AGCYCLEGAN: ATTENTION-GUIDED CYCLEGAN FOR SINGLE UNDERWATER IMAGE RESTORATION

Zhenlong Wang *Weifeng Liu* *Yanjiang Wang* *Baodi Liu*⋆

College of Oceanography and Space Informatics, China University of Petroleum (East China)
College of Control Science and Engineering, China University of Petroleum (East China)

ABSTRACT

Underwater image restoration is a fundamental problem in image processing and computer vision. It has broad application prospects for underwater operations, especially underwater robot operations. The challenging work is how to keep the color authenticity of the captured underwater image. In this paper, we propose a novel network architecture based on CycleGAN. Specifically, in the generator part, we adopt the U-Net structure because the long skip connection of U-Net will obtain more detailed information. Besides, we append the pixel-level attention block to provide greater flexibility for detail structure modeling. It assigns different weights to each channel to pay more attention to the critical feature. We also verify its generalization performance on several benchmark datasets. The extensive experiments with comparisons to state-of-the-art approaches demonstrate the superiority of the proposed model.

***Index Terms*—** Underwater image restoration, CycleGAN, Attention

1. INTRODUCTION

In the past decades, underwater image restoration has attracted more and more research work, such as underwater target detection and underwater object recognition[1]. Absorption and scattering, determined by the internal optical properties of water, affect the underwater imaging process when the photographer is underwater. Active illumination of the underwater imaging process with an artificial light source can improve the above phenomena to a certain extent. Still, the light intensity of the artificial light source gradually attenuates along the radial direction centered on the most vital point of light quantity, which will lead to uneven background gray, false contour, false detail, and self shadow of the obtained image.

This research was supported by the Natural Science Foundation of Shandong Province under Grants ZR2019MF073, the Fundamental Research Funds for the Central Universities, China University of Petroleum (East China) under Grant 20CX05001A, the Major Scientific and Technological Projects of CNPC under Grant ZD2019-183-008, and the Creative Research Team of Young Scholars at Universities in Shandong Province under Grant 2019KJN019.

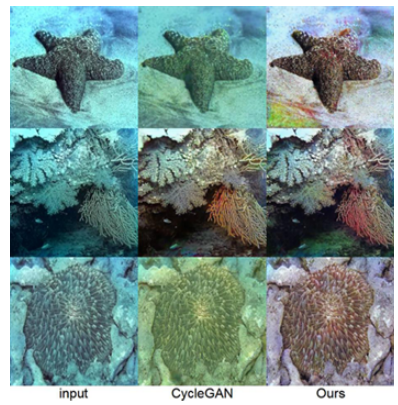

Fig. 1. The comparison on the visual effect of EUVP obtains via the CycleGAN and the proposed model.

The traditional underwater image processing methods[2] are divided into physical model-based and non-physical model-based. Methods based on non-physical models, such as histogram[3], the technique based on view node[4] and strategy based on fusion[5], approaches based on the dark channel a priori[6], refer to image dehazing methods. These underwater image processing methods usually improve the image's visual appearance by modifying the image pixels, regardless of the physical degradation mechanism of the underwater image.

These years, deep learning methods have shown significant success in the computer vision field, and the application has become more extensive in image restoration. The deep generation model based on CNNs provides remarkable performance in the image to image translation[7, 8]. As a hot topic of deep learning in recent years, Generative Adversarial Networks(GAN) have received extensive attention in image generation and image to image translation. Progress in previous work on underwater image restoration tasks often adopts CycleGAN[9]. But CycleGAN[10] pays more attention to the conversion of image style when migrating image style and does not pay attention to the recovery of color and detail, which may cause color distortion and loss of detail.

To solve this problem, we propose a novel Cycle Generative Adversarial Network with an attention mechanism called

AGCycleGAN. We use the U-Net [11] network with a long-skipping connection to replace the generator and add the attention module to make up for this defect. It transforms the image from a style to a target style, providing a more flexible underwater image restoration model. We focus not only on the performance of image restoration but also on the consistency of the content and structure between the generated image and the original image. The recovery effect is shown in Fig.1.

The main contributions of this work are summarized as follows:

- An end-to-end underwater image restoration algorithm based on the depth GAN model is proposed to improve the restored image's contrast and color difference. We use the semantic segmentation U-Net network as the backbone of the generator, which enhances the generality and accuracy of the method.

- We use an attention block, which combines channel attention and pixel attention mechanism. This block provides additional flexibility when dealing with various types of information and pays more attention to the underwater targets to be recovered and more critical channel information.

- Adequate experiments on artificial and real datasets show that the proposed model, AGCycleGAN, has superior performance to the most advanced methods. In particular, the results of our proposed model are notable in visual perception.

2. PROPOSED METHOD

This section presents the details of the proposed underwater image restoration. First, we introduce a detailed overview of our method. Then the image restoration module is described in detail. Finally, we provide the loss functions that are applied to train the proposed networks.

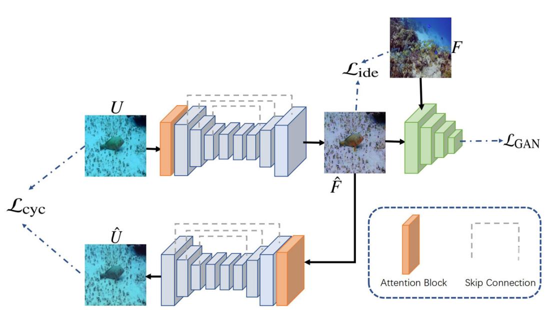

Fig. 2. The network framework of AGCycleGAN.

2.1. Formulation

We propose an attention-guided CycleGAN for underwater image restoration. The AGCycleGAN model we proposed consists of the generator, discriminator, and attention modules, as shown in Fig.2. We define $\{u_i\}_{i=1}^{N} \in U$ (e.g., degraded underwater images) and $\{f_i\}_{i=1}^{N} \in F$ (e.g., free underwater images). The underwater images are not easy to collect the corresponding clear images. It is necessary to train on unpaired datasets.

Through the generator G_1, the turbid underwater image u can be converted into a clear underwater image $\hat{F}$. The purpose of the discriminator D is to distinguish the generated clear underwater image from the real clear underwater image f. Similarly, mapping G_2 represents from source domain F to target domain U.

2.2. Attention Block

An attention mechanism is used increasingly in low-level computer vision tasks such as image dehazing, image translation, and denoise. Natural images are extremely structured, and they represent a stable correlation among the pixel points of the underwater image and the recovered pixel points. In underwater image restoration, attention mechanisms can focus network learning on the detail of restoring underwater images.

Inspired by FFA-Net[12] with high effective FA blocks, We added the attention module to the network. The detail of the proposed network model is shown in Fig.4.

Channel attention(CA) The channel attention mechanism we use focuses on the degree of dependence among various channels. The describe of CA is

$$CA = \sigma\left(\text{Conv}\left(\delta\left(\text{Conv}\left(Avg\left(F_{in}\right)\right)\right)\right)\right) \tag{1}$$

where the σ is the $sigmoid$ function, δ is the $ReLu$ function,

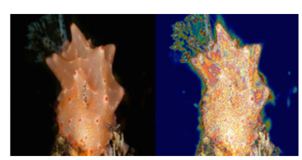

Fig. 3. Characteristic heat map after passing through our attention network, the left is the input image and the right is the corresponding attention heat map.

$Conv$ is the Convolution operation and F_{in} is the input of feature map.

Finally, we conduct element-wise multiply for input and the weights of the channel CA.

$$F^{*} = CA \otimes F_{in} \tag{2}$$

Pixel Attention (PA) We input the result F^{*} output by the CA layer into the PA layer, and pass through the convolutional layer with $sigmoid$ function and $ReLu$ function.

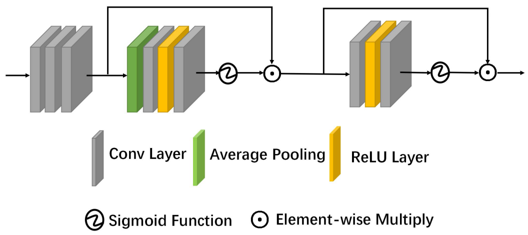

Fig. 4. The framework of Attention Block.

$$PA = \sigma\left(\mathrm{Conv}\left(\delta\left(\mathrm{Conv}\left(F^{*}\right)\right)\right)\right) \tag{3}$$

To intuitively show that the proposed attention module positively affects underwater image restoration tasks, we list the weight heat map of the feature image obtained after passing the attention module. Fig.3 shows our weight heat map image.

2.3. Generator and Discriminator

Underwater image restoration, image denoising, image dehazing, and other low-level vision tasks have higher requirements for image detail restoration. For generative networks, we apply the U-Net as our generator. It can obtain detailed information and low-frequency information of the picture. Then the information at all levels is retained by long-skipping connect so that the whole network can remember all the image information well.

For the discriminator, we use 70×70 PatchGANs to classify whether the image patches are real or fake.

2.4. Loss Functions

The loss function we use refers to CycleGAN, including the identity loss, the adversarial loss, and the cycle consistency loss.

We apply the adversarial loss to the generator G and the discriminator D. We define the loss as follows:

$$\begin{aligned}\mathcal{L}_{\mathrm{GAN}}(G, D, U, F) = & \ \mathtt{E}_{f\sim p_{(f)}}[\log D(f)] \\ & + \mathtt{E}_{u\sim p_{(u)}}[\log(1 - D(G(u))]\end{aligned} \tag{4}$$

In addition, we apply the cycle-consistency loss to standardize the training of translation networks. Specifically, when passing an image u to G_1 and G_2sequentially, we expect the output should be the same, and vice versa for f. The cycle consistency loss can be expressed as:

$$\begin{aligned}\mathcal{L}_{\mathrm{cyc}}(G_1, G_2) = & \ \mathtt{E}_{u\sim p_{(u)}}[\|G_2(G_1(u)) - u\|_1] \\ & + \mathtt{E}_{f\sim p_{(f)}}[\|G_1(G_2(f)) - f\|_1]\end{aligned} \tag{5}$$

Underwater image restoration is mainly to restore the original color of the underwater target to be detected. We input the underwater image. Generator G should generate an image closer to the F domain, and $G(u)$ should be similar to F. We use the identity loss as follows:

$$\begin{aligned}\mathcal{L}_{\text{Identity}}(G_1, G_2) = & \ \mathtt{E}_{f\sim p_{(f)}}[\|G_1(f) - f\|_1] \\ & + \mathtt{E}_{u\sim p_{(u)}}[\|G_2(u) - u\|_1]\end{aligned} \tag{6}$$

Total Loss $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{GAN}} + \lambda_1 \mathcal{L}_{\text{cyc}} + \lambda_2 \mathcal{L}_{\text{Identity}}$

3. EXPERIMENTS

3.1. Datasets

In the experiments, we adopt two benchmark datasets, such as the EUVP[9] dataset and the U-45[13] dataset, to evaluate the proposed AGCycleGAN.

EUVP This dataset is collected by various cameras, such as GoPros, low light level USB, and other cameras during ocean exploration under different visibility conditions. It includes $5,550$ pairs of paired test datasets, $3,200$ unpairs of unpaired test datasets, and 515 pairs of test datasets. The resolution size of the picture is 256×256.

U45 This dataset is a publicly available underwater test dataset that contains underwater images with degraded color casts, low contrast, and haze effects. This dataset includes green light images, blue light images, and images with haze, respectively. It contains 45 degraded images, and the resolution size of the picture is 256×256.

3.2. Quantitative Evaluation

We consider the two standard measures of Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) to quantitatively compare the underwater images generated by the proposed network model with their respective ground truths. We also use the non-reference underwater image quality measure (UIQM) to analyze the quality of the output image. UIQM includes three underwater image attribute measures: color measurement (UICM), sharpness measurement (UISM), and contrast measurement (UIConM). Each attribute evaluates one aspect of underwater image degradation. Uiqm indicators are determined by:

$$\mathrm{UIQM} = c_1 \times \mathrm{UICM} + c_2 \times \mathrm{UISM} + c_3 \times \ \mathrm{UIConM} \tag{7}$$

where $c_1 = 0.0282, c_2 = 0.2953, c_3 = 3.5753$ set according to original paper[14].

3.3. Network Implementation

For all experiments, we set $\lambda_1 = 10$, $\lambda_2 = 5$ in Eqn.6. We use the Adam optimizer with a batch size of 1. All networks are trained from scratch, and the initial learning rate is 0.0002. The size of the picture we input on the network is 256×256. We maintain the same learning rate in the first 100 epochs and decay linearly to zero in the following 100 epochs. We used PyTorch as the deep learning framework on an Nvidia Tesla V100 GPU.

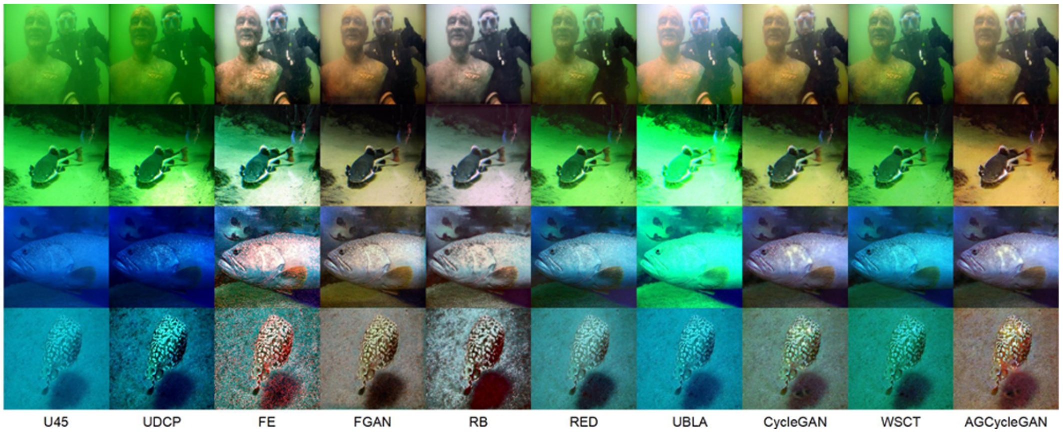

Fig. 5. Comparison of the visual effects to previous work on the U45 dataset.

Table 1. Performance evaluation on the U45 dataset.

Method	U45	FE	UCDP	FGAN	CycleGAN	RB	RED	UIBLA	WSCT	AGCycleGAN
UIQM	2.494	2.984	2.339	3.158	3.138	3.101	2.979	2.401	2.890	**3.183**

Table 2. Quantitative results on synthetic underwater images in terms of SSIM, PSNR,and UIQM.

Method	PSNR	SSIM	UIQM
WaterNet[15]	24.43 ± 4.64	0.82 ± 0.08	2.97 ± 0.32
FUnIE-GAN[9]	26.19 ± 2.87	0.82 ± 0.08	2.84 ± 0.46
DeepSESR[16]	25.30 ± 2.63	0.81 ± 0.07	2.95 ± 0.32
Shallow-UWnet[17]	27.39 ± 2.70	0.83 ± 0.07	2.98 ± 0.38
CycleGAN[18]	26.39 ± 2.53	0.93 ± 0.01	$\mathbf{3.02 \pm 0.08}$
AGCycleGAN	$\mathbf{30.30 \pm 2.27}$	$\mathbf{0.95 \pm 0.01}$	3.02 ± 0.01

Table 3. Comparison between CycleGAN and AGCycleGAN of quantitative results on synthetic underwater images in terms of SSIM, PSNR,and UIQM.

Method	PSNR	SSIM	UIQM
CycleGAN	25.01	0.983	**3.093**
AGCycleGAN	**35.44**	**0.990**	3.064

Table 4. Ablation study on the EUVP dataset with various configurations.

Method	PSNR	SSIM	UIQM
CycleGAN	25.01	0.983	3.093
CycleGAN+U-Net	32.77	0.987	**3.101**
AGCycleGAN	**35.44**	**0.990**	3.064

3.4. Experimental Results

In this subsection, we display a quantitative and qualitative analysis of the experimental results. To prove that our experimental results reach state-of-the-art, we have carried out experiments and tests on multiple datasets. The results show that our effect reaches state-of-the-art. Table 2 shows the results and Table 3 list the performance evaluation.

We list the visual effect in Fig.5 and performance evaluation in Table 1. The competitive methods include Fusion Enhance (FE)[19], Retinex-Based (RB)[4], UDCP[20], UIBLA[21], CycleGAN[18], Weakly Supervised Color Transfer (WSCT)[22], and UGAN[8].

3.5. Ablation Experiment

To further demonstrate the superiority of our network structure, we conducted ablation experiments considering the different modules of our proposed network. We mainly focus on the following factors: 1) attention block module. 2) U-Net network structure. We tested the influence of these modules on the experimental results. Table 4 shows the experimental results. Our model has performance progressively.

4. CONCLUSION

In this paper, we present a degraded underwater image restoration network architecture called AGCycleGAN. This network is an end-to-end unpaired adversarial system. Subjective and objective results show that the image generated by this method displays marvelous quality in contrast enhancement and color correction. Moreover, it is evaluated on multiple datasets to ensure its robustness and application in the real world.

5. REFERENCES

[1] Wei-Hong Lin, Jia-Xing Zhong, Shan Liu, Thomas Li, and Ge Li, "Roimix: Proposal-fusion among multiple images for underwater object detection," *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2588–2592, 2020.

[2] M. Mangeruga, F. Bruno, M. Cozza, P. Agrafiotis, and D. Skarlatos, "Guidelines for underwater image enhancement based on benchmarking of different methods," *Remote Sensing*, vol. 10, pp. 1652, 2018.

[3] L. Sun, X. Wang, X. Liu, P. Ren, P. Lei, J. He, S. Fan, Y. Zhou, and Y. Liu, "Lower-upper-threshold correlation for underwater range-gated imaging self-adaptive enhancement," *Applied Optics*, vol. 55, pp. 8248–8255, 2016.

[4] X. Fu, P. Zhuang, H. Yue, Y. Liao, X. P. Zhang, and X. Ding, "A retinex-based enhancing approach for single underwater image," *2014 IEEE international conference on image processing (ICIP)*, pp. 4572–4576, 2014.

[5] H. Lu, Y. Li, S. Nakashima, H. Kim, and S. Serikawa, "Underwater image super-resolution by descattering and fusion," *IEEE Access*, vol. 5, pp. 670–679, 2017.

[6] K. He, S. Jian, Fellow, IEEE, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 33, pp. 2341–2353, 2011.

[7] Youssef Alami Mejjati, Christian Richardt, James Tompkin, Darren Cosker, and Kwang In Kim, "Unsupervised attention-guided image-to-image translation," *Advances in neural information processing systems*, vol. 31, 2018.

[8] Cameron Fabbri, Md Jahidul Islam, and Junaed Sattar, "Enhancing underwater imagery using generative adversarial networks," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 7159–7165.

[9] M. J. Islam, Y. Xia, and J. Sattar, "Fast underwater image enhancement for improved visual perception," *IEEE Robotics and Automation Letters*, vol. 5, pp. 3227 – 3234, 2020.

[10] J. Lu, N. Li, S. Zhang, Z. Yu, H. Zheng, and Z. Bing, "Multi-scale adversarial network for underwater image restoration," *Optics & Laser Technology*, vol. 110, pp. 105–113, 2018.

[11] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, 2015.

[12] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "Ffa-net: Feature fusion attention network for single image dehazing," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, pp. 11908–11915, 2020.

[13] H. Li, J. Li, and W. Wang, "A fusion adversarial underwater image enhancement network with a public test dataset," *arXiv:1906.06819v2*, 2019.

[14] K. Panetta, C. Gao, and S. Agaian, "Human-visual-system-inspired underwater image quality measures," *IEEE Journal of Oceanic Engineering*, vol. 41, pp. 541–551, 2016.

[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] 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.

[17] A. Naik, A. Swarnakar, and K. Mittal, "Shallow-uwnet : Compressed model for underwater image enhancement," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, pp. 15853–15854, 2021.

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

[19] C. Ancuti, C. O. Ancuti, T. Haber, and P. Bekaert, "Enhancing underwater images and videos by fusion," *IEEE Conference on Computer Vision & Pattern Recognition*, pp. 81–88, 2012.

[20] 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*, vol. 36, no. 2, pp. 24–35, 2016.

[21] Y. T. Peng and P. C. Cosman, "Underwater image restoration based on image blurriness and light absorption," *IEEE Trans Image Process*, vol. 26, pp. 1579–1594, 2016.

[22] C. Li, J. Guo, and C. Guo, "Emerging from water: Underwater image color correction based on weakly supervised color transfer," *IEEE Signal Processing Letters*, vol. 25, pp. 323–327, 2018.