Supplementary Material – Five A$^+$ Network: You Only Need 9K Parameters for Underwater Image Enhancement

Jingxia Jiang[1†]
202021114006@jmu.edu.cn
Tian Ye[2†]
owentianye@hkust-gz.edu.cn
Jinbin Bai[3†]
jinbin.bai@u.nus.edu
Sixiang Chen[2]
sixiangchen@hkust-gz.edu.cn
Wenhao Chai[4]
reselfchai@gmail.com
Shi Jun[5]
junshi2022@gmail.com
Yun Liu[6]
yunliu@swu.edu.cn
Erkang Chen[1,7*]
ekchen@jmu.edu.cn

[1] School of Ocean Information Engineering, Jimei University, Xiamen, China
[2] Hong Kong University of Science and Technology (GZ), Guangzhou, China
[3] Department of Computer Science, National University of Singapore, Singapore
[4] University of Washington, State of Washington, USA
[5] the School of Information Science and Engineering, Xinjiang University, Uruqmi, China
[6] College of Artificial Intelligence, Southwest University, Chongqing, China
[7] Fujian Provincial Key Laboratory of Oceanic Information Perception and Intelligent Processing, China

This is a supplementary material for Five A$^+$ Network: You Only Need 9K Parameters for Underwater Image Enhancement.

We provide the following materials in this manuscript:

- Sec. 1 performance vs. run-time.
- Sec. 2 additional ablation experiments on the element configurations.
- Sec. 3 more visual comparisons.
- Sec. 4 future works and broader impacts.

Table 1: Comparison of speed, GFLOPs and processing frames of previous SOTA methods in different resolutions on RTX 3090 (24G). The best and second best results are highlighted in red and blue, respectively. When there is a tied result, it is indicated in green. Furthermore, ↑ represents the higher is the better as well as ↓ represents the lower is the better.

	256 × 256 patch			512 × 512 patch			1280 × 720 patch			1920 × 1080 patch			2560 × 1440 patch		
	#Flops(G)↓	#Runtime(s)↓	FPS(f/s)↑	#Flops(G)↓	#Runtime(s)↓	FPS(f/s)↑	#Flops(G)↓	#Runtime(s)↓	FPS(f/s)↑	#Flops(G)↓	#Runtime(s)↓	FPS(f/s)↑	#Flops(G)↑	#Runtime(s)↓	FPS(f/s)↑
PRW-Net(ICCVW'21) [■]	15.88G	0.059s	16.799	63.54G	0.074s	13.401	223.40G	0.216s	4.624	Out of Memory			Out of Memory		
Shallow-uwnet(AAAI'21) [■]	21.67G	0.002s	396.948	86.69G	0.008s	115.600	304.75G	0.031s	31.836	685.70G	0.074s	13.418	1219.02G	0.129s	7.741
UIEC^2-Net(SPIC'21) [■]	26.14G	0.026s	37.739	104.54G	0.072s	13.722	367.53G	0.174s	5.742	826.93G	0.383s	2.607	Out of Memory		
PUIE-Net(ECCV'22) [■]	30.08G	0.009s	105.649	120.34G	0.020s	48.035	423.05G	0.071s	14.194	Out of Memory			Out of Memory		
NU2Net(AAAI'23,Oral) [■]	10.43G	0.002s	428.321	41.71G	0.007s	130.850	146.64G	0.024s	42.345	329.95G	0.051s	19.349	586.58G	0.090s	11.077
Ours	0.59G	0.003s	326.372	2.37G	0.007s	133.156	8.33G	0.016s	60.724	18.74G	0.033s	29.943	33.31G	0.057s	17.503

Table 2: Ablation study on the Multi-branch Color Enhancement Module. Specifically, w/o MCEM and w/o IN in MCEM refer to deleting MCEM and remove IN layers in MCEM, respectively. The best result is indicated by underlines.

Configuration details of MCEM	PSNR↑	SSIM↑
a)w/o MCEM	22.528	0.897
b)w/o IN in MCEM	22.892	0.909
c)w/MCEM(IN->BN)	22.717	0.907
d)Ours	23.061	0.911

1 Run-time Comparison

In this section, we demonstrate our speed and computational advantages of our proposed Five A^+ Network (FA^+ Net). All inference stages are conducted on an RTX 3090 to ensure fairness in comparison. We use the torch.cuda.synchronize() API function to accurately measure feed-forward running times. The results in Table 1 show that FA^+ Net outperforms the existing methods, such as PRW Net [3], Shallow-uwnet [6], UIEC^2-Net [7], PUIE-Net [1] and NU2Net [2], in terms of GFLOPs, inference time and FPS, particularly for high resolution images. Notably, even the state-of-the-art UIEC^2-Net [7] architecture encountered a "CUDA out of memory" issue when processing high-resolution images.

2 Additional Ablation Experiments on the Element Configurations

2.1 Effectiveness of the Multi-branch Color Enhancement Module

To showcase the effectiveness of the proposed Multi-branch Color Enhancement Module(MCEM) structure, we present the experimental results in Table 2. It can be observed that the PSNR gain achieved by the Instance Normalization (IN) layer is higher than that of the Batch Normalization (BN) layer, which can be attributed to their distinct normalization methods. BN does not take into consideration the channel differences while IN utilizes different means and variances for each channel. To provide a more intuitive understanding of the MCEM's utility, we visualize the channel histogram before and after applying the enhancement. As

Table 3: Ablation study on the different value of α. Underline and **bold** indicate the best result and the second best result, respectively.

α	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
PSNR	22.831	**23.051**	22.900	22.938	23.061	22.810	0.909	22.812	22.842	22.936
SSIM	0.908	**0.910**	0.907	0.908	0.911	0.904	22.985	0.906	0.907	0.909

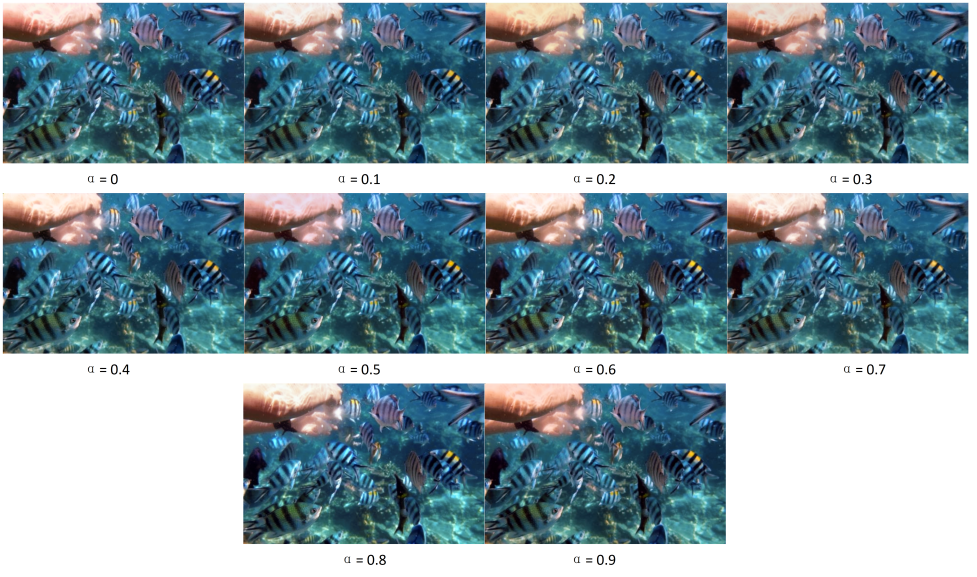

Figure 1: Visual comparison of different α values.

illustrated in Fig. 2 and Fig. 3, MCEM has effectively reduced the difference between the three channels, as anticipated.

2.2 Effectiveness and Visual Effects of Different Value of α

Our Spatial-frequency Domain Feature Interaction Module (SDFIM) employs adjustable hyperparameters α to regulate the fusion of spatial domain and frequency domain information. Table 3 illustrates the effect of distinct α values on model performance. Notably, the model attains the best performance at α of 0.4. Different values of α yield divergent outcomes for the network. As shown in Fig 1, different values of α yield varying results in the network. Notably, the desired chroma can be regulated according to the value of α.

2.3 Effectiveness of downsampling size in Multi-Scale Pyramid Module

In order to explore the influence of down-sampling size on Multi-Scale Pyramid Module (MPM) performance, we adjusted the dimensions and examined their impact. As shown in Table 4, the effectiveness of the erosion model increases with a growing down-sampling size, though at the cost of increased computation time. Once the down-sampling size reached 256, the model's performance gradually shifts toward a plateau. Taking into account efficiency and performance, we eventually had set the multi-scale feature pyramid at the 32×32, 64×64, and 128×128 size.

3 More Visual Comparisons

To further demonstrate the strong generalization performance of FA$^+$ Net on underwater image enhancement, we conduct additional visual comparisons with other state-of-the-art

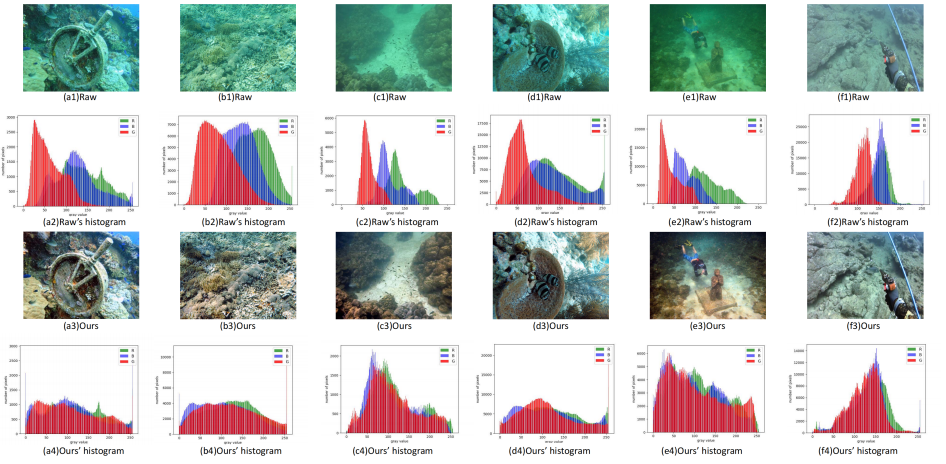

Figure 2: Color correction results of degenerated underwater image on T90 datasets by MCEM.

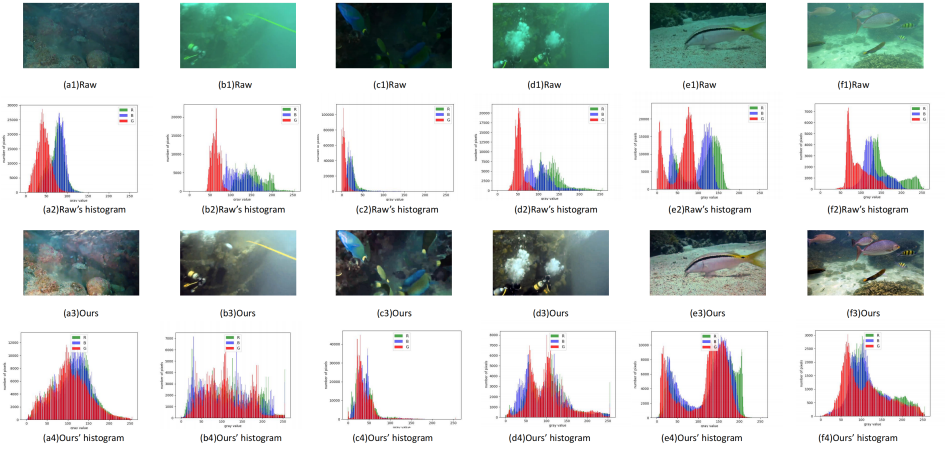

Figure 3: Color correction results of degenerated underwater image on C60 datasets by MCEM.

Table 4: Ablation study on the influence of various down-sampling size of MPM. The efficiency of the models is measured on th a single NVIDIA A100 Tensor Core GPU (40GB), ✓means that the feature size is selected for the Multi-scale Pyramid Module. The best results are underlined. The second-best results are in **bold**.

Model	Multi-scale Pyramid Configuration								Metrics		Efficiency	
	2	4	8	16	32	64	128	256	PSNR↑	SSIM↑	#Runtime(s)↓	FPS(f/s)↑
a)	✓	✓	✓						22.766	0.906	0.0132s	75.398
b)		✓	✓	✓					22.811	0.908	**0.0137s**	**72.941**
c)			✓	✓	✓				22.937	0.908	0.0146s	68.167
d)				✓	✓	✓			22.896	0.910	0.0158s	62.952
e)						✓	✓	✓	**23.001**	0.912	0.0321s	31.060
Ours					✓	✓	✓		23.061	**0.911**	0.0251s	39.751

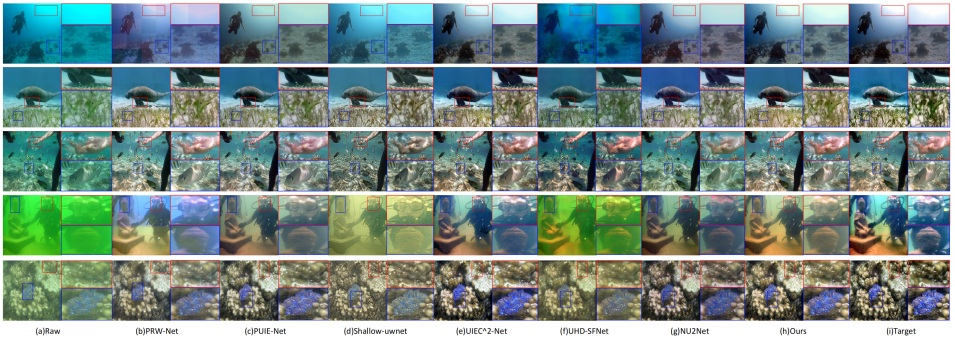

Figure 4: Visual comparisons of degenerated images from the T90 [5] testing dataset. The images can be zoomed in for improved visualization.

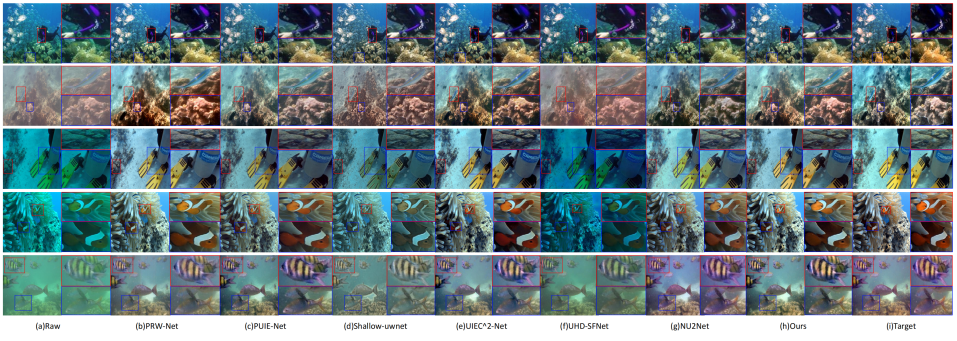

Figure 5: Visual comparisons of degenerated images from the T90 [5] testing dataset. The images can be zoomed in for improved visualization.

methods on T90 [4], C60 [4] and U45 [5] datasets. Additionally, we also present the rendering and augmentation capabilities of the proposed method for virtual scenes.

3.1 Real-world Dataset:T90, C60 and U45

To showcase the remarkable performance of our model in real-world scenarios, we present comprehensive visual comparisons for real-world underwater image enhancement in Fig. 4, 5, 6 and 7. Our observations indicate that the images processed by FA$^+$ Net exhibit improved quality and effectively removed complex degradations, while other algorithms often struggled with complex mixed degradations. Moreover, as shown in Fig.8, the visual and perceptual evaluation of various methods shows that FA$^+$Net consistently achieves superior results, providing further evidence of its efficacy and practical applicability. FA$^+$ Net surpasses previous state-of-the-art methods in effectively handling fine-grained degradations.

3.2 Rendering and Enhancement of Virtual Scenes

The effectiveness of FA$^+$ Net in eliminating complex and challenging virtual underwater scenes is demonstrated in Fig. 9. Owing to its robust high-quality phased design, FA$^+$ Net excels in underwater scene rendering.

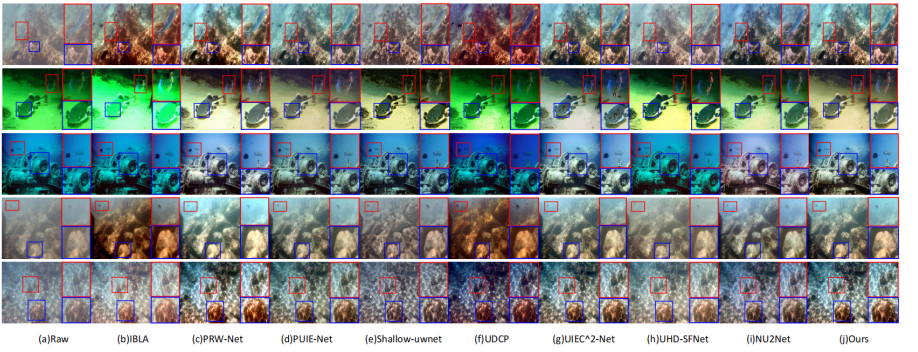

Figure 6: Visual comparisons of degenerated images from the U45 [5] testing dataset. The images can be zoomed in for improved visualization.

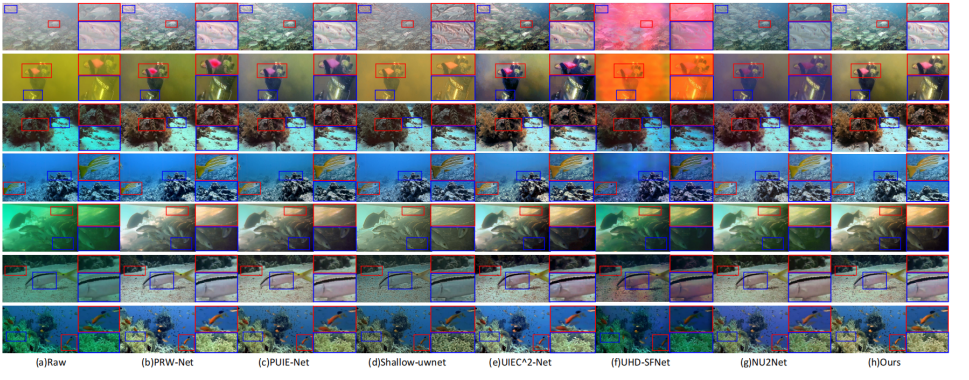

Figure 7: Visual comparisons of degenerated images from the C60 [2] testing dataset. The images can be zoomed in for improved visualization.

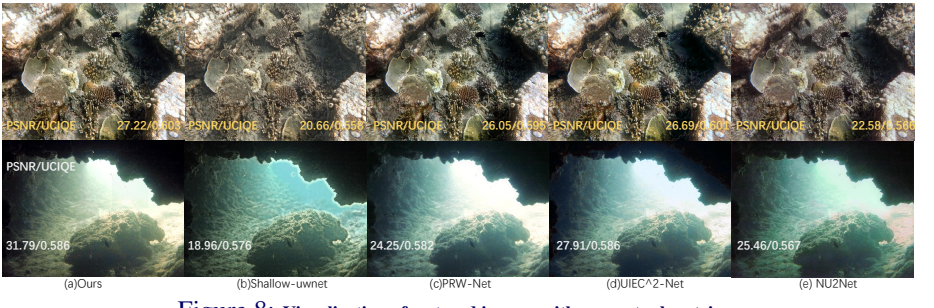

Figure 8: **Visualization of restored images with perceptual metrics.**

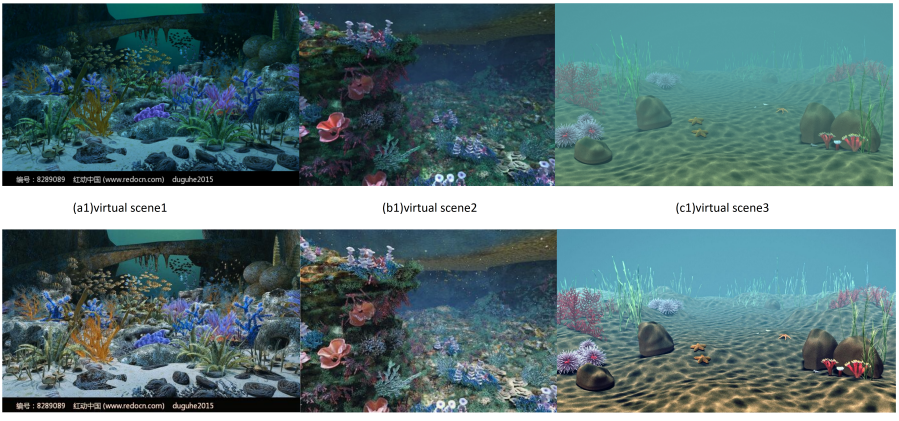

Figure 9: The rendering and enhancing visualization results of virtual scenes achieved by the proposed method.

4 Future Works and Broader Impacts

Moving forward, we plan to explore the full potential of FA$^+$ Net for various tasks, aiming to make our overall architecture more efficient and less computationally complex. Our ultimate goal is to develop a real-time high resolution image processing system that can effectively handle other tasks such as denoising and low-light image enhancement. Our FA$^+$ Net has shown promising results with various challenging underwater conditions.

FA$^+$ Net has demonstrated strong performance on real-world underwater image enhancement, indicating potential applications in various industrial tasks and applications, such as underwater photography, underwater archaeology and marine life detection. Consequently, our work has the potential to contribute positively to both academia and industry. However, it is important to note that this kind of technology should be used responsibly, and potential adverse impacts must be considered. For instance, extreme or uncommon cases of degradation may led to Five A$^+$ Network not functioning optimally, and it should not be solely relied on as a navigation or decision-making aid in such situations.

References

[1] Zhenqi Fu, Wu Wang, Yue Huang, Xinghao Ding, and Kai-Kuang Ma. Uncertainty inspired underwater image enhancement. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XVIII*, pages 465–482. Springer, 2022.

[2] Chunle Guo, Ruiqi Wu, Xin Jin, Linghao Han, Zhi Chai, Weidong Zhang, and Chongyi Li. Underwater ranker: Learn which is better and how to be better. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023.

[3] Fushuo Huo, Bingheng Li, and Xuegui Zhu. Efficient wavelet boost learning-based multi-stage progressive refinement network for underwater image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1944–1952, 2021.

[4] 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.

[5] Hanyu Li, Jingjing Li, and Wei Wang. A fusion adversarial underwater image enhancement network with a public test dataset. *arXiv preprint arXiv:1906.06819*, 2019.

[6] Ankita Naik, Apurva Swarnakar, and Kartik Mittal. Shallow-uwnet: Compressed model for underwater image enhancement (student abstract). In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 15853–15854, 2021.

[7] Yudong Wang, Jichang Guo, Huan Gao, and Huihui Yue. Uiecˆ 2-net: Cnn-based underwater image enhancement using two color space. *Signal Processing: Image Communication*, 96:116250, 2021.