A Efficient Learning-based Framework for Underwater Image Enhancement

ZhangKai Lv, Long Chen, and Dandan Ding
School of Information Science and Engineering, Hangzhou Normal University
Hangzhou, China, 311121
Email: DandanDing@hznu.edu.cn

***Abstract*—Underwater image enhancement, which targets removing the noise, scattering effect, as well as the bluish, greenish, or yellowish tone from the underwater images, has attracted much attention. Earlier rule-based methods generally assume some prior knowledge when performing constrained optimization. Recently, the learning-based methods have exhibited superior performance over the rule-based methods. This paper proposes a simple yet effective convolution neural network (CNN)-based framework to enhance underwater images. Our method consists of two stages, CNN-based enhancement and YUV-based post-processing. In the first stage, a lightweight CNN network is developed to extract the latent features from the input images for enhancement. Specifically, our CNN cascades three residual groups which utilize channel attention blocks to strengthen the feature extraction capability. In the second stage, the output of the CNN model is transformed to the YUV color space where the luminance component is further corrected, improving the brightness of the whole image. Our model is trained using the UIEB data set and tested on 100 underwater images. Experimental results show that our method outperforms state-of-the-art methods both qualitatively and quantitatively. Moreover, in terms of computational complexity, our CNN model costs 437k parameters, which is only 39.5% and 22.5% of the existing CNN-based methods WaterNet and UWGAN.**

***Index Terms*—CNN, Underwater image enhancement, Channel attention, Color space**

I. Introduction

With the widespread aggravation of land resource shortage, population expansion, and environmental degradation, which pose potential hazards to the ocean——the crucial core of the life-support system and the development of human society. Consequently, various coastal countries have aroused their environmental awareness and accelerated the research, development, and utilization of the ocean. When exploiting marine resources, the vision system is deemed as one of the most effective detection methods for underwater robots. However, the complexity of imaging in water and the difference of law of light propagation cause inevitable degradation to the underwater images such as color cast, low light, and low contrast. Thus, underwater image enhancement with the aim of restoring visually pleasing and clear images has become a key issue in marine missions.

Earlier underwater image enhancement methods are *rule-based*. The non-physical methods, such as histogram equalization [1], auto white balance and fusion these two methods to enhance underwater images [2], directly adjust the dynamic value range and enhance the contrast by changing the gray value of the pixels in either spatial or frequency domain, without considering the physical procedure of underwater degradation. These methods could remove certain noise and enhance edge-details, whereas their performance is circumscribed due to the physical characteristics of underwater imaging are not taken into account.

Physical methods resolve such problems by abstracting and modelling the underwater degradation process of imaging. As such, undegraded images are accessible through inverting the degradation process. For instance, B.L. McGlamery [3] proposed the calculation model of underwater imaging system; Jules S. Jaffe [4] optimized an underwater imaging system; based on the imaging model, Nicholas c [5] proposed a method through eliminating light scattering in underwater images; He [6] proposed a single-image defogging method based on dark channel prior (DCP). Generally, these physical methods gain more superiority than the non-physical ones. However, they are incapable of removing noise and blurring artifacts from the underwater images, causing restrictions on the quality of the processed images. To address this issue, researchers introduced restoration techniques to obtain more realistic underwater images. Lu [7] proposed the color correction method based on spectral characteristics to restore distorted colors; Peng [8] proposed a depth estimation method based on blur and absorption (IBLA); Song [9] proposed a depth estimation method combined underwater light attenuation prior with linear regression model to estimate background light and transmitted light in RGB color space. Nevertheless, these methods of image enhancement are usually based on the premise of obtaining prior knowledge and information, thus can only be managed under specific scenarios.

Benefiting from the development of neural networks, *learning-based* methods gain extensive implementation in image defogging, deblurring, and restoration, etc. Research demonstrates that Convolutional Neural Network (CNN) could exert better performance than the traditional methods both objectively and subjectively. Inspired by the previous work, researchers introduced CNN to improve the quality enhancement of underwater images. Chen [10] proposed a scheme combined a filtering-based with a GAN-based restoration. Li [11] designed an end-to-end WaterGAN consisting of a depth estimation module followed by a color correction module. Wang [12] proposed UWGAN learning the nonlinear mapping between undistorted and distorted images. All above methods generate enhanced results by utilizing end-to-end and data-driven training mechanisms and obtaining better performance than rule-based methods. On the other hand, these CNN models are computationally expensive because they generally cost millions of parameters, which challenges practical applications.

Instead, this paper systematically proposes an *efficient, low complexity* framework for underwater image enhancement. Our framework consists of two stages, CNN-based enhancement and YUV-based post-processing. Specifically, we develop a simple yet effective network structure to extract the latent features of input low-quality underwater images. The network cascades three residual groups for feature enhancement and channel attention block is incorporated into each residual group. Afterwards, the output of the CNN models is transformed into YUV color space for luminance correction. Besides, we design a collaborative loss function containing perceive loss, MSE loss, gradient loss, and SSIM loss to preserve the texture features. Results show that our proposed framework outperforms state-of-the-art methods both quantitatively and qualitatively.

II. Related Work

A. Learning-based Underwater Image Enhancement

As an effective deep learning method, CNN has been widely adopted in the field of visual enhancement. To continuously promote, scientists also try to apply this method to the enhancement of underwater images.

Due to the arduousness of obtaining high-quality underwater reference images, Cameron. [13] proposed to use Cycle Generative Adversarial Network (CycleGAN) [14] to improve the quality of underwater visual scenes. Similarly, introducing gradient descent Loss to enhance the predictive ability of the generated network; Wang [12] proposed an unsupervised generative opposition network (UWGAN), which utilized aerial image, depth map pairs, and optimized imaging models to generate realistic underwater images. Specifically, the obtained comprehensive underwater data further used the U-Net network to directly reconstruct the underwater clear image, while maintaining the structural similarity of the scene content; Li [15] constructed an Underwater Image Enhancement Benchmark (UIEB) including 950 real world underwater images, 890 of which have the corresponding reference images, and propose an underwater image enhancement network (called Water-Net); Zhang [16] proposed an attention-based neural network to generate high-quality enhanced low-light images from raw sensor data; Li [17] proposed an underwater image enhancement with image colorfulness measure. This model encompasses a pre-processed non-parametric layer and an adaptively refined parameter layer, optimizing the enhancement network based on the joint of pixel level and feature level; Li [18] proposed a lightweight convolutional enhancement network (UWCNN) with good applicability. The method is based on the synthetic underwater image training set to directly reconstruct clear underwater images, which has a good real-time characteristic and can be easily extended to the frame-by-frame enhancement of the video.

Compared with the conventional methods, the CNN-based methods above could achieve notable performance. Nevertheless, their computational complexity is greatly increased owing to the larger cost of parameter and longer runtime than the rule-based methods. Some networks like UWCNN has fewer parameters than other methods like UWGAN and WaterNet, whereas its performance is also compromised.

B. Underwater Image Dataset

Deep learning is a data-driven learning method. FUniE-GAN opens source data set EUVP which is divided into three subsets. These real underwater images from seven different camera equipment are cropped part of images from the Youtube videos, and the ground truth is generated by a trained CycleGAN [19]. UFO120 [20] is a new dataset divided into training dataset and test benchmark dataset. It can be used for model training tasks for underwater image super-division. Li [15] provided paired dataset called UIEB where the original image comes from the real underwater image, and the ground truth is selected from a variety of traditional methods by comparing. The dataset includes 890 training pictures and 90 test pictures, and small pictures are very convenient to test and are commonly used. The author further used this dataset to design WaterNet for underwater image enhancement. In this paper, we perform our experiments using UIEB dataset.

III. Proposed Method

In this section, we will describe the framework of our proposed underwater image enhancement methods in details, and introduce the loss function employed in our work. Ultimately, we will present the post-processing employed, i.e., luminance correction in YUV color space.

A. Network Architecture

Overall Framework. As illustrated in Fig. 1, the method proposed consists of a single branch network and a post-processing operation. The input image is an unprocessed RGB underwater image of size $256\times256\times3$. Our network first uses $3\times3\times32$ and $3\times3\times64$ convolution kernels to perform flat convolution operations to extract shallow features, and then three residual groups are used to extract deeper features. In each residual group, three channel attention blocks are utilized to learn the distribution weights of different channels to strengthen useful features and suppress invalid features. In the CA block module, input data is squeezed into size of $1\times1\times$C data by average pooling, then this $1\times1\times$C descriptors are put through the convolutional layers and activated by sigmoid function. The output of the third residual group is $256\times256\times16$. After processing by the convolution kernel with a size of 3×3, an enhanced image is generated with a size of $256\times256\times3$.

Channel attention module With notable breakthroughs in both image and natural language processing, the Attention

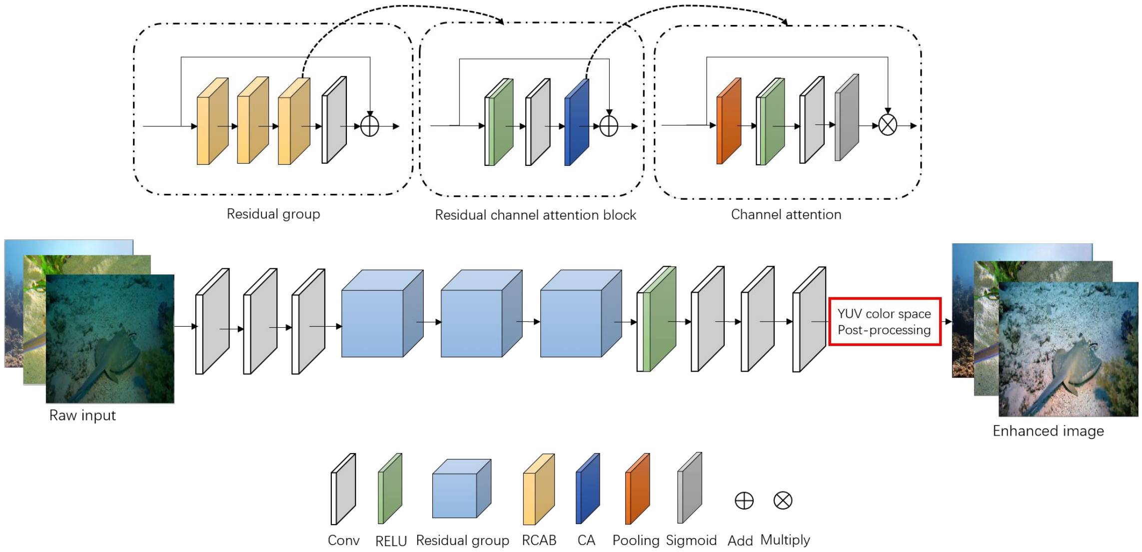


Fig. 1: Structure of our proposed network which concatenates three residual groups. Each residual group has three channel attention blocks for latent feature extraction and texture preservation.

mechanism has been substantiated in improving network performance. Drawing on the human brain and the human eye perception mechanism, the attention mechanism intelligently focuses attention on the local information of interest and neglects extraneous information as much as CNN could. The mechanism was proposed in 2018 by Jie Hu [21] and extensively utilised in deep learning scenarios.

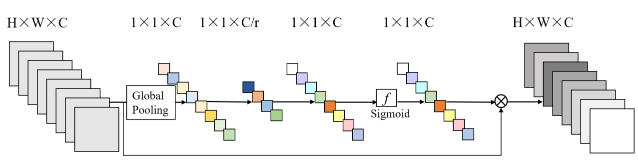


Fig. 2: Channel attention module used in our residual group.

As demonstrated in Fig. 2, H, W, and C are height, width, and the number of channels, respectively. r is the reduction factor. At first, the input of size H×W×C features is processed to generate channel weight of 1×1 for each channel through pooling operation. Next, through using the convolutional and sigmoid layers, the network learns the degree of dependence across channels to obtain the weight of each channel. Finally, the feature map of each channel with the corresponding weighted value will be multiplied together to generate the adjusted feature map of the same size as the input.

B. Loss Function

To ensure the generation of high-quality visual images, we adopt a collaborative loss function in our work: content perception loss, mean square error(MSE) loss, structural similarity (SSIM) loss, and gradient descent loss. The final loss function is presented as:

$$\begin{aligned} L = & w_{vgg} \times L_{vgg} + w_{mse} \times L_{mse} \\ & + w_{ssim} \times L_{ssim} + w_{gdl} \times L_{gdl}, \end{aligned} \tag{1}$$

where w indicates the weighting factor.

L_{vgg} is a content loss item which drives the network to get training results with similar content. The ReLU activation layer based on the pre-trained 19-layer VGG network defines L_{vgg} as:

$$L_{vgg}(E,G) = \frac{1}{C_j H_j W_j} \sum_{i=1}^{N} ||\varphi_j(E^i) - \varphi_j(G^i)||, \tag{2}$$

where E is the enhanced image and G is the reference image. N denotes batch size, and C_j, H_j, W_j indicate the number, height, and width of feature maps, respectively.

L_{mse} is the MSE loss between the enhanced image and the reference image. Let n and m represent the width and height of the image, respectively. We use L_{mse} to calculate the average pixel-wise distance on the enhanced image and the ground truth. It is written as:

$$L_{mse} = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} ||E(i,j) - G(i,j)||^2. \tag{3}$$

L_{ssim} denotes the SSIM loss between enhanced image and reference image:

$$SSIM(E,G) = \frac{(2\mu_E \mu_G + c_1)(\sigma_{EG} + c_2)}{(\mu_E^2 + \mu_G^2 + c_1)(\sigma_E^2 + \sigma_G^2 + c_2)} \tag{4}$$

$$L_{ssim}(E,G) = 1 - SSIM(E,G) \tag{5}$$

where μ_E, μ_G represents the average of image E and G, respectively, σ_E, σ_G means standard deviation of E and G, σ_{EG} means the covariance of E and G, c_1 and c_2 are constant to avoid situations where the denominator is 0.

L_{gdl} is the gradient descent loss which introduces enhanced image, reference image, and correlation between adjacent pixels. α means an integer greater than or equal to 1, in our experiment, we set α to 1. The relevant formula is expressed as follows:

$$L_{gdl} = \sum_{i,j} ||G_{i,j} - G_{i-1,j}| - |E_{i,j} - E_{i-1,j}||^{\alpha} + \sum_{i,j} ||G_{i,j} - G_{i,j-1}| - |E_{i,j} - E_{i,j-1}||^{\alpha}. \tag{6}$$

C. Post-processing in YUV Color Space

Our proposed method adds a post-processing module to improve the overall brightness of the image generated by the network. To elaborate further, we convert the original image from RGB space to YUV space. YUV includes three components Y, U, and V. Y stands for luma, U and V stand for chroma respectively. Then we perform maximum and minimum normalization operation on the Y component and convert the original data linearization method to the range of [0, 1]. The corresponding formula is as follows, MAX and MIN in the formula respectively represent the maximum and minimum of the Y channel component:

$$Y' = 255.0 \times \frac{Y - MIN}{MAX - MIN}. \tag{7}$$

IV. Experimental Results

A. Implementation Details

The dataset of our project is mainly based on UIEB's 890 high-definition underwater images with corresponding reference images. Specifically, 800 are randomly selected as the training set, while the remaining 90 images along with 10 images randomly selected from the UFO120 dataset [20] constitute the test set. We resized the original picture and cropped them to size 256×256 for better training and test. We train the model use Adam, set the segmented learning rate from 0.001, and set the batch size to 16 and epoch to 500. We use Tensorflow as the deep learning framework on an Inter(R)i7-9700k CPU@3.60GHz, 32GB RAM, and a NVIDIA GeForce RTX 2080Ti GPU. The total training time is around 10 hours.

B. Evaluation Metrics

We employ Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) to measure the image quality, and use Underwater Image Quality Measure (UIQM) [22] to evaluate non-reference underwater quality assessments. The definition of PSNR can be expressed as:

$$PSNR = 10 \times \log_{10}(\frac{Max_i^2}{MSE}), \tag{8}$$

where Max_i^2 means the max pixel value of image provided ranging from 0 to 255 for uint8 data type. For any picture, PSNR reflects the degree of distortion of the image, and SSIM objectively reflects how much the image is structural similarity on scale. For the particularity of underwater image quality assessment, the predecessors proposed UIQM to evaluate the overall underwater image quality, which can intuitively reflect image color, sharpness and contrast. UIQM is a reference-free index based on human visual system excitation. Aimed at the degradation of underwater images, the UIQM is expressed as a linear combination of three parts: Color Measurement index (UICM), Degree Measurement Index (UISM), and Contrast Measurement Index (UIConM). The calculation of UIQM is as follows:

$$UIQM = c_1 \times UICM + c_2 \times UISM + c_3 \times UIconM, \tag{9}$$

where c_1, c_2, and c_3 indicate the weighting factors.

$$UICM = -0.0268 \times \sqrt{\mu_{\alpha,RG}^2 + \mu_{\alpha,YB}^2} + 0.1586 \times \sqrt{\sigma_{\sigma,RG}^2 + \sigma_{\alpha,YB}^2} \tag{10}$$

$$UISM = \Sigma_{c=1}^{3} \lambda_c EME(grayscale\, edge_c) \tag{11}$$

$$EME = \frac{2}{k_1 k_2} \Sigma_{l=1}^{k_1} \Sigma_{k=1}^{k_2} \log(\frac{I_{max,k,l}}{I_{min,k,l}}) \tag{12}$$

$$UIConM = \log AMEE(Intensity) = \frac{1}{k_1 k_2} \bigotimes \sigma_{l=1}^{k_1} \sigma_{k=1}^{k_2} (\frac{I_{max,k,l} \Theta I_{min,k,l}}{I_{max,k,l} \bigoplus I_{min,k,l}}) \tag{13}$$

In UICM calculation, $\mu_{\alpha,RG}$, and $\mu_{\alpha,YB}$ mean asymmetrical trimmed chroma intensity average in RG and YB. $\sigma_{\alpha,RG}^2$ and $\sigma_{\alpha,RG}^2$ mean the pixel variance in RG and YB.

In UISM calculation, grayscale edge representative gray edge map of each R, G, B, EME is a method of measuring edge sharpness and the values of the linear combination are λ_R=0.299, λ_G=0.587, λ_B=0.114. In EME, k_1, k_2 mean number of image divisions and $\frac{I_{max,k,l}}{I_{min,j,l}}$ means the relative contrast.

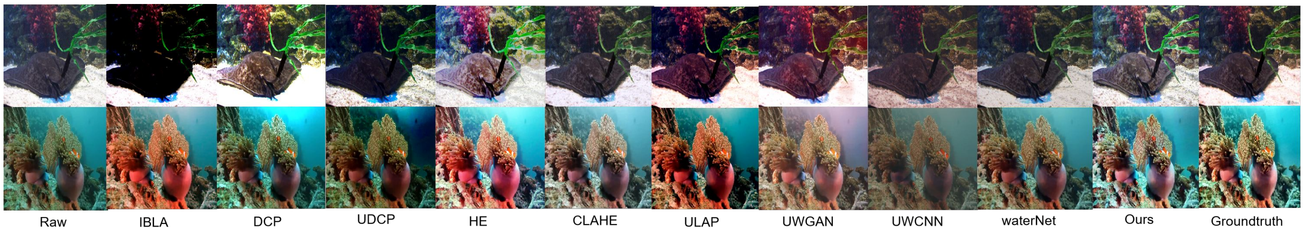
In UIConM, k_1, k_2 mean number of image divisions and $\bigotimes$, $\bigoplus$ and Θ mean the special operations on images.

C. quantitative comparison

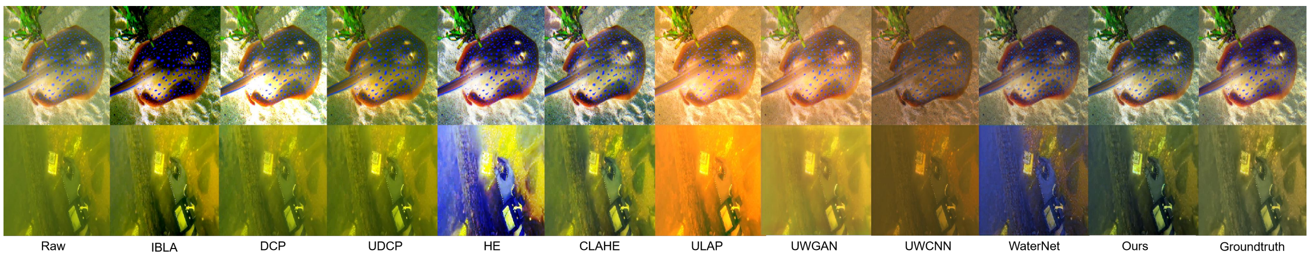
We compare our work with ten state-of-the-art methods, as shown in Table. I. Findings from this project indicate that our proposed method achieves the highest SSIM and PSNR of all, significantly improves the SSIM, PSNR, and UIQM metrics, e.g. For instance, the average SSIM, PSNR, and UIQM of recent WaterNet is 0.8000, 18.7366dB, 2.7165, whereas ours is as high as 0.8720, 21.9273, and 3.0522.

TABLE I: Performance comparison with state-of-the-art methods

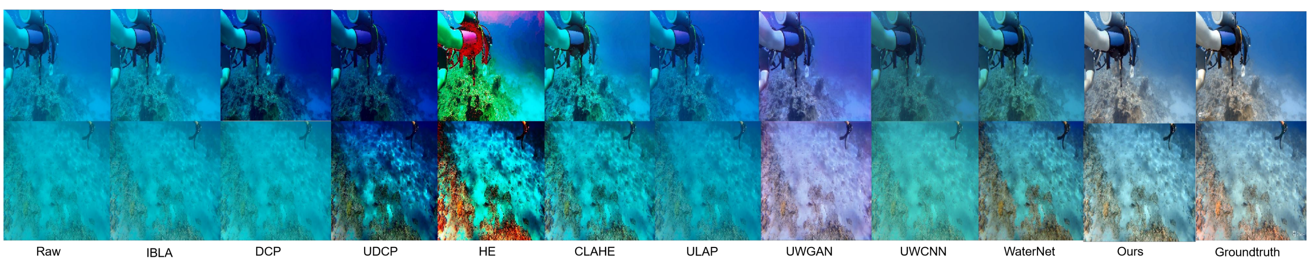
Method	SSIM	PSNR (dB)	UIQM
IBLA [8]	0.6672	15.5341	1.7402
DCP [6]	0.6657	13.5033	1.9462
UDCP [23]	0.5349	12.0763	1.7544
HE [24]	0.7453	15.6528	2.6197
CLAHE [25]	0.8218	18.2440	2.8704
ULAP [9]	0.6870	15.1370	1.7686
UWGAN [12]	0.7937	17.0210	**3.1170**
UWCNN [26]	0.7179	15.5202	2.8419
WaterNet [15]	0.8000	18.7366	2.7165
Ours	**0.8720**	**21.9273**	3.0522



(a) Underwater scenes with creatures

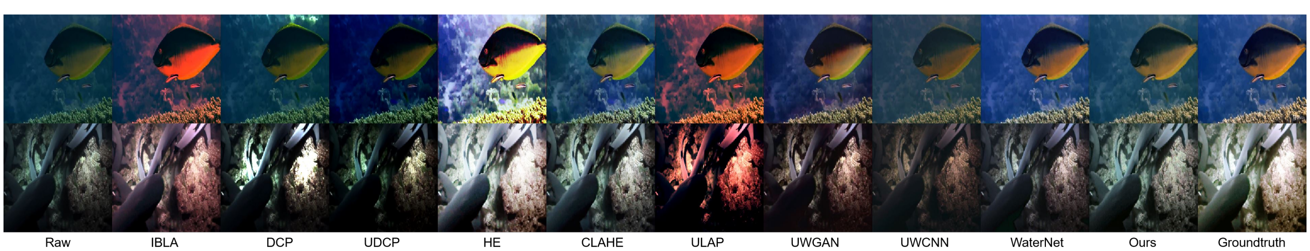


(b) Underwater scenes with yellowish tone



(c) Underwater scenes with bluenish tone

(d) Underwater scenes with greenish tone



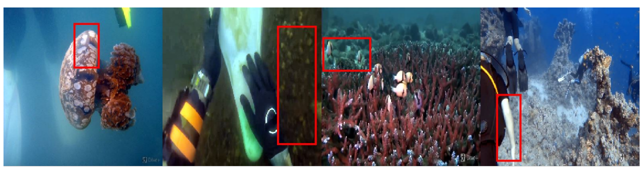
(e) Low-light illumination underwater scenes

Fig. 3: Visualization results of different methods in low-light illumination underwater scenes.

(a) Input Raw underwater images

(b) Underwater images enhanced without post-processing

(c) Underwater images enhanced with post-processing



(d) Ground truth underwater images

Fig. 4: Visual quality comparison of using and without using post-processing in our approach.

D. Qualitative Comparison

The enhanced images of different methods are visualized in Fig. 3 for comparison. In the figure, the last column provides the ground truth images for reference. There are several diverse underwater scenes such as (a) Scenes with creatures, (b) Yellowish scenes. (c) Blueish scenes. (d) Greenish scenes. (e) low-light illumination scenes. UW-GAN, UWCNN, and WaterNet are three state-of-the-art CNN-based methods. It is shown in the figure that UW-GAN fails in blurring removal, color cast correcting, and chroma and brightness restoring, e.g., in Fig. 3d, the visualization of UW-GAN is unclear and unnatural. UWCNN is a simple network with a small number of parameters. The images enhanced by UWCNN have distinct color cast (Fig. 4c and Fig. 4d) and blurry details (Fig. 3d). Among all the other methods, WaterNet and ours conduct more efficient in visual quality, yet WaterNet performs worse than ours, especially in scenes like Yellowish and low-light scenes. By contrast, the underwater images enhanced by our method are much closer to the ground truth and visually pleasing. Ultimately, the qualitative comparison proves again the superiority of our proposed method.

E. Ablation Study

In order to verify the contribution of each stage in our proposed framework, we conduct ablative experiments on the loss function and post-processing respectively.

1) Loss function: To verify the effectiveness of the perceptual loss, gradient loss, structural similarity loss, and mean square error loss used in our method, we conduct a series of ablation experiments on the four loss functions respectively. Each time a certain loss function item is disabled. The experimental results are shown in Table II.

TABLE II: Loss function ablation experiment

Method	SSIM	PSNR (dB)	UIQM
loss in ours	**0.8720**	**21.9273**	3.0522
w/o L_{vgg} loss	0.8544	20.2642	3.0068
w/o L_{mse} loss	0.8242	19.8349	**3.1453**
w/o L_{ssim} loss	0.8151	19.8073	3.0888
w/o L_{gdl} loss	0.8651	21.3996	3.0661

Comparing the data Table II, we find that the loss proposed can obtain the highest SSIM and PSNR. Without using L_{gdl}, the UIQM value is slightly improved at the expense of decreasing UIQM and SSIM. Without using L_{mse}, nothing could be higher with a dramatic decline except UIQM. In general, our proposed collaborative loss successfully reach a balance across SSIM, PSNR, and UIQM, leading to high-quality underwater images.

2) Post-processing: In order to verify the contribution of the post-processing operation, we conduct a post-processing ablation experiment in which the post-processing is removed.

We can see from Table. III that without the post-processing operation, both SSIM and PSNR are greatly decreased although there is a slight decline in UIQM. We also visualize the results in Fig. 4. It is obvious that using the post-processing significantly improves the the sharpness and brightness of the images.

TABLE III: Efficiency comparison experiment results

Method	SSIM	PSNR (dB)	UIQM
w/o post-processing	0.8606	21.3234	**3.0963**
Ours	**0.8720**	**21.9273**	3.0522

F. Computational Complexity

In addition to the quantitative and qualitative comparisons above, we analyze the number of parameters and the runtime of these CNN-based methods for comparison, as shown in Table IV. Although UWCNN has fewer parameters and UWGAN has less runtime than ours, their performance is much lower than ours. The number of parameters in WaterNet is 1106k, which is 2.5 times our number. The runtime of WaterNet is also 3 times that of ours. Instead, our method reaches the highest performance of all with low complexity cost.

V. Conclusion

This paper proposes an efficient, low-complexity framework for underwater image quality enhancement. Our framework

TABLE IV: computational complexity comparison with existing CNN-based methods

Method	Parameter	Runtime
UWCNN [18]	97901	0.2894
UWGAN [12]	1939194	0.0170
WaterNet [15]	1106274	0.6096
Ours	436682	0.2025

consists of two stages: CNN-based enhancement and post-processing operation. In terms of the CNN-based enhancement, we design a CNN structure using cascading residual groups for feature extraction and enhancement. In each residual group, the channel attention mechanism is incorporated to self-adjust the channel weight coefficients of each channel. We observe that the images generated from CNN are of low contrast. To this end, we feed the images to the post-processing stage for further brightness improvement. Specifically, the image is transformed into YUV space and the Y component is normalized and corrected. We conduct our experiments on the conventional UIEB dataset. Results demonstrate that compared with state-of-the-art methods, our proposed approach achieves the highest SSIM and PSNR values of all. Besides, we also compare the visual quality of all methods. Results confirm that the underwater images enhanced by our approach look much better than those of other methods and much closer to the ground truth. In terms of the computational complexity, our method costs only 436k parameters and our runtime is only $1/3$ of existing method WaterNet.

REFERENCES

[1] C.-Y. Li, J.-C. Guo, R.-M. Cong, Y.-W. Pang, and B. Wang, "Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior," *IEEE Transactions on Image Processing*, vol. 25, no. 12, pp. 5664–5677, 2016.

[2] C. Ancuti, C. O. Ancuti, T. Haber, and P. Bekaert, "Enhancing underwater images and videos by fusion," in *2012 IEEE conference on computer vision and pattern recognition*. IEEE, 2012, pp. 81–88.

[3] B. McGlamery, "A computer model for underwater camera systems," in *Ocean Optics VI*, vol. 208. International Society for Optics and Photonics, 1980, pp. 221–231.

[4] J. S. Jaffe, "Computer modeling and the design of optimal underwater imaging systems," *IEEE Journal of Oceanic Engineering*, vol. 15, no. 2, pp. 101–111, 1990.

[5] N. Carlevaris-Bianco, A. Mohan, and R. M. Eustice, "Initial results in underwater single image dehazing," in *Oceans 2010 Mts/IEEE Seattle*. IEEE, 2010, pp. 1–8.

[6] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 12, pp. 2341–2353, 2010.

[7] H. Lu, Y. Li, L. Zhang, and S. Serikawa, "Contrast enhancement for images in turbid water," *JOSA A*, vol. 32, no. 5, pp. 886–893, 2015.

[8] Y.-T. Peng and P. C. Cosman, "Underwater image restoration based on image blurriness and light absorption," *IEEE transactions on image processing*, vol. 26, no. 4, pp. 1579–1594, 2017.

[9] W. Song, Y. Wang, D. Huang, and D. Tjondronegoro, "A rapid scene depth estimation model based on underwater light attenuation prior for underwater image restoration," in *Pacific Rim Conference on Multimedia*. Springer, 2018, pp. 678–688.

[10] X. Chen, J. Yu, S. Kong, Z. Wu, X. Fang, and L. Wen, "Towards quality advancement of underwater machine vision with generative adversarial networks," 2018.

[11] J. Li, K. A. Skinner, R. M. Eustice, and M. Johnson-Roberson, "Watergan: Unsupervised generative network to enable real-time color correction of monocular underwater images," *IEEE Robotics and Automation letters*, vol. 3, no. 1, pp. 387–394, 2017.

[12] N. Wang, Y. Zhou, F. Han, H. Zhu, and J. Yao, "Uwgan: underwater gan for real-world underwater color restoration and dehazing," *arXiv preprint arXiv:1912.10269*, 2019.

[13] C. Fabbri, M. J. Islam, and J. Sattar, "Enhancing underwater imagery using generative adversarial networks," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 7159–7165.

[14] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[15] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao, "An underwater image enhancement benchmark dataset and beyond," *IEEE Transactions on Image Processing*, vol. 29, pp. 4376–4389, 2019.

[16] C. Zhang, Q. Yan, Y. Zhu, X. Li, J. Sun, and Y. Zhang, "Attention-based network for low-light image enhancement," in *2020 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2020, pp. 1–6.

[17] H. Li, X. Yang, Z. Li, and T. Zhang, "Underwater image enhancement with image colorfulness measure," *arXiv preprint arXiv:2004.08609*, 2020.

[18] C. Li, S. Anwar, and F. Porikli, "Underwater scene prior inspired deep underwater image and video enhancement," *Pattern Recognition*, vol. 98, p. 107038, 2020.

[19] Q. You, C. Wan, J. Sun, J. Shen, H. Ye, and Q. Yu, "Fundus image enhancement method based on cyclegan," in *2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE, 2019, pp. 4500–4503.

[20] M. J. Islam, P. Luo, and J. Sattar, "Simultaneous enhancement and super-resolution of underwater imagery for improved visual perception," *arXiv preprint arXiv:2002.01155*, 2020.

[21] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[22] K. Panetta, C. Gao, and S. Agaian, "Human-visual-system-inspired underwater image quality measures," *IEEE Journal of Oceanic Engineering*, vol. 41, no. 3, pp. 541–551, 2015.

[23] P. L. Drews, E. R. Nascimento, S. S. Botelho, and M. F. M. Campos, "Underwater depth estimation and image restoration based on single images," *IEEE computer graphics and applications*, vol. 36, no. 2, pp. 24–35, 2016.

[24] R. Hummel, "Image enhancement by histogram transformation," *Unknown*, 1975.

[25] S. M. Pizer, "Contrast-limited adaptive histogram equalization: Speed and effectiveness stephen m. pizer, r. eugene johnston, james p. ericksen, bonnie c. yankaskas, keith e. muller medical image display research group," in *Proceedings of the First Conference on Visualization in Biomedical Computing, Atlanta, Georgia*, vol. 337, 1990.

[26] Y. Guo, H. Li, and P. Zhuang, "Underwater image enhancement using a multiscale dense generative adversarial network," *IEEE Journal of Oceanic Engineering*, vol. 45, no. 3, pp. 862–870, 2019.