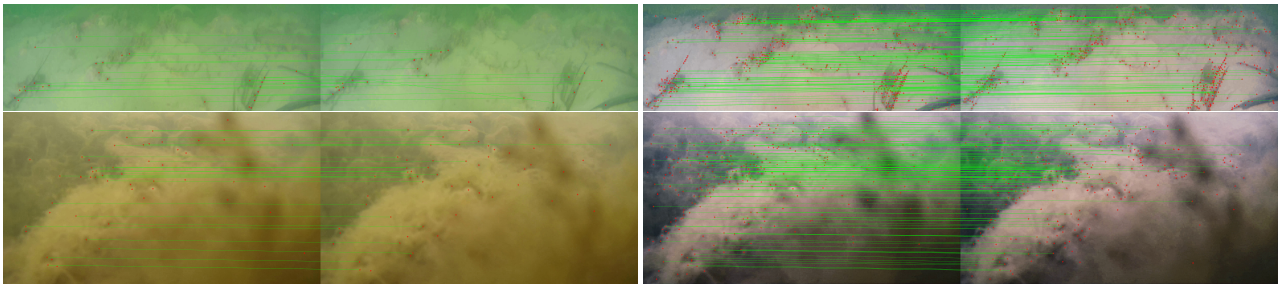
2.1. No-reference underwater image quality evaluation metrics

Unlike other works, we did not use no-reference underwater image quality evaluation metrics to evaluate the enhancement results because these evaluation metrics are not accurate enough for underwater video quality evaluation.

We show the evaluation results of three existing no-reference underwater image quality evaluation metrics in Fig. 3. The top row shows the image with better quality, while the bottom row shows the image with poorer quality. We can see that three metrics give higher ratings to inferior-quality images. Generally, the better the image quality, the higher the UCIQE [13], CCF [15], and FUDM [14] assessment scores of the image. These metrics may give incorrect ratings because the image contrast in the bottom row is higher. These evaluation metrics overly rely on image contrast and appear inflexible. We also measured the enhancement results of 20 methods on the UVEB test dataset using the UCIQE[13] metric. We compared the evaluation

Table 2. Performance in resource-constrained environments. (INT8 stands for Tensort's accelerated INT8 quantization model)

GPU	A40	RTX3090	RTX3060	RTX3070 (laptop)	Orin NX(INT8)
CPU	Xeon Silver 4314	Xeon Silver 4314	i7-12700H	i7-11800H	Arm Cortex-A78AE
2K Inference time (s)↓	0.0445	0.0404	0.090	0.120	0.162



Raw Enhanced

Figure 2. Visual comparisons on feature point matching tasks.

Table 3. Evaluate results on video quality metric (**VQM** for short).

Method	CLAHE	USUIR	Uranker	UVE-Net (ours)
VQM↓	1.301	1.027	0.6520	0.6305

results of UCIQE[13] with those of PSNR in Tab. 8. We find that the poorer-performing methods UDCP [16] and HE [17] obtain higher UCIQE[13] scores. This abnormal phenomenon also indicates that the erroneous evaluation examples in Fig. 3 are not accidental individual cases, and the existing no-reference underwater image quality evaluation methods need to be further developed and improved.

Designing learning-based video quality evaluation methods trained with large-scale data may be a good solution. The video quality scores with the UVEB dataset also provide convenience for future works on no-reference underwater video quality assessment.

2.2. Discussions on UVEB and UVE-Net

Similar to previous underwater image enhancement datasets, the quality of GT for some samples in the UVEB dataset may not be perfect due to the limitations of existing underwater image enhancement methods. This problem is difficult to solve unless sufficiently excellent underwater image/video enhancement methods appear. To some extent, developing datasets and learning-based underwater image/video enhancement methods are mutually promoting. Finely annotated datasets bring better methods, and better methods can build higher-quality new datasets. Our main purpose in constructing the UVEB dataset is also to promote the development of underwater video/image enhancement methods.

The large-scale underwater videos contained in the UVEB dataset also provide rich raw materials for other underwater visual tasks. The video score information attached to the UVEB dataset can not only be used for the underwater video quality assessment task but also serve as additional auxiliary information to facilitate future work in designing better underwater video enhancement methods.

The value of UVE-Net lies in completing inter-frame information exchange by passing action instructions (convolutional kernels) and proposing a practical underwater video enhancement framework. People can replace residual modules with existing network structures as needed. We use the residual module to facilitate the construction of an underwater video enhancement baseline.

3. More details in Labeled Sample Generation

The 15 observers are aged between 22 and 27, including seven males and eight females. Due to the degradation types of enhanced underwater videos, including various types of color deviation, insufficient lighting, artifacts, blurring, noise, etc. Before observers score the quality of all videos, We first select 1743 videos (83×21) covering various scenarios and degradation types for the assessment test. We noted three issues during the test:

1. Before observers understand the overall video quality, their ratings are unstable and inaccurate.
2. When observers are required to double-check videos after a period, the ratings before and after vary greatly.
3. The workload of annotating 21×1308 videos (20 methods) is enormous, and the enhancement results of some methods are generally poor. Scoring multiple enhancement results for a video is not necessary.

For issue 1, we designed a "pretext task" to establish the observers' perception of the overall video quality before labeling. Firstly, we chose 150 videos with rating spans with various types of videos to construct the example li-

Table 4. Comparison with video enhancement methods. (C:channels R:Resnet blocks)

Method setting	DST-Net C:64,R:15(default)	DST-Net C:24,R:15	UVE-Net-s	UVE-Net
Memory(G)↓	9.7832	8.2695	2.6162	5.404
2K Inference time(s)↓	0.4154	0.141	0.0404	0.4533
PSNR(dB)↑	25.3352	24.8576	24.43	26.27

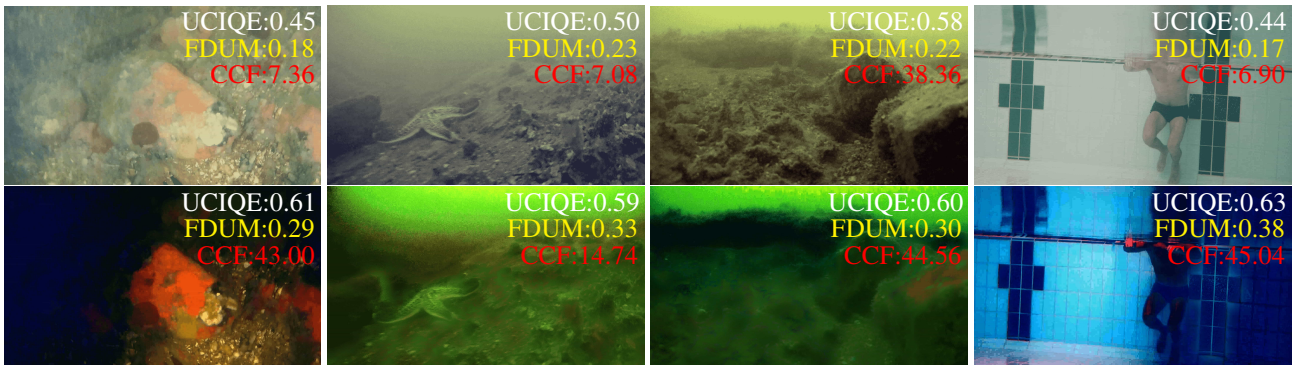


Figure 3. Visual comparisons in terms of UCIQE [13], FUDM [14], and CCF [15] metrics. Better image quality does not receive higher evaluation scores.

Table 5. The impact of scene change speed on UVE-Net performance.

Scene change speed	×1(default)	×2	×4	×8
Results PSNR (dB)↑	23.7137	23.7114	23.7045	23.7095

Table 6. The impact of adjacent frame water body changes on enhancement effect.

Guidance way	C guide B	A guide B	B guide B (default)
Adjacent frame sequence	...BCBBCB...	...BABBAB...	...BBBBBB...(default)
Results PSNR(dB)↑	21.36	11.36	23.37

Table 7. Ablation on the number of convolutional kernels used for guidance.

Number	PSNR(dB)↑	MSE($\times 10^3$)↓
96	26.80	0.4273
384	26.27	0.4059
1536	23.11	0.5432

brary. Secondly, we asked observers to score and sort the 150 videos in increasing order of quality. Next, we asked the observers about the reasons for ranking and scoring any videos. When observers could not give reasons, they were asked to reorder the videos. We repeated the process until the observers were able to justify their scoring. After scoring through the videos in the example library, observers can understand the overall video quality and form reasonable evaluation logic.

Table 8. Quantitative comparisons in terms of UCIQE [13] and PSNR metrics on UVEB dataset.

Methods	PSNR(dB)↑	UCIQE [13]↑
PUIE [18]	24.21	0.5486
URanker [19]	23.93	0.5682
USUIR [10]	21.64	0.6199
LANet [20]	21.49	0.5365
CLAHE [21]	19.71	0.5592
Red Channle [22]	19.61	0.5410
CLUIE [23]	19.44	0.5517
MLLE [9]	18.79	0.5891
retinex-based [3]	18.75	0.5915
FspiralGAN [24]	18.67	0.6288
fusion-based [2]	17.73	0.6350
WWPF [11]	17.67	0.6049
GC [25]	16.61	0.4634
MetaUE [26]	15.91	0.5535
HE [17]	15.78	0.6596
FA$^+$Net [27]	15.34	0.5483
GDCP [4]	13.33	0.5591
MSCNN [28]	13.17	0.5478
DCP [29]	13.03	0.5432
UDCP [16]	10.75	0.5697

For issue 2, we disrupted the order of the 150 videos

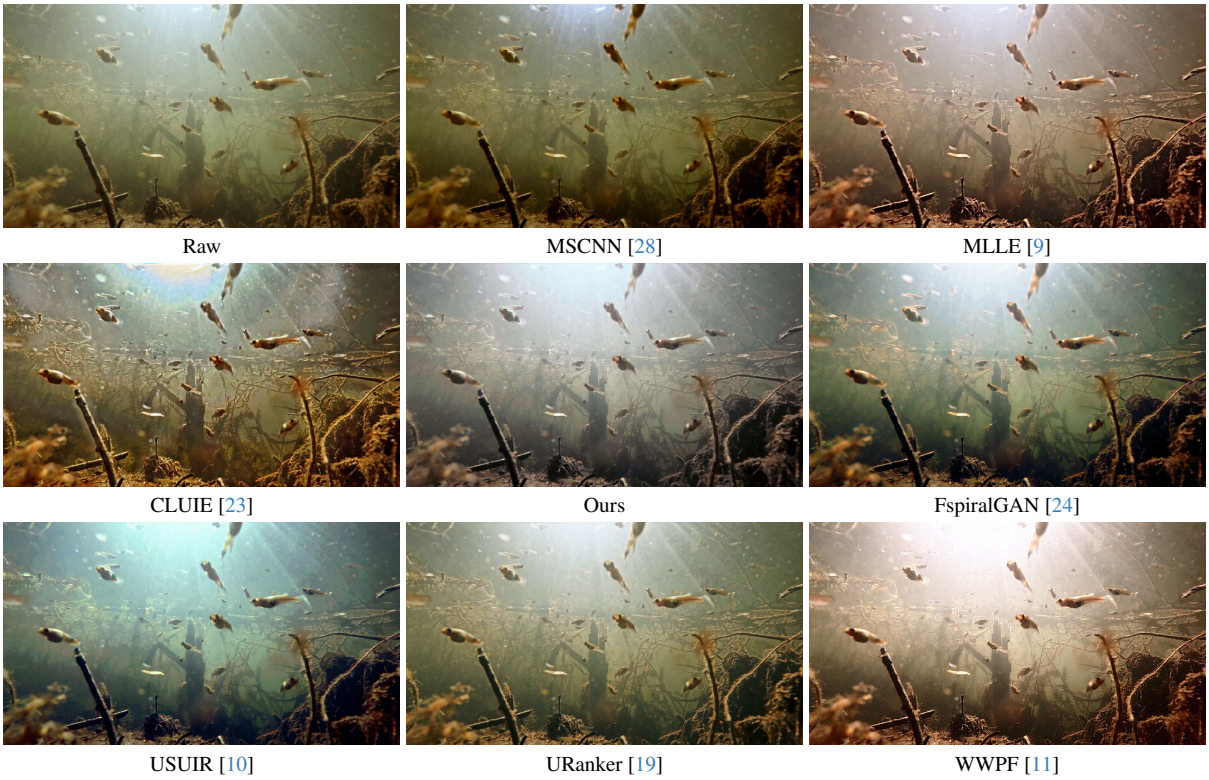


Figure 4. Visual comparisons with state-of-the-art methods on real underwater scenes.

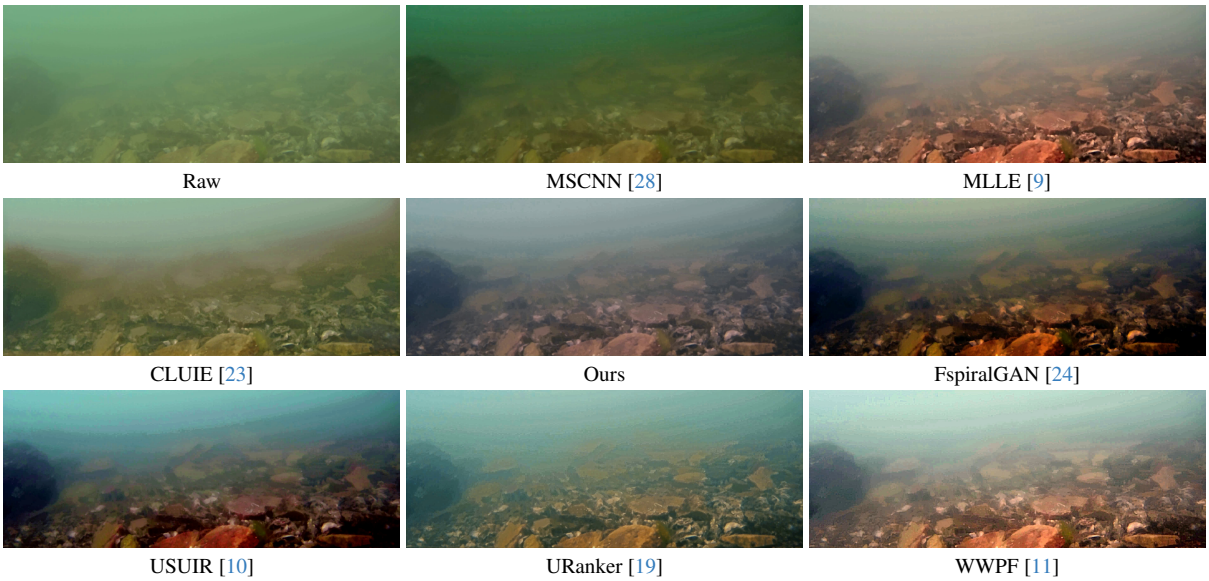


Figure 5. Visual comparisons with state-of-the-art methods on real underwater scenes.

and asked observers to reorder the videos one day later. If the difference between the observer's ratings for the same video was greater than 5 points, the observer was asked to reorder the videos and repeat this checking process the next day until the observer gave a relatively stable rating for the same video. The sorted example libraries obtained by each observer through the process were used as respective video quality scales. Observers could view their scales if they were unsure of their ratings and watch the videos many times to give more definitive ratings.

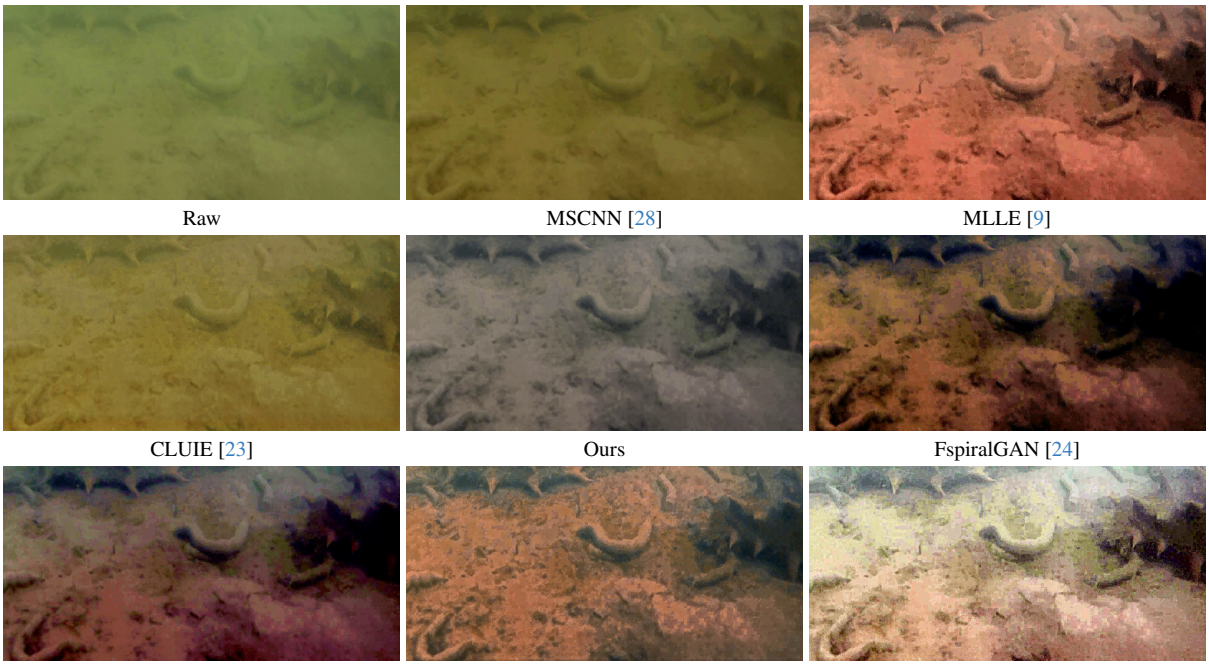


Figure 6. Visual comparisons with state-of-the-art methods on real underwater scenes.

For issue 3, to reduce the workload, we asked observers to score the 83 sets of videos and selected the best 10 methods based on the total points of enhancement results. For the remaining videos, the observers only needed to select the optimal enhancement result from 10 better methods and score the optimal enhancement result and the raw video.

Observers can use the raw video and 10 better enhancement results as a set. These observers were advised to complete one set of video assessments in about 5 minutes and 40 to 50 sets per day. Observers could schedule their own annotation time for free but were required to watch their example scales before each day's annotation to stabilize their ratings further.

References

[1] Chongyi Li, Chunle Guo, Wenqi Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao. An underwater image enhancement benchmark dataset and beyond. *IEEE Transactions on Image Processing*, 29:4376–4389, 2019. 1, 2

[2] Cosmin Ancuti, Codruta Orniana Ancuti, Tom Haber, and Philippe Bekaert. Enhancing underwater images and videos by fusion. In *2012 IEEE conference on computer vision and pattern recognition*, pages 81–88. IEEE, 2012. 2, 4

[3] Xueyang Fu, Peixian Zhuang, Yue Huang, Yinghao Liao, Xiao-Ping Zhang, and Xinghao Ding. A retinex-based enhancing approach for single underwater image. In *2014 IEEE international conference on image processing (ICIP)*, pages 4572–4576. IEEE, 2014. 2, 4

[4] Yan-Tsung Peng, Keming Cao, and Pamela C Cosman. Generalization of the dark channel prior for single image restoration. *IEEE Transactions on Image Processing*, 27(6):2856–2868, 2018. 2, 4

[5] Chong-Yi Li, Ji-Chang Guo, Run-Min Cong, Yan-Wei Pang, and Bo Wang. Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior. *IEEE Transactions on Image Processing*, 25(12):5664–5677, 2016. 2

[6] Yan-Tsung Peng and Pamela C Cosman. Underwater image restoration based on image blurriness and light absorption. *IEEE transactions on image processing*, 26(4):1579–1594, 2017. 2

[7] Chongyi Li, Jichang Guo, and Chunle Guo. Emerging from water: Underwater image color correction based on weakly supervised color transfer. *IEEE Signal processing letters*, 25(3):323–327, 2018. 2

[8] Yecai Guo, Hanyu Li, and Peixian Zhuang. Underwater image enhancement using a multiscale dense generative adversarial network. *IEEE Journal of Oceanic Engineering*, 45(3):862–870, 2019. 2

[9] Weidong Zhang, Peixian Zhuang, Hai-Han Sun, Guohou Li, Sam Kwong, and Chongyi Li. Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement. *IEEE Transactions on Image Processing*, 31:3997–4010, 2022. 2, 4, 5, 6

[10] Zhenqi Fu, Huangxing Lin, Yan Yang, Shu Chai, Liyan Sun, Yue Huang, and Xinghao Ding. Unsupervised underwater image restoration: From a homology perspective. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 643–651, 2022. 2, 4, 5, 6

[11] Weidong Zhang, Ling Zhou, Peixian Zhuang, Guohou Li, Xipeng Pan, Wenyi Zhao, and Chongyi Li. Underwater image enhancement via weighted wavelet visual perception fusion. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023. 2, 4, 5, 6

[12] Ashraf Saleem, Sidike Paheding, Nathir Rawashdeh, Ali Awad, and Navjot Kaur. A non-reference evaluation of underwater image enhancement methods using a new underwater image dataset. *IEEE Access*, 11:10412–10428, 2023. 1, 2

[13] Miao Yang and Arcot Sowmya. An underwater color image quality evaluation metric. *IEEE Transactions on Image Processing*, 24(12):6062–6071, 2015. 2, 3, 4

[14] Ning Yang, Qihang Zhong, Kun Li, Runmin Cong, Yao Zhao, and Sam Kwong. A reference-free underwater image quality assessment metric in frequency domain. *Signal Processing: Image Communication*, 94:116218, 2021. 2, 4

[15] Yan Wang, Na Li, Zongying Li, Zhaorui Gu, Haiyong Zheng, Bing Zheng, and Mengnan Sun. An imaging-inspired no-reference underwater color image quality assessment metric. *Computers & Electrical Engineering*, 70:904–913, 2018. 2, 4

[16] Paulo LJ Drews, Erickson R Nascimento, Silvia SC Botelho, and Mario Fernando Montenegro Campos. Underwater depth estimation and image restoration based on single images. *IEEE computer graphics and applications*, 36(2):24–35, 2016. 3, 4

[17] Robert Hummel. Image enhancement by histogram transformation. *Unknown*, 1975. 3, 4

[18] Zhenqi Fu, Wu Wang, Yue Huang, Xinghao Ding, and Kai-Kuang Ma. Uncertainty inspired underwater image enhancement. In *European Conference on Computer Vision*, pages 465–482. Springer, 2022. 4

[19] Chunle Guo, Ruiqi Wu, Xin Jin, Linghao Han, Weidong Zhang, Zhi Chai, and Chongyi Li. Underwater ranker: Learn which is better and how to be better. In *AAAI Conference on Artificial Intelligence*, volume 37, pages 702–709, 2023. 4, 5, 6

[20] Shiben Liu, Huijie Fan, Sen Lin, Qiang Wang, Naida Ding, and Yandong Tang. Adaptive learning attention network for underwater image enhancement. *IEEE Robotics and Automation Letters*, 7(2):5326–5333, 2022. 4

[21] Karel Zuiderveld. *Contrast Limited Adaptive Histogram Equalization*, page 474–485. Academic Press Professional, Inc., USA, 1994. 4

[22] Adrian Galdran, David Pardo, Artzai Picón, and Aitor Alvarez-Gila. Automatic red-channel underwater image restoration. *Journal of Visual Communication and Image Representation*, 26:132–145, 2015. 4

[23] Kunqian Li, Li Wu, Qi Qi, Wenjie Liu, Xiang Gao, Liqin Zhou, and Dalei Song. Beyond single reference for training: underwater image enhancement via comparative learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022. 4, 5, 6

[24] Yang Guan, Xiaoyan Liu, Zhibin Yu, Yubo Wang, Xingyu Zheng, Shaoda Zhang, and Bing Zheng. Fast underwater image enhancement based on a generative adversarial framework. *Frontiers in Marine Science*, 9:964600, 2023. 4, 5, 6

[25] Christophe Schlick. Quantization techniques for visualization of high dynamic range pictures. In *Photorealistic rendering techniques*, pages 7–20. Springer, 1995. 4

[26] Zhenwei Zhang, Haorui Yan, Ke Tang, and Yuping Duan. Metaue: Model-based meta-learning for underwater image enhancement. *arXiv preprint arXiv:2303.06543*, 2023. 4

[27] Jingxia Jiang, Tian Ye, Jinbin Bai, Sixiang Chen, Wenhao Chai, Shi Jun, Yun Liu, and Erkang Chen. Five a $^+$ network: You only need 9k parameters for underwater image enhancement. *arXiv preprint arXiv:2305.08824*, 2023. 4

[28] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, pages 154–169. Springer, 2016. 4, 5, 6

[29] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1956–1963, 2009. 4