RAUNE-Net: A Residual and Attention-Driven Underwater Image Enhancement Method

Wangzhen Peng[1], Chenghao Zhou[2], Runze Hu[3], Jingchao Cao[1], and Yutao Liu[1,*]

Department of Computer Science and Technology, Ocean University of China, Qingdao, China
School of Management Engineering, Qingdao University of Technology, China
School of Information and Electronics, Beijing Institute of Technology, Beijing 100080, China
*Email: *liuyutao@ouc.edu.cn*

Abstract

Underwater image enhancement (UIE) poses challenges due to distinctive properties of the underwater environment, including low contrast, high turbidity, visual blurriness, and color distortion. In recent years, the application of deep learning has quietly revolutionized various areas of scientific research, including UIE. However, existing deep learning-based UIE methods generally suffer from issues of weak robustness and limited adaptability. In this paper, inspired by residual and attention mechanisms, we propose a more reliable and reasonable UIE network called ***RAUNE-Net*** by employing residual learning of high-level features at the network's bottle-neck and two aspects of attention manipulations in the down-sampling procedure. Furthermore, we collect and create two datasets specifically designed for evaluating UIE methods, which contains different types of underwater distortions and degradations. The experimental validation demonstrates that our method obtains promising objective performance and consistent visual results across various real-world underwater images compared to other eight UIE methods. Our example code and datasets are publicly available at https://github.com/fansuregrin/RAUNE-Net.

Proposed Method

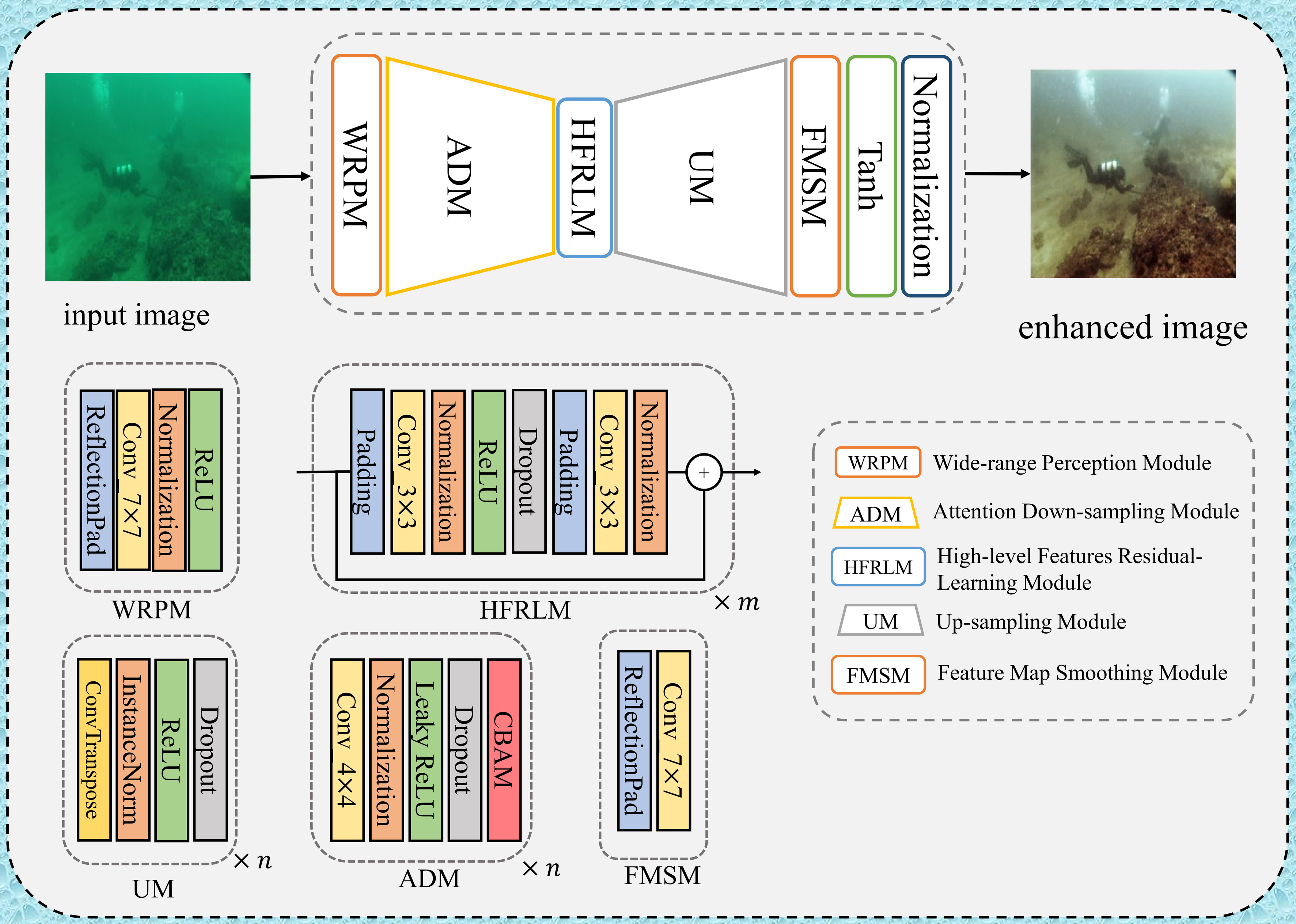

Fig. 1 The data-flow and architecture of RAUNE-Net. the proposed network consists of a wide-range perception module (WRPM), an attention down-sampling module (ADM), a high-level features residual-learning module (HFRLM), an up-sampling module (UM), and a feature map smoothing module (FMSM) followed by a Tanh activation layer.

A. Pixel Content Loss

The pixel content loss is designed to measure the discrepancy in pixel values between the output image Y_e generated by the model and the reference image Y_{ref} across the red, green, and blue color channels. The mathematical formula is defined as:

$$\mathcal{L}_{\text{pcont}} = \mathbb{E}\left[\left\|Y_e - Y_{ref}\right\|_1\right].$$

Acknowledgment

This work was supported by the National Science Foundation of China under grant 62201538 and Natural Science Foundation of Shandong Province under grant ZR2022QF006.

B. Structural Similarity Loss

The main purpose of incorporating SSIM metric into the loss function is to encourage the network to generate enhanced images that are closer to the ground truth in luminance, contrast, and structural information. The expression for SSIM Loss is:

$$\mathcal{L}_{\text{ssim}} = \mathbb{E}\left[\frac{1-\text{SSIM}(Y_e - Y_{ref})}{2}\right],$$

where Y_e are enhanced images, are Y_{ref} reference images, and the $SSIM\ (*,*)$ represents structural similarity index map between two images.

B. Semantic Content Loss

We consider using high-level features obtained from last several layers of a classification network to evaluate the semantic discrepancy between the enhanced image and the reference image. We extract feature maps obtained from the last five convolutional layers for both the enhanced image and the reference image. Finally, we calculate the weighted sum of the L1 distances between each pair of feature maps, which serves as the semantic content loss. Mathematically, this loss can be written as:

$$\mathcal{L}_{scont} = \sum_{i=1}^{5} k_i \cdot \ell_1\left(\Omega_i(Y_e), \Omega_i(Y_{ref})\right),$$

where k_i is the i-th weight for summing, $\ell_1(*,*)$ is the L1 loss and the $\Omega_i(*)$ is the last i-th convolutional layer in *VGG19_BN*.

Results

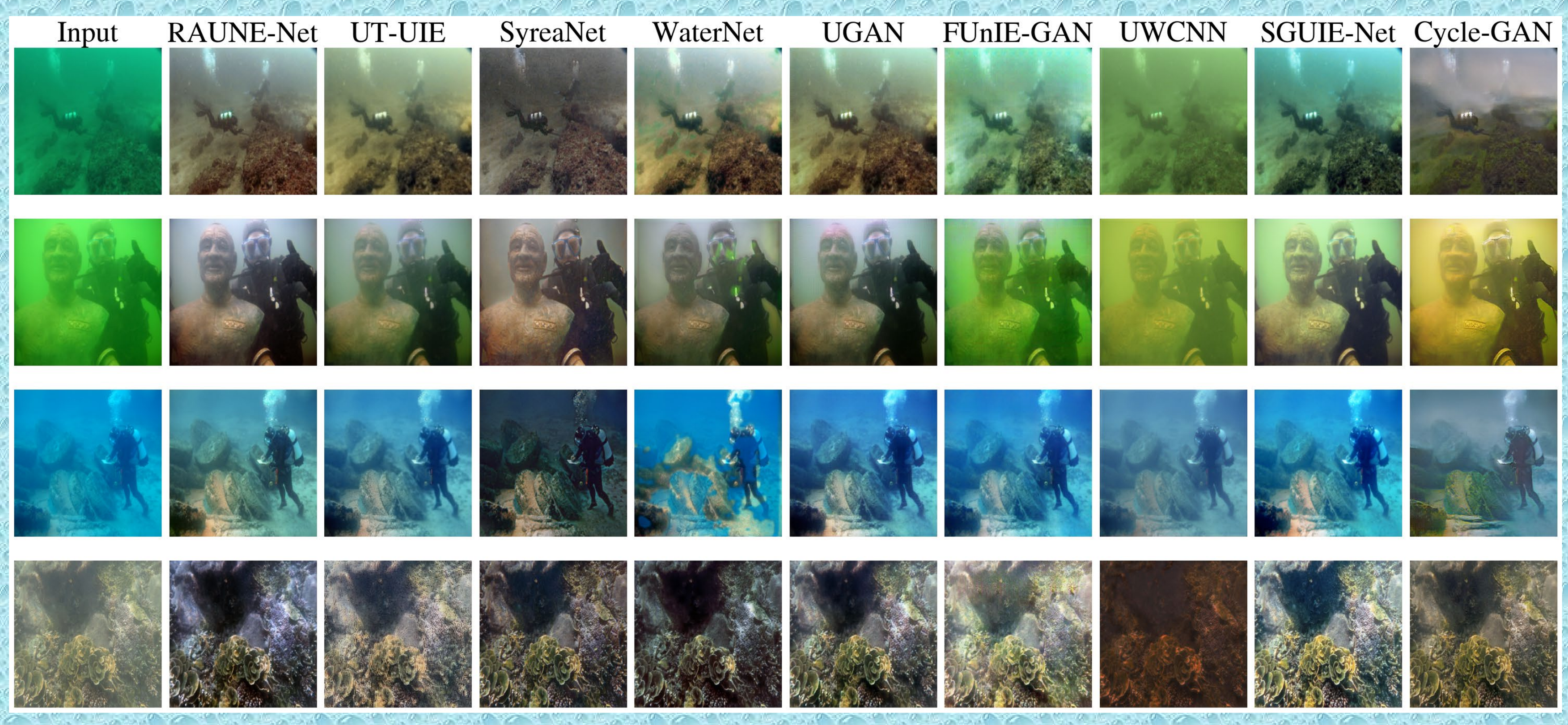

Fig. 2 Subjective evaluation results of different methods on *U45*

Table 1. Objective evaluation results of different UIE methods

Methods	LSUI400		EUVP_Test515		UIEB100		OceanEx	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
UT-UIE [21]	24.351	0.829	25.214	0.813	20.916	0.764	21.270	0.822
SyreaNet [29]	18.050	0.766	17.721	0.743	16.501	0.836	20.243	0.865
WaterNet [8]	26.688	0.874	25.285	0.833	22.279	0.868	22.132	0.887
UGAN [3]	25.117	0.846	23.636	0.805	21.368	0.825	22.436	0.822
FUnIE-GAN [6]	23.272	0.818	24.077	0.794	19.614	0.813	20.448	0.855
UWCNN [7]	17.366	0.725	17.725	0.704	14.155	0.686	15.960	0.724
SGUIE-Net [22]	19.910	0.819	19.187	0.760	21.178	0.872	18.677	0.834
Cycle-GAN [9]	18.320	0.749	17.963	0.709	17.714	0.758	21.007	0.828
RAUNE-Net	26.812	0.876	26.331	0.845	22.751	0.879	22.728	0.876

As can be observed from Figure 2, our network is capable of enhancing bluish, greenish and hazy underwater images, while the other methods have exhibited improper or insufficient handling of these images. From the table 1, our method achieves highest PSNR values and SSIM values.

Conclusion

- We incorporate residual and attention mechanisms into UIE to construct an effective and efficient network called RAUNE-Net, which is capable of handling a wide range of underwater scenes comprehensively.
- Our method achieves remarkable objective performance and consistent subjective results on both referenced and non-referenced real-world underwater image sets.