

Contrastive Semi-supervised Learning for Underwater Image Restoration via Reliable Bank Supplementary Material

Shirui Huang^{1*}, Keyan Wang^{1*†}, Huan Liu², Jun Chen², Yunsong Li¹

¹ Xidian University, ² McMaster University, * Equal Contribution, † Corresponding Author
srhuang@stu.xidian.edu.cn, {kywang, ysli}@mail.xidian.edu.cn
{liuh127, chenjun}@mcmaster.ca

In the supplementary material, we first introduce the overall structure of our AIM-Net, details of validation set and loss functions, and then analyze the influence of the optimizer and NR-IQA metrics during training. Meanwhile, inherent advantage of Semi-UIR and additional experimental results on non-reference benchmarks are provided as well.

A. Details of the Network Structure

The goal of underwater image restoration is to conquer the problems such as low contrast, color distortion and blur. To this end, we propose an Asymmetric Illumination-aware Multi-scale Network (AIM-Net) to restore underwater images, as illustrated in Fig. 1. AIM-Net contains two important branches, namely *illumination-aware restoration branch* and *gradient branch*. The restoration branch adopts a multi-scale structure containing three parallel streams to exact precise spatial information and abundant contextual information in different scales. In addition, illumination prior is incorporated into the restoration branch to enhance the perception capability of color and light source. The simple gradient branch aims to recover high quality gradient map. The recovered gradient features are merged into the restoration branch to provide edge and structure prior.

The entire pipeline is as follows. The degraded image x passes through a convolution layer to extract preliminary features. After that, the preliminary features are fed into the illumination-aware restoration branch to extract and integrate precise spatial information and contextual information. The illumination map x_L is converted to illumination feature x'_L after a convolution layer. Then x'_L participates in the restoration branch via illumination guidance block (IGB) [8]. On the other side, x is fed into the gradient branch to restore fine gradient information. The gradient branch incorporates intermediate representations from the restoration branch because the intermediate representations carry rich structure information and are helpful for gradient map recovery. Later, attention feature fusion (AFF) block [3] fuses the fine features obtained by the two branches. Finally, the restored clear image x_{out} and restored gradient map g_{out} are both obtained after a convolution layer.

Illumination-aware Restoration Branch The task of the restoration branch is to reconstruct overall structure and color information of a degraded image. To complete the task, we introduce a parallel multi-scale structure incorporating illumination prior. The multi-scale structure can effectively integrate local details from different resolutions to maintain edge features and suppress halo artifacts [18]. Therefore, we adopt the structure of Multiscale Residual Block (MRB) [23] as the backbone. MRB contains three parallel streams of different scales. Each stream utilizes residual contextual block (RCB) to distill useful spatial information and selective kernel feature fusion (SKFF) block to integrate contextual information. Based on the structure of MRB, we add several fundamental modules to further enhance feature extraction capability, including: multi-dilated-convolution block (MDB), IGB, non-local spatial attention (NLSA) block. Meanwhile, we replace the SKFF block with AFF block to better aggregate contextual information. These modules are described in detail as follows.

MDB works at the front of each stream to exact spatial features. It expands the receptive field via combing four dilated convolutions. Dilation rates are set as 1, 2, 3 and 4, consistent with [2]. RCB [23] distills useful contextual features by modeling and transforming the inter-channel dependencies via attention mechanism.

In the highest resolution stream, we introduce IGB between MDB and RCB to incorporate illumination prior. As is noted in [22], the illumination map reflects underwater light field information including ambient light and scene-dependent degradation that are essential for underwater image formation. As a result, we introduce the illumination map to help the

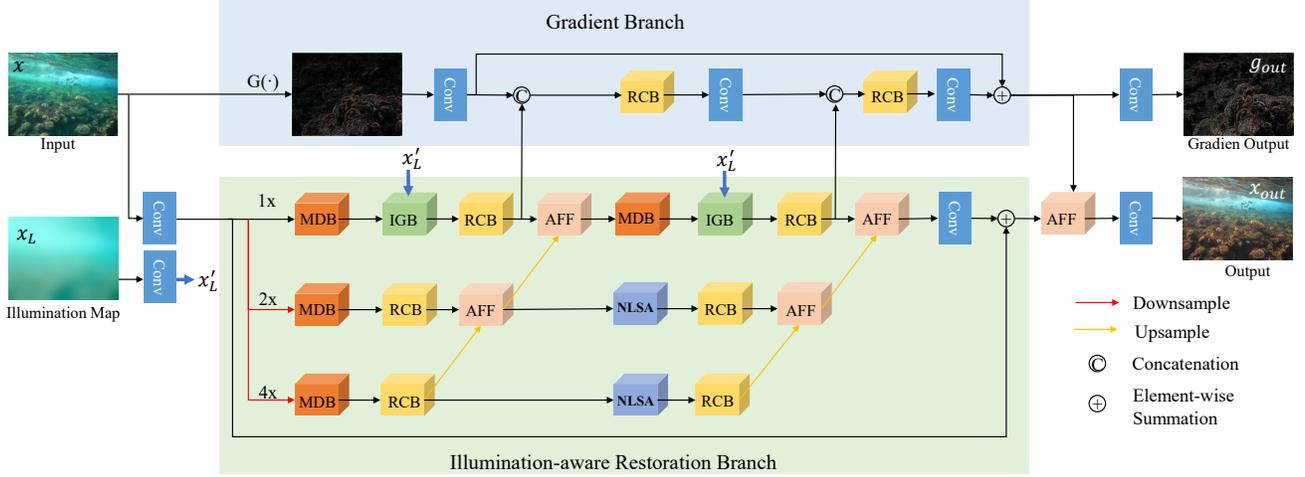


Figure 1. An overview of the proposed Asymmetric Illumination-aware Multi-scale Network (AIM-Net). AIM-Net mainly contains an illumination-aware restoration branch and a gradient branch.

network perceive color and light source information. The illumination map is estimated in accordance with [22]. IGB consists of a spatial feature transformation layer [20] and a deformable convolution [25]. In this way, our network can adapt to different color and light degradation types of underwater images.

To better integrate features from different scales, we replace SKFF module with AFF module. SKFF captures channel-wise dependencies via global-scale channel attention, whereas AFF squeezes local and global information into channel attention because local information is helpful to highlight local small targets [3]. The AFF module follows the RCB to aggregate local and global contextual features. Furthermore, we add NLSA module [14] in the middle of the two lower resolution streams. NLSA models long-range feature correlations and enjoys robustness from sparse representation. Compared with the standard non-local attention [19], it reduces computation expenses significantly.

With the help of the above well-designed modules, the network is able to extract fine feature representations. These feature representations are important for texture and color reconstruction.

Gradient Branch Image gradient map contains rich edge information and can guide the network to focus on local regions with sharp edges. As a result, we adopt the gradient map to enhance the edges of a restored image. Similar to [13], the image gradient prior is introduced to promote the restoration of underwater images from two aspects: 1) a gradient branch to restore a high-quality gradient map and provide structural information for the restoration branch; 2) a supplementary gradient loss to constrain the second-order relationship of adjacent pixels, guiding the underwater image restoration to focus more on the geometry. The gradient branch first estimates a coarse gradient map of x via gradient operation $G(\cdot)$ [13], and then enhances the gradient map. Thanks to the intermediate representations from the restoration branch, the gradient branch can recover a fine gradient map with a very simple structure only including three convolution layers and two RCBs. Gradient loss is detailed in Eq. (2). Under these two types of guidance, the structure features can be better preserved, and the restoration results with sharper edges, higher perceptual quality and less geometric-inconsistent textures can be obtained.

In a word, AIM-net takes a degraded underwater image x and its corresponding illumination map x_L as input and outputs a restored clear image x_{out} and a restored gradient map g_{out} :

$$x_{out}, g_{out} = f_{\theta}(x, x_L), \quad (1)$$

where f_{θ} represents the AIM-net parameterized by θ . AIM-Net has 1.675M parameters, and its inference speed is 33.3 FPS on the images with a resolution of 256×256 .

B. Details of Validation Set and Loss Functions

Our validation set, independent of testS and testR, contains 200 pairs of full reference underwater images from [10, 11] with ratio 12:8.

The detailed perceptual loss and gradient penalty are as follows:

$$\begin{aligned}
 L_{per} &= \sum_{j=1}^{K'} \sum_{i=1}^N |\varphi'_j(x_{out_i}) - \varphi'_j(y_i^l)| \\
 L_{grad} &= \sum_{i=1}^N |g_{out_i} - G(y_i^l)|,
 \end{aligned}
 \tag{2}$$

where L_{per} denotes perpetual loss based on pretrained VGG-16 [17] network and L_{grad} denotes gradient loss. φ'_j refer to ReLU1-2, ReLU2-2, and ReLU3-3 layers of the VGG-16 model. y^l denotes clear ground truth. $G(\cdot)$ stands for the operation to extract a gradient map [13]. L_{grad} constrains the restored gradient map of AIM-Net to approach the ground truth’s gradient map.

C. Influence of Optimizer and NR-IQA Metrics during Training

Table 1 shows the total training epochs required by Adam and AdamP to reach a similar accuracy. It can be observed that the training time required by Adam is longer than that by AdamP. Thus we choose AdamP as training optimizer.

Table 1. Evaluation the influence of adopting different optimizers on testR in terms of PSNR and SSIM.

Optim	PSNR	SSIM	Epochs
Adam	24.48	0.902	240
AdamP	24.59	0.901	200

To compare the efficiency of NR-IQA metrics, we additionally show the performance of adopting all seven NR-IQA metrics in Table 2. It is easy to check that using MUSIQ can achieve the best performance. Moreover, Fig. 2 presents some examples of pseudo labels (images in the reliable bank) selected by seven NR-IQA methods in the training process. It can be observed that MUSIQ can help select more visually pleasing pseudo labels (rightmost) over other metrics.



Figure 2. Examples of pseudo labels selected by NR-IQA metrics during training.

D. Inherent Advantage of Semi-UIR

In order to verify the inherent advantage of our proposed semi-supervised restoration framework Semi-UIR, we replace the AIM-Net with 5-layer Unet used by FUnIE-GAN [6], and compare the performance using and not using Semi-UIR. Training details and datasets are unchanged. The quantitative results are shown in Table 3. It’s obvious that our semi-supervised framework Semi-UIR is beneficial to improve the generalizability of general model like Unet on real-world underwater benchmarks. It also demonstrates that Semi-UIR is extensible.

E. Additional Experimental Results on Non-reference Benchmark

In Fig. 3-5, we present more results of our Semi-UIR on non-reference benchmarks Seathru [1], RUIE [12], UIEB [11] and EUVP [6], and compare with the state-of-the-art methods including GDCP [16], MMLE [24], WaterNet [11], Ucolor [9], FUnIE-GAN [6], PRWNet [5] and CWR [4]. Our proposed Semi-UIR outperforms other algorithms in restoring underwater images with rich details and natural color.

Table 2. Evaluation the influence of adopting different NR-IQA metrics on testS [10] and testR [11] in terms of PSNR and SSIM.

	NIQE	NIMA	UCIQE	BRISQUE	UIQM	PAQ2PIQ	MUSIQ
Reliability	13.45%	41.05%	48.16%	48.69%	76.87%	82.11%	91.21%
testS	22.83/0.811	23.01/0.815	22.90/0.813	23.15/0.820	23.24/0.820	23.08/0.818	23.40/0.821
testR	22.98/0.887	23.88/0.888	23.64/0.890	24.00/0.900	23.80/0.897	24.28/0.893	24.59/0.901

Table 3. Evaluations on non-reference benchmarks UIEB [11], EUVP [6], RUIE [12] and Seathru [1] in terms of UIQM [15], UCIQE [21] and MUSIQ [7]. Unet-base refers to training Unet without semi-supervised learning and unlabeled data. Unet-semi denotes training Unet with our proposed Semi-UIR.

Method	UIQM (higher, better)				UCIQE (higher, better)				MUSIQ (higher, better)			
	UIEB	EUVP	RUIE	Seathru	UIEB	EUVP	RUIE	Seathru	UIEB	EUVP	RUIE	Seathru
Unet-base	4.215	4.442	4.529	4.970	0.585	0.583	0.554	0.603	39.33	43.87	31.17	62.92
Unet-semi	4.329	4.512	4.763	5.037	0.586	0.597	0.570	0.620	41.06	49.11	34.41	64.13

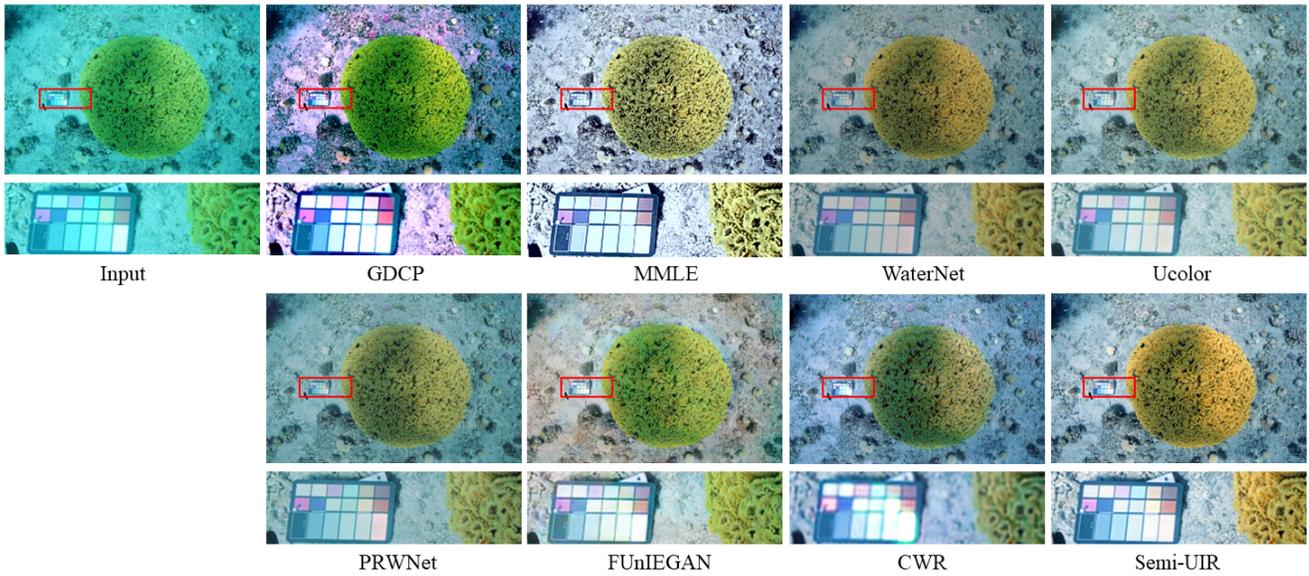


Figure 3. Visual comparisons on non-reference benchmark Seathru [1].

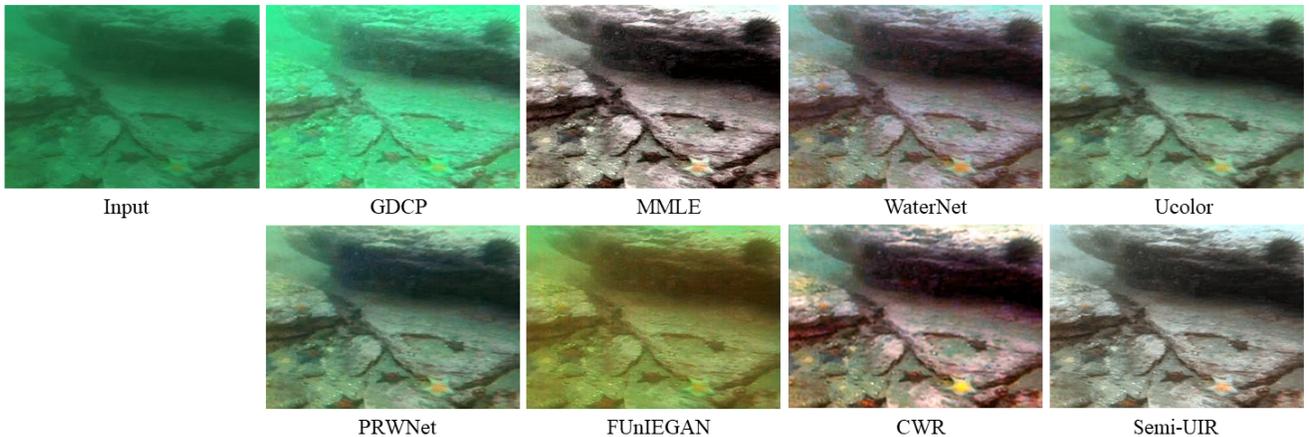


Figure 4. Visual comparisons on non-reference benchmark RUIE [12].

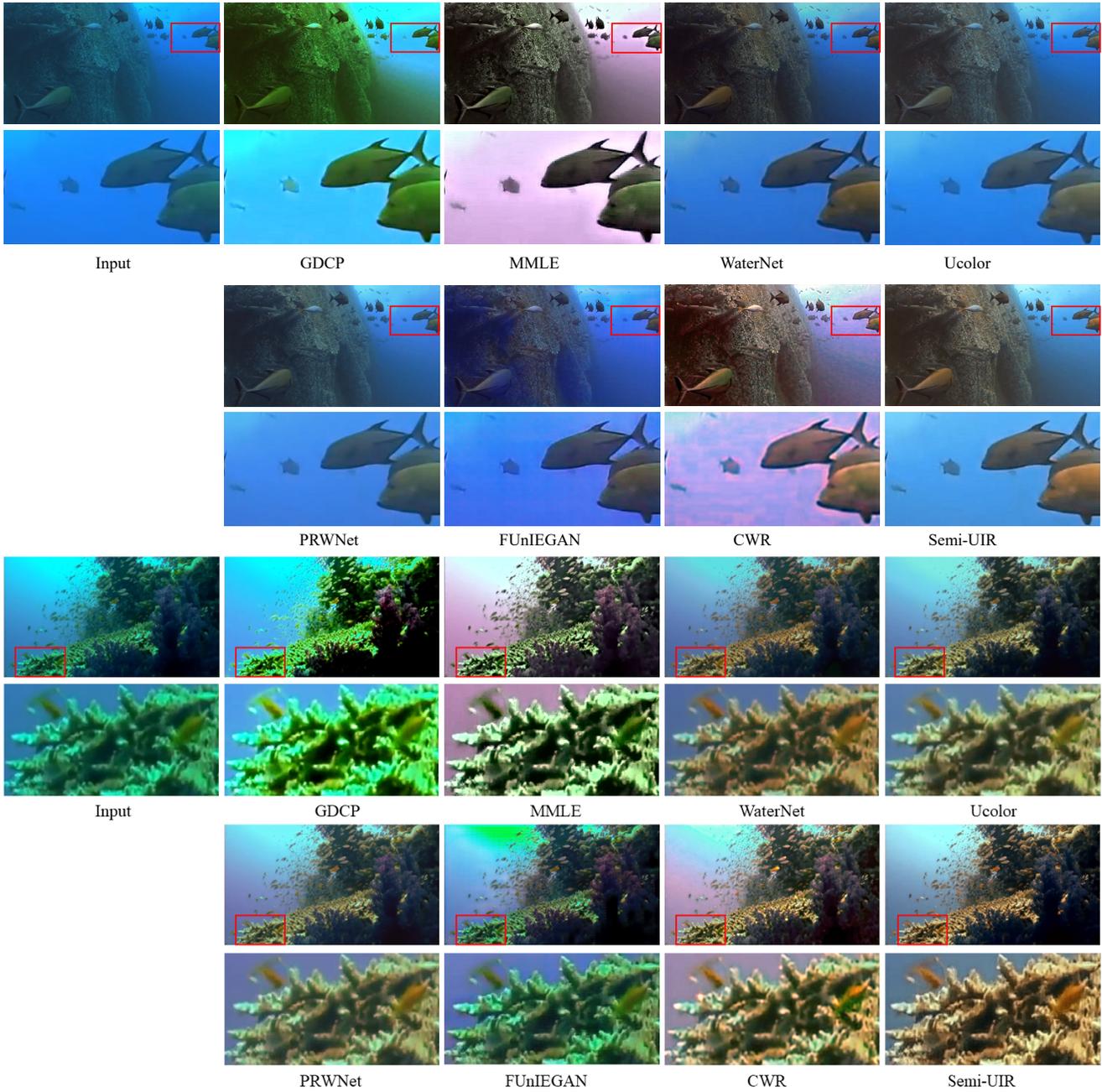


Figure 5. Visual comparisons on non-reference benchmark UIEB [11].

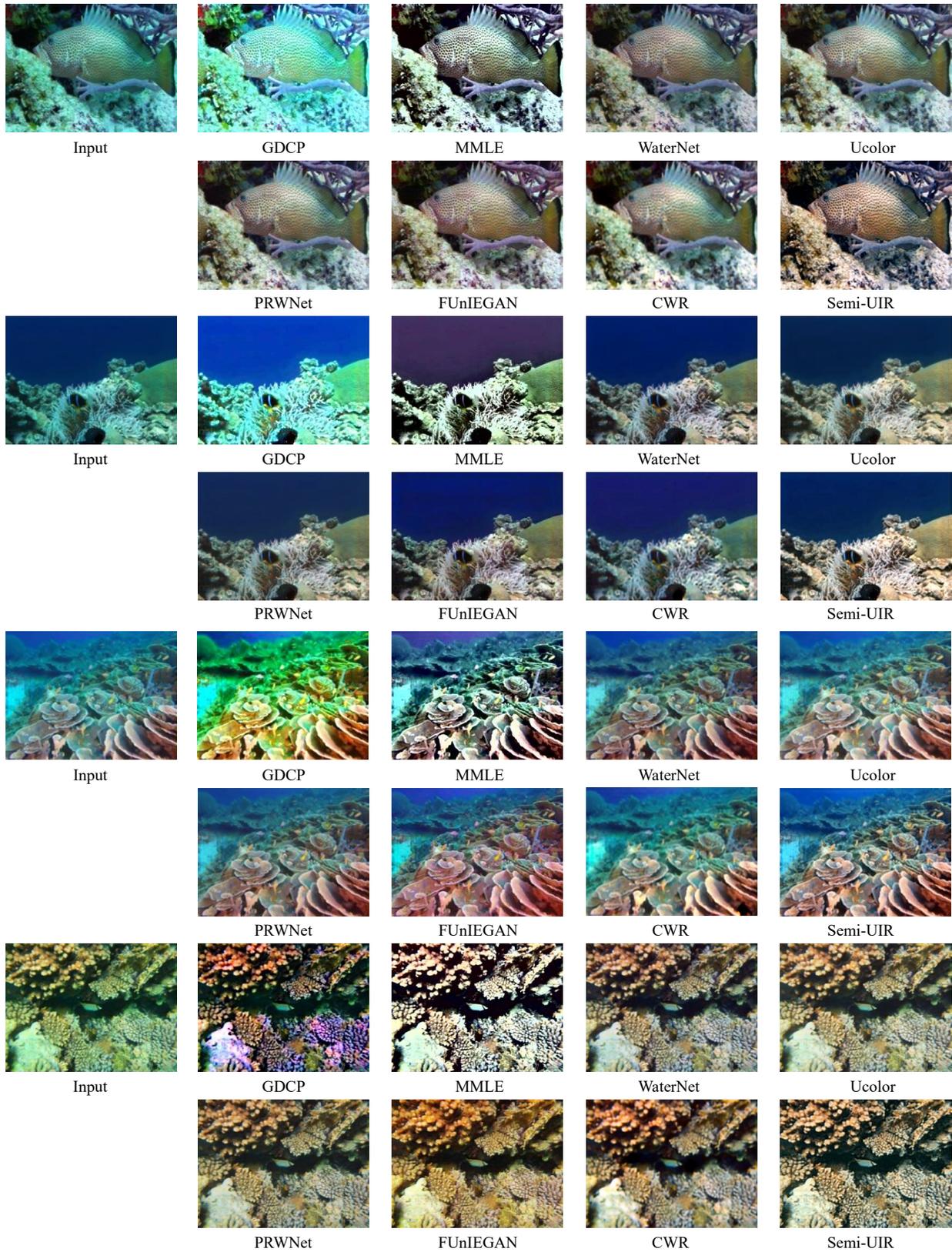


Figure 6. Visual comparisons on non-reference benchmark EUVP [6].

References

- [1] Derya Akkaynak and Tali Treibitz. Sea-thru: A method for removing water from underwater images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1682–1691, 2019. [3](#), [4](#)
- [2] Stephan Brehm, Sebastian Scherer, and Rainer Lienhart. High-resolution dual-stage multi-level feature aggregation for single image and video deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 458–459, 2020. [1](#)
- [3] Yimian Dai, Fabian Gieseke, Stefan Oehmcke, Yiquan Wu, and Kobus Barnard. Attentional feature fusion. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3560–3569, 2021. [1](#), [2](#)
- [4] Junlin Han, Mehrdad Shoeiby, Tim Malthus, Elizabeth Botha, Janet Anstee, Saeed Anwar, Ran Wei, Mohammad Ali Armin, Hongdong Li, and Lars Petersson. Underwater image restoration via contrastive learning and a real-world dataset. *Remote Sensing*, 14(17):4297, 2022. [3](#)
- [5] 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. [3](#)
- [6] Md Jahidul Islam, Youya Xia, and Junaed Sattar. Fast underwater image enhancement for improved visual perception. *IEEE Robotics and Automation Letters*, 5(2):3227–3234, 2020. [3](#), [4](#), [6](#)
- [7] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5148–5157, 2021. [4](#)
- [8] Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-in-one image restoration for unknown corruption. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17452–17462, 2022. [1](#)
- [9] Chongyi Li, Saeed Anwar, Junhui Hou, Runmin Cong, Chunle Guo, and Wenqi Ren. Underwater image enhancement via medium transmission-guided multi-color space embedding. *IEEE Transactions on Image Processing*, 30:4985–5000, 2021. [3](#)
- [10] Chongyi Li, Saeed Anwar, and Fatih Porikli. Underwater scene prior inspired deep underwater image and video enhancement. *Pattern Recognition*, 98:107038, 2020. [2](#), [4](#)
- [11] 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, 2020. [2](#), [3](#), [4](#), [5](#)
- [12] Risheng Liu, Xin Fan, Ming Zhu, Minjun Hou, and Zhongxuan Luo. Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(12):4861–4875, 2020. [3](#), [4](#)
- [13] Cheng Ma, Yongming Rao, Jiwen Lu, and Jie Zhou. Structure-preserving image super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021. [2](#), [3](#)
- [14] Yiqun Mei, Yuchen Fan, and Yuqian Zhou. Image super-resolution with non-local sparse attention. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3517–3526, 2021. [2](#)
- [15] Karen Panetta, Chen Gao, and Sos Agaian. Human-visual-system-inspired underwater image quality measures. *IEEE Journal of Oceanic Engineering*, 41(3):541–551, 2015. [4](#)
- [16] 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. [3](#)
- [17] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. [3](#)
- [18] Keyan Wang, Yan Hu, Jun Chen, Xianyun Wu, Xi Zhao, and Yunsong Li. Underwater image restoration based on a parallel convolutional neural network. *Remote sensing*, 11(13):1591, 2019. [1](#)
- [19] Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He. Non-local neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7794–7803, 2018. [2](#)
- [20] Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. Recovering realistic texture in image super-resolution by deep spatial feature transform. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 606–615, 2018. [2](#)
- [21] Miao Yang and Arcot Sowmya. An underwater color image quality evaluation metric. *IEEE Transactions on Image Processing*, 24(12):6062–6071, 2015. [4](#)
- [22] Tian Ye, Sixiang Chen, Yun Liu, Yi Ye, Erkang Chen, and Yuche Li. Underwater light field retention: Neural rendering for underwater imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 488–497, June 2022. [1](#), [2](#)
- [23] Syed Waqas Zamir, Aditya Arora, Salman Hameed Khan, Hayat Munawar, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for fast image restoration and enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1, 2022. [1](#)
- [24] 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. [3](#)
- [25] Xizhou Zhu, Han Hu, Stephen Lin, and Jifeng Dai. Deformable convnets v2: More deformable, better results. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9308–9316, 2019. [2](#)