Advances in Machinery, Materials Science and Engineering Application IX
M. Chen et al. (Eds.)

doi:10.3233/ATDE230554

Underwater Image Enhancement Algorithm Based on UWGAN and U-Net

Xinyan YIN [a,1], Xiwei CHEN [a], Yihang YANG [a], Ying LIU [a] and Lei BI [b]
[a] *Beijing Institute of Technology, Zhuhai, 519088, P. R. China*
[b] *93199 Troops of the PLA, Harbin, 150000, P. R. China*

Abstract. The underwater environment presents unique characteristics that often result in defects like low contrast and blurred edges. To address these issues and improve underwater target recognition for vehicles, a new method that combines UWMAN and U-Net algorithms has been developed. One key advantage of this algorithm is its ability to counteract the impact of underwater environmental factors. It achieves this by enhancing the color of the images and improving various details, leading to higher scores in evaluations of underwater color image quality and overall image quality. This means that the algorithm effectively enhances the clarity and visual appeal of underwater images. Additionally, the enhanced images obtained through this method demonstrate improved matching effects in feature point matching experiments. This indicates that the algorithm enhances the identification and alignment of crucial features within the images, resulting in more accurate target recognition. These capabilities are vital for underwater vehicles operating in challenging underwater environments. Overall, the combination of UWMAN and U-Net algorithms represents a significant advancement in enhancing image clarity and improving the accuracy of underwater target recognition. The algorithm's ability to mitigate the impact of underwater environmental factors and produce visually pleasing images holds great potential for various applications related to underwater exploration, marine research, and inspection tasks beneath the water's surface.

Keywords. Deep learning, neural network, GAN, U-Net, underwater image enhancement

1. Introduction

Underwater image enhancement technology holds significant potential for various applications, including underwater robot operations, as it contributes to enhancing the accuracy of underwater target recognition and improving the endurance of underwater robots [1]. The primary function of an underwater robot's vision system is to collect surrounding environmental data and analyze the corresponding information [2]. By utilizing the location information of the target, the vision system facilitates target tracking, monitoring, and helps in obtaining real-time environmental data for real-time analysis and management [3,4]. In the context of underwater robots, target detection serves as a crucial component of the vision system, primarily employed for close-range detection purposes [5, 6].

[1] Xinyan YIN, Corresponding author, Beijing Institute of Technology, Zhuhai, 519088, P. R. China; E-mail: xiaowen122@yeah.net.

In the common underwater operation scene, the underwater target detection and recognition technology based on optical vision has relatively good performance in short-range target recognition and accurate positioning, and is more widely used in the application of underwater robots. Wang proposed a method for enhancing underwater images captured under natural lighting conditions by combining a refined underwater imaging model [7] with scene depth estimation. This method effectively eliminates the haze and blur caused by backscattering, enhances image contrast, and corrects color deviations. However, it requires prior knowledge and scene depth information to estimate the backscatter component of individual pixels. Chen discussed an underwater image enhancement algorithm [8] that utilizes depth learning techniques and a physical imaging model. The algorithm employs a neural network with expansion convolution and parameter activation functions to estimate background scattering and direct transmission in underwater scenes. The model is trained using the UIEB dataset, which consists of real-world underwater images. However, the dataset's limited size poses a challenge in terms of training diversity [9].

2. Model and Algorithm

2.1. Underwater Scene Atlas Reconstruction Based on UWGAN-Fast

On the basis of UWGAN (Underwater GAN), the lightweight UWGAN-Fast is obtained by online multigranularity distillation (OMGD) method to reconstruct images with underwater style. The generative countermeasure network (GAN) has achieved great success in generating excellent images. However, due to the high computing cost and large memory usage, it is very difficult to deploy GAN on equipment with limited resources. The Online Multi-granularity Distillation (OMGD) scheme is a method for obtaining a lightweight Generative Adversarial Network (GAN). This scheme enables the generation of high-fidelity images while reducing the computational requirements. Promote single-stage online distillation to GAN-oriented compression, and gradually upgrade the teacher generator will help improve the student generator based on discriminator. The combination of a complementary teacher generator and network layer offers a comprehensive and multi-granularity approach to enhancing visual fidelity from various perspectives. Experimental results on four benchmark data sets show that OMGD has successfully compressed 40% on Pix2Pix and CycleGAN × MAC and 82.5 ×MAC parameters, and no image quality loss. This shows that OMGD provides a feasible solution for deploying real-time image translation on resource-constrained devices. There are three main problems in the existing compression algorithms.

Initially, there is a tendency to directly utilize well-established model compression technologies that are not specifically tailored for Generative Adversarial Networks (GANs). Consequently, there is a lack of exploration of the intricate characteristics and structure unique to GANs. Moreover, conventional approaches typically depict GAN compression as a multi-phase process that involves pre-training, distillation, evolution, and fine-tuning conducted sequentially. Specifically, the distillation-based approach necessitates training the teacher generator beforehand, followed by the distillation of knowledge to the student generator. However, an end-to-end methodology becomes vital in order to mitigate the complexities associated with time and computational resources within a multi-phase framework. Furthermore, even the state-of-the-art

methods still incur significant computational costs, which pose challenges for their deployment on resource-constrained edge devices [10]. To address this issue, the Online Multi-granularity Distillation (OMGD) approach extends the online distillation strategy to a multi-granularity solution, as illustrated in figure 1. It leverages teacher generators with diverse structures to capture more complementary information, thereby enhancing visual fidelity from multiple dimensions. In this approach, the student model is expanded into a teacher model using two complementary dimensions: depth and width. To achieve this, the channel of the student generator is extended to obtain a wider teacher generator. Additionally, several Resnet blocks are inserted after each lower sampling layer and upper sampling layer of the student generator to create a deeper teacher generator.

To incorporate the complementary teacher generator into the multi-teacher setup, this study combines the two distillation losses into the knowledge distillation losses. Besides considering the output layer, OMGD utilizes the granularity information of the intermediate channel as an auxiliary monitoring signal for distillation optimization. This is achieved by calculating attention weights on the channel dimension to measure the importance of each channel in the feature map, and transferring this attention information to the student model.

Numerous experiments demonstrate that OMGD can significantly reduce the computational costs of Pix2Pix and CycleGAN without causing substantial loss in visual fidelity. This provides a viable solution for deploying real-time GAN on devices with limited resources. Therefore, this paper aims to utilize OMGD to compress UWGAN and obtain the compressed UWGAN-Fast model.

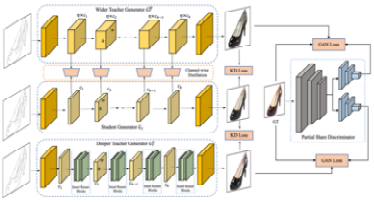

Figure 1. OMGD neural network model.

Various underwater activities, such as seabed resource exploration, underwater archaeology, and underwater fishing, heavily rely on sensor technology. Among these sensors, the visual sensor holds paramount importance due to its non-invasiveness, high information content, and passive nature. However, the underwater environment poses challenges to visual sensing, primarily attributed to wavelength-dependent light attenuation and backscattering phenomena. These factors cause color distortion and haze effects that significantly reduce image visibility.

Gatys [14] used a style cost model to achieve image style conversion tasks. The Gram matrix concept represents the correlation between various channels of an image in the form of a matrix. The definition of the Gram matrix is as follows,

$$G_{kk'}^{[l][S]} = \sum_{i=1}^{n_H^{[l]}} \sum_{j=1}^{n_W^{[l]}} a_{i,j,k}^{[l][S]} a_{i,j,k'}^{[l][S]} \tag{1}$$

where, S is the style image. l is convolutional layer depth. $G_{kk'}^{[l][S]}$ represents the value of an element with a row column coordinate of (k,k') using style images as input in

the Gram matrix of the output features in the l_{th} layer. $a_{i,j,k}^{[l][S]}$ is the activation item at position (i, j, k) in l_{th} layer. $n_H^{[l]}$ is the height of the output feature of the l_{th} layer. $n_W^{[l]}$ is the width of the output feature of the l_{th} layer. i, j, k indicate the height, width, and corresponding number of channels of the position.

And, the definition of the style cost function as,

$$J_{style}^{[l]}(S,G) = \frac{1}{(2n_H^{[l]}n_W^{[l]}n_C^{[l]})^2}\sum_k\sum_{k'}(G_{kk'}^{[l][S]} - G_{kk'}^{[l][G]}) \tag{2}$$

where, G is the generated image. $n_C^{[l]}$ is the number of channels for the output feature of the l_{th} layer. $G_{kk'}^{[l][G]}$ represents the value of an element with a row column coordinate of (k,k') using generated images as input in the Gram matrix of the output features in the l_{th} layer. Combined with L_2 normal form Loss function and style cost function, composite Loss function is designed [15],

$$L(x) = \frac{1}{n}\sum_{x\in X}(g(x) - r(x))^2 + J_{style}^{[l]}(g(x), r(x)) \tag{3}$$

where, x is the coordinate of a single pixel within the X range, X is the coordinate set of all pixels in the input image, N is the sum of the number of pixels in the input image, $g(x)$ is the pixel value at the model output image coordinate x, and $r(x)$ is the pixel value at the true value image coordinate x.

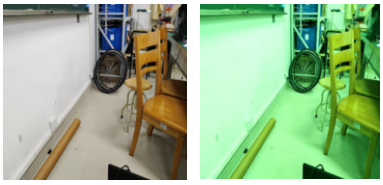

Figure 2. Comparison between the original image (left) and the underwater style image generated by UWGAN (right).

UWGAN utilizes an enhanced underwater imaging model, depicted in figure 2, to generate realistic underwater images with color distortion and haze effects. It achieves this by processing aerial images and depth maps as input pairs. To enhance color recovery and eliminate mist, U-Net technology is employed, leveraging synthetic underwater datasets for effective training. The model employs an automatic encoder network, ensuring an end-to-end approach to directly reconstruct clear underwater images while preserving the structural similarity of the scene content.

To deploy the UWGAN model in an underwater robot system with limited computing resources, several optimizations are implemented. Initially, the original UWGAN generator is transformed into a teacher generator and a student generator to compress the model. The intermediate presentation layer of the teacher generator is migrated to the corresponding compressed student generator layer. Furthermore, pseudo-pairs are generated using the outputs of the teacher model to facilitate non-pair training and transform it into paired learning.

Additionally, neural architecture search (NAS) is employed to automatically discover an efficient network architecture with reduced computational costs and parameters. To minimize training expenses, this paper introduces a "once-for-all" network that encompasses all possible channel number configurations. Through weight sharing, the once-for-all network generates multiple subnets, each of which can be evaluated for performance without requiring additional training. Figure 3 illustrates the enhanced generator model resulting from these optimizations.

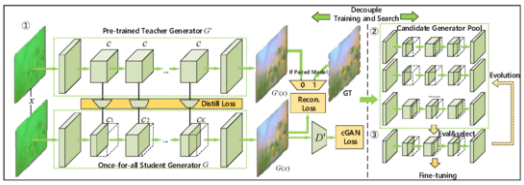

Figure 3. The improved generator model of the once-for-all net [11].

UWGAN-Fast is an enhanced version of UWGAN that offers the capability to generate underwater-style images with fewer computing resources. The process begins by importing a dataset into UWGAN-Fast, which is then used to generate a significant number of underwater-style images, as depicted in figure 4. These generated images serve as an augmented dataset for training the improved U-Net network.

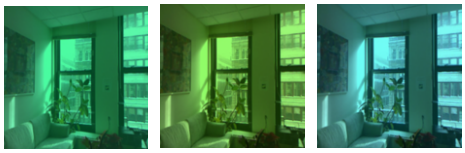

Figure 4. Images of different underwater styles generated by UWGAN-Fast.

2.2. Create Improved U-Net to Repair Underwater Images

To enhance the quality of underwater images, an improved U-Net architecture is employed. U-Net is a convolutional neural network specifically designed for biomedical image segmentation. It derives its name from its "U-shaped" network structure, as depicted in figure 5. U-Net has proven to be highly effective for tasks that require similar input and output sizes, making it suitable for tasks like image processing, generation, and segmentation.

Unlike traditional convolutional neural networks used for image classification, U-Net utilizes a two-step convolution process with repeated downsampling to reduce the spatial resolution of the input image. However, to generate an output image that matches or exceeds the size of the input, an upsampling path is necessary to increase the resolution. This layout gives U-Net its distinct U-shaped appearance, with the downsampling/encoder path forming the left side of the U, and the upsampling/decoder path forming the right side.

The upsampling/decoder path employs several transposed convolutions, which add pixels between and around existing pixels to reverse the downsampling process. The improved U-Net is trained using a large dataset generated by UWGAN-Fast, which produces underwater-style images, and the original images are utilized as validation sets to fine-tune the network. This training process ensures that the improved U-Net

effectively repairs and enhances the underwater images, providing visually improved results. The initial step in the process involves data preparation, where the training data from UWGAN is utilized. The training dataset consists of a subset of images, specifically those that depict underwater scenes. This subset of underwater images is categorized into two distinct categories based on subjective visual evaluation: the x set comprising undistorted underwater images, and the y set comprising distorted underwater images. UWGAN-Fast can learn mapping functions $f: x \to y$ and $g: x \to y$ use x to degraded $f: x \to y$ images to generate 6514 sets of training data, and has three different styles of image pairs. From the NYU depth dataset, a set of 114 authentic underwater images is chosen. Additionally, a test set of 114 images exhibiting an underwater style is selected, resulting in a total of 228 real underwater images in the test set.

During the training process, both the training and test images are resized to dimensions of 256 × 256 × 3. To achieve different rendering effects, different loss functions are employed using various training sets. Figure 6 illustrates these different loss functions and their corresponding rendering effects.

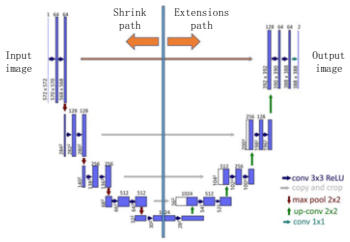

Figure 5. The model of the U-Net.

This paper uses $\lambda_1 = 60, \lambda_g = 10$ and ReLU (slope is 0.2) and Adam algorithm. In this study, the U-Net model was trained using the Tensor-Flow framework on a Tesla P100-PCIE-16GB GPU in Colab. The training process involved 60 epochs, with a learning rate of 0.0001 and a batch size of 32.

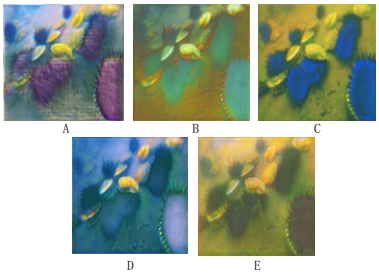

Figure 6. (A~E) show the rendering effects caused by different loss functions.

To evaluate the quality of the trained images, a comparison method was employed, which involved comparing the results obtained from different loss functions used in the improved U-Net model. To ensure the credibility of the findings, two non-reference indicators were utilized for evaluating underwater image quality. The first indicator is

the Underwater Color Image Quality Assessment (UCIQE), which quantifies uneven color deviation, blurring, and low contrast by employing a linear combination of chromaticity, saturation, and contrast [12]. The second indicator is the Underwater Image Quality Measurement (UIQM), which includes three attributes: Underwater Image Color Measurement (UICM), Underwater Image Sharpness Measurement (UISM), and Underwater Image Contrast Measurement (UIConM) [13]. A higher UCIQE value indicates superior image quality.

During the evaluation test, the test set is assessed using UCIQE and UIQM metrics. The results of the evaluation are presented in table 1. Based on the table, it can be concluded that the U-Net model with the specified loss function performs the best in terms of image quality restoration.

Table 1. Image Quality Score Table.

Different loss functions	UCIQE	UICM	UIConM	UISM
Gradient Descent	0.6028	-35.7297	0.7976	7.1897
Gradient Descent+L_1	0.5998	-1.3505	0.8140	7.2454
L_1	0.5533	-67.5051	0.9196	6.9193
L_2	0.5745	-14.2970	0.7916	6.9622
L_1+ L_2	0.6262	5.2901	0.8035	5.0973

2.3. Using ESRGAN to Enhance Underwater Image Resolution

To enhance the super-resolution of the repaired underwater image, ESRGAN (Enhanced Super-Resolution GAN) is employed. ESRGAN is an advanced image algorithm that builds upon SRGAN and further improves the network structure and super-resolution processing with anti-loss and perception loss techniques.

ESRGAN offers several advantages: 1) Enhanced Network Structure: It introduces the Residual-in-residual Dense Block (RRDB) with a larger capacity and easier training, improving the overall network structure. 2) Improved Training Strategies: The BN (Batch Normalization) layer is removed, and Residual-in-residual scaling and small initialization are utilized to enhance training in deep networks. 3) RaGAN Discriminator: The discriminator is upgraded using RaGAN to predict relative authenticity rather than absolute values between the high-resolution image and the original image. This enables the generator to restore more realistic texture details from the original image. 4) Enhanced Perception Loss: The perception loss is improved by altering the VGG features to execute before activation, resulting in enhanced edge definition and texture authenticity in the output image.

The effectiveness of ESRGAN in enhancing the resolution of underwater images can be observed in figure 7.

3. Using ESRGAN to Enhance Underwater Image Resolution

The rendered image is input to the edge computing device, and the lightweight Yolo Fastest neural network is used as the recognition algorithm of the edge computing device to identify the image data collected by the underwater robot.

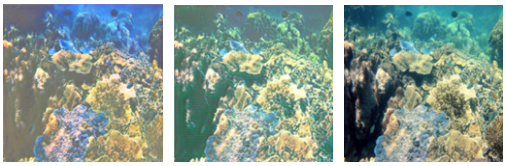

Figure 7. Original image (left), Reconstruction image (middle), Super resolution (right).

Yolo-Fastest is one of the fastest and lightest Yolo universal target detection algorithms known for open source. Its original intention is to break the bottleneck of computing power and run target detection algorithms in real time on more low-cost edge devices, such as raspberry pie 3b, 4-core A53 1.2Ghz. BF16s is enabled in the latest NCNN based reasoning framework. The single reasoning time of 320x320 images is 60 ms, while the single reasoning time of raspberry pie 4b is 33 ms, It achieves full real-time at 30 fps. In contrast, the most widely used lightweight target detection algorithm, MobileNet-SSD, runs about 200ms in Raspberry pie 3b, and the speed of Yolo-Fastest is exactly 3 times faster, and the model is only 1.3MB, while the MobileNet-SSD model reaches 23.2MB, and Yolo-Fastest is exactly 20 times smaller than it. Of course, there is also a cost. In the map on Pascal voc, MobileNet-SSD is 72.7, and Yolo-Fastest is 61.2, resulting in nearly 10 points of accuracy loss. However, on edge computing devices with low load and low computing power, the general detection task itself is not as complex as VOC's detection of 20 categories, which is generally several or even single category detection. In this way, the demand for the accuracy of the model itself is not high. Therefore, Yolo Fastest is adopted as the target detection algorithm for these edge computing devices. Use Yolo-Fastest to detect the sea urchin in the image. Compare the images before and after model rendering to visually reflect the effect of image enhancement, as shown in figure 8. The reconstructed image is clearer, and Yolo-Fastest has more accuracy and quantity of target recognition.

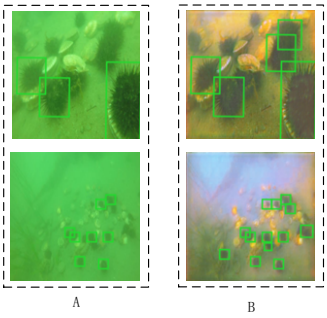

Figure 8. Target detection comparison before (showing as A) and after (showing as B) enhancement.

4. Conclusion

This paper proposes a method that combines UWMAN and U-Net to enhance image clarity. The approach utilizes the improved UWGAN-Fast model to generate underwater-style images efficiently, requiring fewer computational resources. The U-Net network is trained using these generated images as the original dataset to

reconstruct the images. Prior to and after model rendering, Yolo-Fastest is employed to detect sea urchins in the images. The results of the detection indicate that the reconstructed image is clearer, and Yolo-Fastest achieves higher accuracy and target recognition quantity. The algorithm has several advantages. It eliminates the influence of underwater environmental factors, enhances image color to some extent, enriches various details, and achieves higher scores in underwater color image quality and underwater image quality measurement evaluations. Moreover, the enhanced images demonstrate better matching effects in feature point matching experiments.

Acknowledgments

This work is supported by the Characteristic Innovation Project of Colleges and Universities in Guangdong Province (Grant 2020 KTSCX186), the Special Projects in Key Areas of Guangdong Province (Grant 2021ZDZX4050) and the Key Project of Online Open Courses in Guangdong universities (Grant 2022ZXKC550).

References

[1] Li C, Guo C, Ren W, et al. An underwater image enhancement benchmark dataset and beyond. IEEE Transactions on Image Processing. 2019; 29: 4376-4389.
[2] Ancuti CO, Ancuti C, De Vleeschouwer C, et al. Color balance and fusion for underwater image enhancement. IEEE Transactions on Image Processing. 2017; 27(1): 379-393.
[3] Raveendran S, Patil MD, Birajdar GK. Underwater image enhancement: a comprehensive review, recent trends, challenges and applications. Artificial Intelligence Review. 2021; 54: 5413-5467.
[4] Gao, Zhang M, Zhao Q, et al. Underwater image enhancement using adaptive retinal mechanisms. IEEE Transactions on Image Processing. 2019; 28(11): 5580-5595.
[5] Islam MJ, Xia Y, Sattar J. Fast underwater image enhancement for improved visual perception. IEEE Robotics and Automation Letters. 2020; 5(2): 3227-3234.
[6] Fu X, Fan Z, Ling M, et al. Two-step approach for single underwater image enhancement. International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), 2017; pp.789–794.
[7] Wang D, Zhang Z, Zhao J, et al. An underwater image enhancement method under natural light based on scene depth estimation. Robot. 2021; 43 (3): 364–372.
[8] Chen X, Zhang P, Quan L, et al. Underwater image enhancement algorithm combining depth learning and imaging model. Computer Engineering. 2022; 048-002.
[9] Fan X, Yang X, Shi P, et al. Feature fusion generates underwater image enhancement of countermeasure network. Journal of Computer Aided Design and Graphics. 2022; 034-002.
[10] Ren Y, Wu J. Online multi-granularity distillation for GAN compression. arXiv Preprint. 2021; arXiv:2108.06908.
[11] He K, Chen X, Xie S, et al. Masked autoencoders are scalable vision learners. arXiv Preprint. 2021; arXiv: 2111.06377.
[12] Zhao H. An aerial image defogging algorithm combining a priori dense attention network. Value Engineering. 2021; 40(4):5.
[13] Yong Z, Guo J, Li C. Weak supervised underwater image enhancement algorithm incorporating attention mechanism. Journal of Zhejiang University: Engineering Edition. 2021.
[14] Gatys LA, Ecker AS, Bethge M. A neural algorithm of artistic style. arXiv Preprint. 2015; arXiv:1508.06576.
[15] Wen P, Chen J, Xiao Y, et al. Underwater image enhancement algorithm based on GAN and multi-level wavelet CNN. Journal of ZheJiang University (Engineering Science). 2022; 56(2): 213-224.