

Article

Real-World Underwater Image Enhancement Based on Attention U-Net

Pengfei Tang ¹, Liangliang Li ² , Yuan Xue ³, Ming Lv ³, Zhenhong Jia ³ and Hongbing Ma ^{2,*} ¹ School of Software, Xinjiang University, Urumqi 830091, China; pengfei_tang@foxmail.com² Department of Electronic Engineering, Tsinghua University, Beijing 100084, China; leeliangliang@163.com³ College of Information Science and Engineering, Xinjiang University, Urumqi 830046, China; 18309886088@163.com (Y.X.); lvming09@163.com (M.L.); jzh@xju.edu.cn (Z.J.)

* Correspondence: hbma@tsinghua.edu.cn

Abstract: In recent years, with the increasingly serious problems of resource shortage and environmental pollution, the exploration and development of underwater clean energy were particularly important. At the same time, abundant underwater resources and species have attracted a large number of scientists to carry out research on underwater-related tasks. Due to the diversity and complexity of underwater environments, it is difficult to perform related vision tasks, such as underwater target detection and capture. The development of digital image technology has been relatively mature, and it has been applied in many fields and achieved remarkable results, but the research on underwater image processing technology is rarely effective. The underwater environment is much more complicated than that on land, and there is no light source underwater. Underwater imaging systems must rely on artificial light sources for illumination. When light travels through water, it is severely attenuated by water absorption, reflection, and scattering. The collected underwater images inevitably have problems such as limited visible range, blur, low contrast, uneven illumination, incoherent colors, and noise. The purpose of image enhancement is to improve or solve one or more of the above problems in a targeted manner. Therefore, underwater image enhancement technology has become one of the key contents of underwater image processing technology research. In this paper, we proposed a conditional generative adversarial network model based on attention U-Net which contains an attention gate mechanism that could filter invalid feature information and capture contour, local texture, and style information effectively. Furthermore, we formulate an objective function through three different loss functions, which can evaluate image quality from global content, color, and structural information. Finally, we performed end-to-end training on the UIEB real-world underwater image dataset. The comparison experiments show that our method outperforms all comparative methods, the ablation experiments show that the loss function proposed in this paper outperforms a single loss function, and finally, the generalizability of our method is verified by executing on two different datasets, UIEB and EUVP.



Citation: Tang, P.; Li, L.; Xue, Y.; Lv, M.; Jia, Z.; Ma, H. Real-World Underwater Image Enhancement Based on Attention U-Net. *J. Mar. Sci. Eng.* **2023**, *11*, 662. <https://doi.org/10.3390/jmse11030662>

Academic Editor: Rafael Morales

Received: 21 February 2023

Revised: 17 March 2023

Accepted: 20 March 2023

Published: 21 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: underwater image; image enhancement; generative adversarial network

1. Introduction

Nowadays, people turn their eyes to the ocean because of the shortage of land resources. This has greatly promoted the study and exploration of the ocean, and the development of underwater resources. In order to perform relevant underwater tasks, some devices with visual sensors, such as underwater robots, have become the preferred tools and play an important role in underwater exploration. However, due to the diversity and complexity of the underwater environment, the directly captured visual information is usually mixed with a lot of noise, which makes it difficult to carry out object detection, sample capture, and other visual tasks. It has greatly hindered the process of ocean exploitation. Therefore, underwater image enhancement technology becomes particularly important.

Due to the degradation of underwater light and the different absorption rates of various colors of light by water, the underwater image always appears blue or green. At the same time, the underwater image is noisy to a certain extent because of the scattering effect of suspended particles on the light. Additionally, it gets more complicated as the depth, light, and suspended objects change in the water. Directly captured underwater images are always blurry and appear blue or green, such as in Figure 1. In order to obtain high-quality images in water, underwater image enhancement technology comes into being and plays an important role. With this technique, we can obtain clear and high-quality underwater images, as shown in Figure 1.

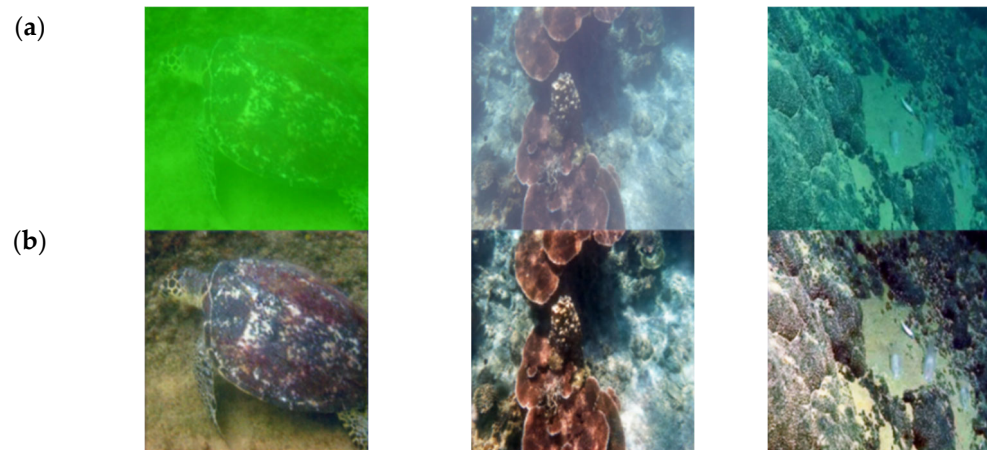


Figure 1. Example of underwater image. From top to bottom are (a,b). (a) Original image and (b) enhanced image by our method.

Many underwater image enhancement methods have been proposed in the past few years. They can be roughly divided into two categories according to whether they are based on deep learning technology. The first kind is the physical model-based method, which is mainly based on the physical model of underwater imaging, they achieve enhancement by eliminating the scattering effect of light and color correction in the imaging process, such as white balance adjustment [1], histogram equalization [2], and fusion technology [3]. However, these physical model-based algorithms usually use simplified underwater imaging models, which are inaccurate since they assume many parameters and cannot reflect the complexity of real underwater environments. With the changes in the water area, water depth, illumination, magazines, and other factors, the underwater imaging process becomes more complex, and it is difficult for such algorithms to achieve good generalization results. On the other hand, this kind of method generally needs a long time to obtain the enhanced image and cannot achieve the real-time visual enhancement task.

The second is deep learning-based methods. In recent years, with the improvement of hardware performance, artificial intelligence has ushered in a new wave. Deep learning [4] as its sub-branch also has produced a technological explosion, especially in computer vision and image processing tasks [5,6]. Those deep learning-based methods mainly use a large number of paired or unpaired underwater image datasets for training, such as the references [7–11]. However, most of these methods are trained on synthetic underwater image datasets, and the generalization of those methods in the real underwater environment is not very good. Additionally, some of these methods need to pre-process the dataset, so it is impossible to train end-to-end and achieve real-time image enhancement.

To tackle these issues, a novel generative adversarial network (GAN) based on the attention-gate (AG) mechanism [12] is proposed in this paper. Our proposed model can be trained end-to-end using public real-world underwater image datasets. The AG mechanism screens the important feature information from the original feature map and eliminates the irrelevant noise, allowing only the filtered feature information to be merged through the skip connection. Moreover, the AG mechanism filters the activation operation of

neurons during both forward and backward propagation, thereby enabling the efficient update of model parameters at a shallow level. To improve generalization performance, we combine three different loss functions to obtain a new objective function, which is used to train our model with a real-world underwater image dataset. Our proposed model, which incorporates the AG mechanism and the new objective function, achieves impressive performance in real-world underwater environments. The contributions of this paper can be summarized as follows:

1. We propose a generative adversarial network (GAN) for enhancing underwater images based on the attention-gate (AG) mechanism. The AG is integrated into the standard U-Net architecture to screen important feature information;
2. We formulate a new objective function and train our model end-to-end on a real-world underwater image dataset. Experiments demonstrate that our model outperforms several state-of-the-art methods in both qualitative and quantitative evaluations.

2. Related Works

In this section, we introduce some related work on generative adversarial networks and give some examples of applications based on these models.

2.1. Generative Adversarial Nets

In the development of deep learning, many generation models have appeared. One of the most widely used is Generative Adversarial Nets (GAN), proposed by Goodfellow et al. [13] in 2014. It introduced a novel way to train a generative model, which usually consists of two main parts. One is the generative model (G), which captures the data distribution and then generates a new object. Another is the discriminative model (D), which estimates whether a sample came from the training dataset or is generated by G. Both G and D networks usually consist of convolutional neural networks with nonlinear activation functions. To learn a generator distribution p_g over data x , the generator builds a mapping function from a prior noise distribution $p_z(z)$ to data space as $G(z, \theta_g)$. Additionally, the discriminator $D(x, \theta_d)$ outputs a single scalar representing the probability from training data rather than p_g . This is the training procedure for G to minimize $\log(1 - D(G(z)))$ and for D to minimize $\log(D(x))$. In other words, D and G play the following two-player:

$$\min_G \max_D \left(E_{x \sim P_{data}} [\log D(x)] + E_{z \sim P_z(z)} [\log(1 - D(G(z)))] \right) \quad (1)$$

2.2. Conditional Adversarial Nets

The object generated by the original GAN is highly uncertain, and it cannot generate the objects we expect. In order to solve this problem, Mirza et al. [14] proposed Conditional Generative Adversarial Nets (CGAN) in 2014. They perform the conditioning by feeding extra information y , which could be any kind of auxiliary information, into both the discriminator and generator as additional input. For the generator, noise $p_z(z)$ and y are combined as new inputs. Additionally, for the discriminator, y is combined with x as inputs. The objective function of CGAN can be summarized below:

$$\min_G \max_D (E_{x \sim P_{data}} [\log D(x, y)] + E_{z \sim P_z(z)} [\log(1 - D(G(z, y)))]) \quad (2)$$

As GAN and CGAN are proposed, there has a great impact on image generation. Many applications in image generation appeared, such as image style conversion [13,15], image generation [16,17], text-image conversion [18] and image enhancement [11,19].

3. Proposed Model

In this paper, we propose a new underwater image enhancement network based on conditional GAN. We use the attention U-Net, proposed by Oktay et al., [12] as the generator network baseline and the Patch-GAN, proposed by Isola et al., [5] as the discriminator

network baseline. Additionally, we combine three different loss functions to form the final objective function. We refer to our network as **AttU-GAN**.

3.1. Generator with Skip

Our generator network is a fully convolutional encoder–decoder structure that resembles a “U”-Net [20] due to the structural similarity between the input and output. The advantage of the “U”-Net is that it retains the original feature information extracted by the encoder and combines it with the reconstruction feature through the “skip connection” with an attention-gate mechanism. This mechanism filters out invalid feature information and effectively restores the structural information of the original image in the reconstructed image. The attention gate mechanism for the skip input x , previous layer input g , convolution kernel W_x and W_g , bias b_g and b_ψ , activation functions σ_1 and σ_2 , and attention coefficient α can be formulated as follows:

$$\alpha = \sigma_2 \left(\Psi^T \left(\sigma_1 \left(W_x^T x_{skip} + W_g^T g + b_g \right) \right) + b_\psi \right) \quad (3)$$

$$x_{att} = \alpha \cdot x_{skip} \quad (4)$$

The architecture of our G net is shown in Figure 2. The encoder consists of six identical down-sampling blocks with the following structure: Convolution layer with kernel size = 3×3 , stride = 1; Instance Normalization layer; GELU Activation layer; Dropout layer; and Max Pooling layer. The decoder consists of five identical up-sampling blocks and a final output block. The up-sampling block structure is as follows: Transpose Convolution layer with kernel size = 4×4 , stride = 2; Instance Normalization layer; GELU activation layer; and Dropout layer. In the final output block, Tanh was used as the active function.

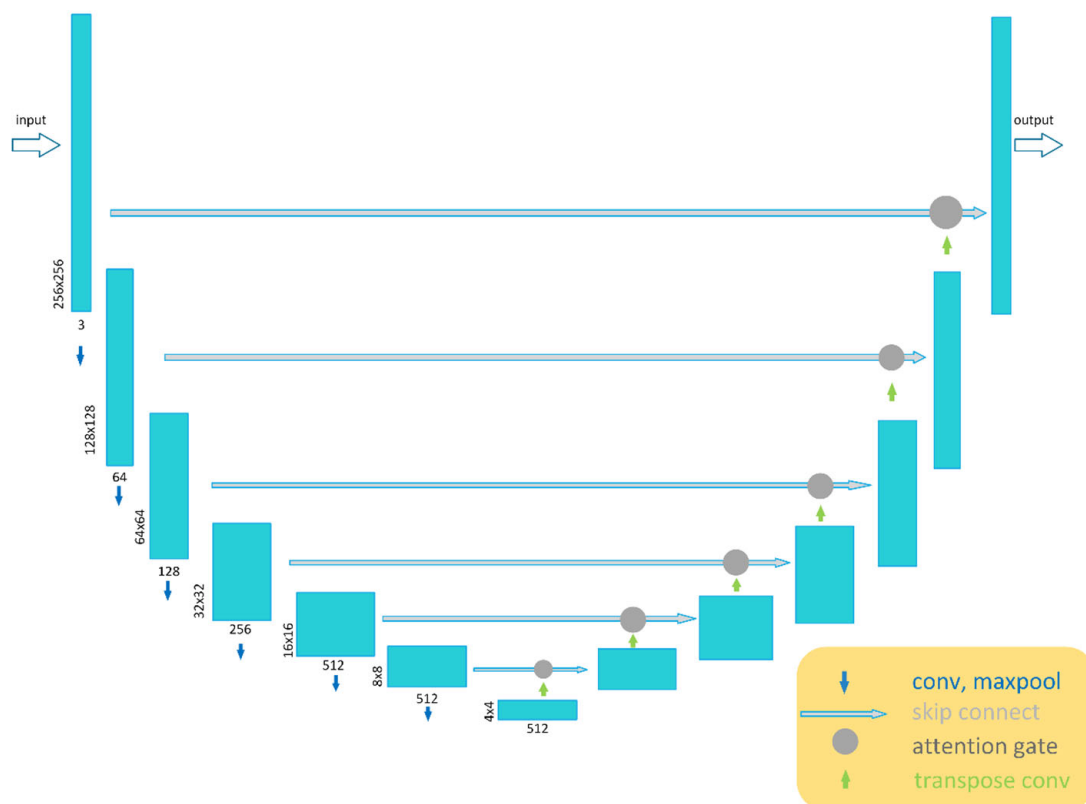


Figure 2. The generator architecture of our model is illustrated as follows. Each blue box represents a multi-channel feature map, and the number of channels is denoted at the bottom of the box. The xy size is provided at the lower left edge of the box.

3.2. Discriminator

We adopt the Patch-GAN [5], a Markovian discriminator, as the baseline for our discriminator. Unlike a regular discriminator that outputs a scalar value indicating whether an image is real or fake, the Patch-GAN outputs an $N \times N \times 1$ feature matrix as a patch-level judgment for an image. N can be set to a smaller value and still produce high-quality results, and a smaller Patch-GAN has fewer parameters, resulting in faster network computation and applicability to arbitrarily large images. For a 256×256 image, we set $N = 16$ and run the discriminator on the image, averaging all the responses to provide the final output of D .

3.3. Loss Function

One of the most difficulties in the training of the GAN model is the construction of the loss function. The loss function, also known as the objective function, is used to calculate the difference between the real value and the predicted value, and the optimizer is an important element of compiling a neural network model. Commonly used loss functions include MSE loss, L1 loss, and the perceptual loss recently proposed by Johnson et al. [21].

MSE Loss is the mean square error loss, also known as the quadratic loss and L2 loss, and is often used in regression prediction tasks. The mean squared error function measures how good a model is by calculating the square of the distance (i.e., the error) between the predicted and actual values. That is, the closer the predicted value is to the real value, the smaller the mean square error between the two. Its formula is as follows:

$$\text{loss}_{\text{MSE}}(x, y) = \frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2 \quad (5)$$

MAE Loss is the mean absolute loss, also known as L1 Loss, and is the absolute error as the distance. Its formula is as follows:

$$\text{loss}_{\text{MAE}}(x, y) = \frac{1}{n} \sum_{i=1}^n |y_i - f(x_i)| \quad (6)$$

Perceptual loss is recently proposed, which is a distance calculation method based on feature maps. It obtains a feature map of reference and enhanced images using a 16-Layer VGG network [22] pre-trained on the ImageNet dataset and then calculates feature reconstruction loss to control the feature similarity between the reference image and enhanced image.

The above-mentioned loss functions have their own strengths and limitations. Given the complexity of real-world underwater datasets, it is desirable to combine them to obtain a more balanced loss function that can effectively and universally fit the dataset. Each loss function has a unique calculation method and plays a distinct role in network optimization. To achieve balance in network optimization, we combine multiple loss functions with a weighted sum to obtain the final loss function. Since each loss function is calculated on a different scale, the weighting coefficient varies in cardinality. Therefore, we perform a weighted sum of these three loss functions to obtain the final loss function, which is shown below:

$$\text{loss} = \alpha_1 \cdot \text{loss}_{\text{MSE}} + \alpha_2 \cdot \text{loss}_{\text{MAE}} + \alpha_3 \cdot \text{loss}_{\text{perceptual}} \quad (7)$$

Through our repeated experiments, we finally concluded that when $\alpha_1 = 0.15$, $\alpha_2 = 10$, $\alpha_3 = 0.8$ in Equation (7), the best results can be achieved.

4. Experimental Results and Analysis

In this section, we will introduce the details of the experiments, including data sets, experimental environment, experimental methods and experimental results. Additionally, we analyze and explain the experimental results.

4.1. Datasets

We use the Underwater Image Enhancement Benchmark (UIEB) dataset proposed by Li et al. [7] to train and test our model. The UIEB dataset includes 950 real-world underwater images, 890 of which have corresponding reference images. The original images of the dataset were captured from different real-world underwater environments. For the reference images, 12 different enhancement methods were used to generate candidate images, and then multiple volunteers voted for each pair of enhanced images, and the image with the highest number of votes was used as the corresponding reference image. We divided the 890 images into 3 parts: 710 images as the training set, 90 images as validation sets, and 90 images as test sets.

4.2. Experimental Environment

In order to evaluate our model objectively and effectively, we conducted qualitative and quantitative comparison experiments with several of the most advanced methods. They include Fusion based [3], Statistical based [23], UGAN [8] and FUnIE [9]. We use the source codes provided by the authors, or according to the description of the original papers, or refer to open-source resources to build the models. All the methods used in the experiment adopted the same testing and training process and were carried out in the following environment, AMD Ryzen 7 5800H CPU, NVIDIA GeForce RTX 3060 Laptop 6GB GPU, and 16GB DDR4 RAM.

4.3. Evaluations

In this part, we conducted both qualitative and quantitative experiments. The qualitative experiment served as a supplement to express subjective evaluation results. Additionally, the quantitative experiment served as the main indicator to show the objective evaluation results.

4.3.1. Subjective Evaluation

We selected images of different underwater environments from 90 test images and tested them using the above methods. The used real-world underwater images have diverse tones, light, and contrast. The comparison results with competitive methods are shown in Figure 3. A first glance at the results may give the impression that the comparison of Fusion [3] is stronger and better. However, after careful observation, it will be found that Fusion only has an obvious comparison with the pictures with relatively regular noise distribution, such as Figure 3d. However, in another picture in Figure 3, you can see that the effect is minimal. The reason why is that the physical model approach used in Fusion is not always valid. Similarly, the images enhanced by Statistics [23] are not natural, and the color is too bright. The disadvantage of UGAN [8] and FUnIE [9] is that they do not have good fitting and generalization ability for the real-world underwater image dataset, which makes their test results obviously inadequate. In contrast, our results show good generalization ability on all the test set, and the comparison with the other images shows that our results are more realistic in color and more accurate in detail reduction.



Figure 3. Subjective comparisons on real-world underwater images. From top to bottom are (a–g). From left to right are raw underwater images, the results of Fusion based [3], Statistics based [23], UGAN [8], FUnIE [9] and Ours.

In addition, we performed a comparative experiment using the DEA-Net algorithm for image dehazing, which was accepted at CVPR2023 [24]. The pre-training weights provided by the authors were used to test the algorithm, and the results of the experiment on the UIEB dataset are presented in Figure 4. While some images showed a denoising effect, the overall performance of the algorithm was not as good as our method, and we

use two underwater image quality evaluation metrics, UCIQE and UIQM, to evaluate the DEA-Net test results, as detailed in Section 4.3.2.

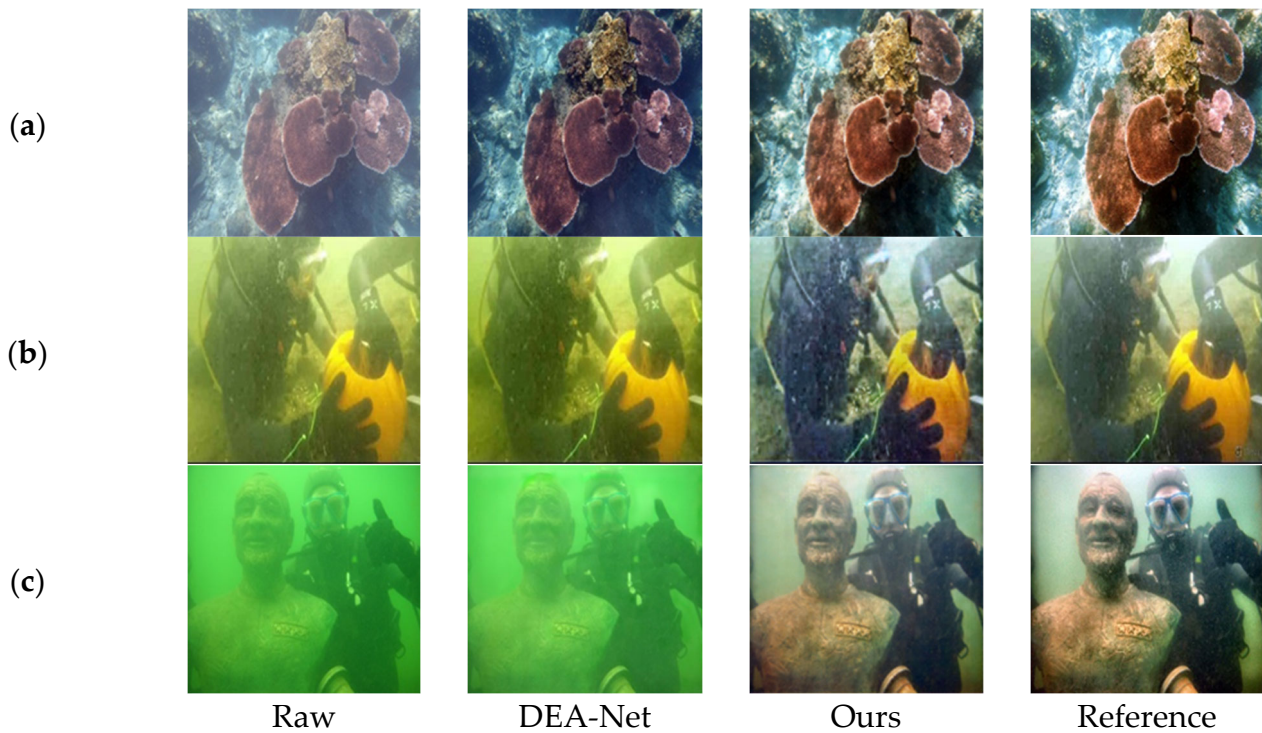


Figure 4. Comparison of DEA-Net dehazing model with Our model on UIEB dataset.

4.3.2. Objective Evaluation

Objective evaluation is mainly divided into two categories: one is full reference, which requires a comparison between the original image and the corresponding reference image; the other is non-reference, which only needs an enhanced image to be evaluated. All test data in the following table are averaged on the test set. We mark the best performer in **red**, the second best in **blue**, and the reference result in **black**.

We chose Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM) as full-reference image quality evaluation indicators. PSNR is an image quality evaluation index based on the error between corresponding pixels calculated by MSE. The larger the value is, the smaller the image distortion. For images I and K , it can be formulated as follows:

$$\text{PSNR}(I, K) = 10 \log_{10} \frac{255^2}{\text{MSE}(I, K)} \quad (8)$$

SSIM is another image quality evaluation index, which measures image similarity from brightness, contrast, and structure. When SSIM calculates the difference between two images at each position, it does not take a pixel from each of the two images at that position but takes a pixel from each region. The higher the SSIM value is, the closer the reference image is to the original image in structure, and $\text{SSIM} \leq 1$. It is defined as below:

$$\text{SSIM} = \frac{(2\mu_X\mu_Y + c_1)(\sigma_{XY} + c_2)}{(\mu_X^2 + \mu_Y^2 + c_1)(\sigma_X^2 + \sigma_Y^2 + c_2)} \quad (9)$$

Table 1 shows the evaluation results of different methods. All the results are computed in the YCrCb color space, using methods provided by Python's third-party library scikit-image, as can be seen from the table. In PSNR, our method achieved the best result, even beyond the reference. Additionally, in SSIM, our method achieved the second best but was very close to the best one. Across the board, our method achieved the best results.

Table 1. Full-reference image quality evaluation.

Method	PSNR (dB)	SSIM
Fusion	20.709	0.886
Statistic	20.466	0.825
FUnIE	21.119	0.787
UGAN	20.734	0.856
AttU-GAN	21.852	0.875

We chose underwater color image quality evaluation (UCIQE) [25] and underwater image quality measure (UIQM) [26,27] as non-reference image quality evaluation indicators. UCIQE evaluates underwater image quality by color density, saturation, and contrast. The result is shown in Table 2.

Table 2. No-reference image quality evaluation (UCIQE).

Method	UCIQE
Fusion	0.926
Statistic	0.713
FUnIE	0.782
UGAN	0.891
DEA-Net	0.649
AttU-GAN	0.936

UIQM is a comprehensive underwater image evaluation index, which is the weighted sum of the underwater image colorfulness measure (UICM), underwater image sharpness measure (UISM), and underwater image contrast measure (UIConM). Its formula is as follows:

$$\text{UIQM} = c_1 \cdot \text{UICM} + c_2 \cdot \text{UISM} + c_3 \cdot \text{UIConM} \quad (10)$$

According to the original paper, we set the parameters as follows: $c_1 = 0.0282$, $c_2 = 0.2953$, and $c_3 = 3.5753$. Results are shown in Table 3. We found that our method achieved the best results on UCIQE and UICM and the second best on UISM and UIQM. Of all the results, three exceeded the reference. So, our method also achieved the best results overall.

Table 3. No-reference image quality evaluation (UIQM).

Method	UICM	UISM	UICONM	UIQM
Fusion	5.271	6.140	0.278	2.957
Statistic	4.482	5.646	0.208	2.537
FUnIE	5.375	6.962	0.246	3.096
UGAN	5.363	6.658	0.251	3.015
DEA-Net	3.420	5.328	0.236	2.513
AttU-GAN	6.587	6.839	0.237	3.053

4.3.3. Ablation Experiments

In the above, we proposed a novel method: training the model by weighting and summing multiple loss functions. In order to verify the effectiveness of the method, we will use different loss functions to compare the ablation experiments below. In the case of keeping other parameters and modules unchanged, only a single loss function is used to train the model. We will analyze the training results using both subjective evaluation (as shown in Figures 5 and 6) and objective evaluation (as shown in Tables 4–6).

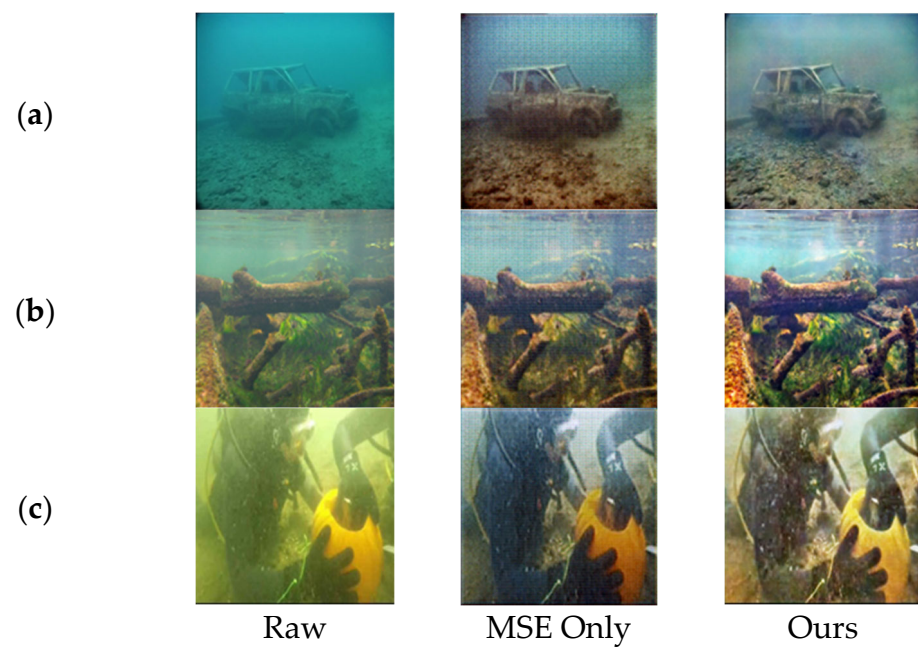


Figure 5. Subjective comparisons on real-world underwater images. From left to right are raw underwater images, the results of using MSE loss function only and using our proposed loss function.

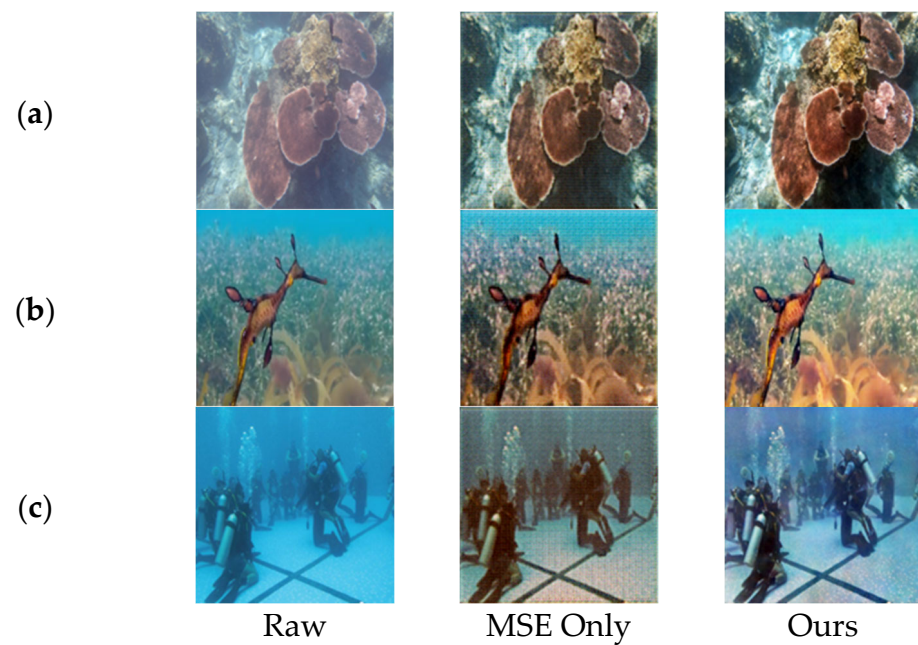


Figure 6. Subjective comparisons on real-world underwater images. From left to right are raw underwater images, the results of using MAE loss function only and using our proposed loss function.

Table 4. Full-reference image quality evaluation.

Method	PSNR (dB)	SSIM
MSE Only	19.760	0.807
MAE Only	21.816	0.870
AttU-GAN	21.852	0.875

Table 5. No-reference image quality evaluation (UCIQE).

Method	UCIQE
MSE Only	0.928
MAE Only	0.713
AttU-GAN	0.936

Table 6. No-reference image quality evaluation (UIQM).

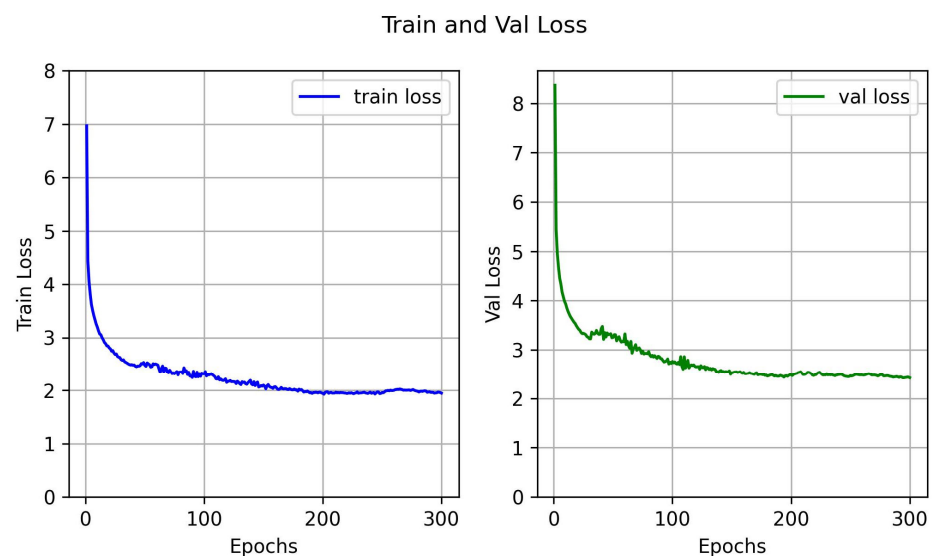
Method	UICM	UISM	UICONM	UIQM
MSE Only	6.794	6.396	0.218	2.860
MAE Only	6.454	6.802	0.241	3.051
AttU-GAN	6.587	6.839	0.237	3.053

From the above experimental results, it is not difficult to see that the combination of multiple loss functions proposed by us is better than a single loss function, whether it is subjective evaluation or objective evaluation. In terms of objective evaluation, the result of a single loss function still has some noise, but the image generated by our development is more accurate and complete. In terms of objective evaluation, it can be seen intuitively that among all seven indicators, we achieved the five best results.

4.3.4. Generalizability Verification

In this section, we will validate the generalization capability of our method on the EUVP [9] dataset.

EUVP's test set contains 515 paired real-world underwater images. We will use the AttU-GAN model trained on the UIEB dataset and other compared methods to test on the EUVP test set to verify the generalization of our method. Additionally, the training loss of our model on the UIEB dataset is shown in Figure 7; our model has a good fitting ability. The test results in EUVP are shown in Figure 8. From the figures, we can see that our method still has excellent enhancement results on unfamiliar data sets, which proves that our method has a strong generalization ability.

**Figure 7.** Training loss and validation loss of our model on the UIEB dataset.

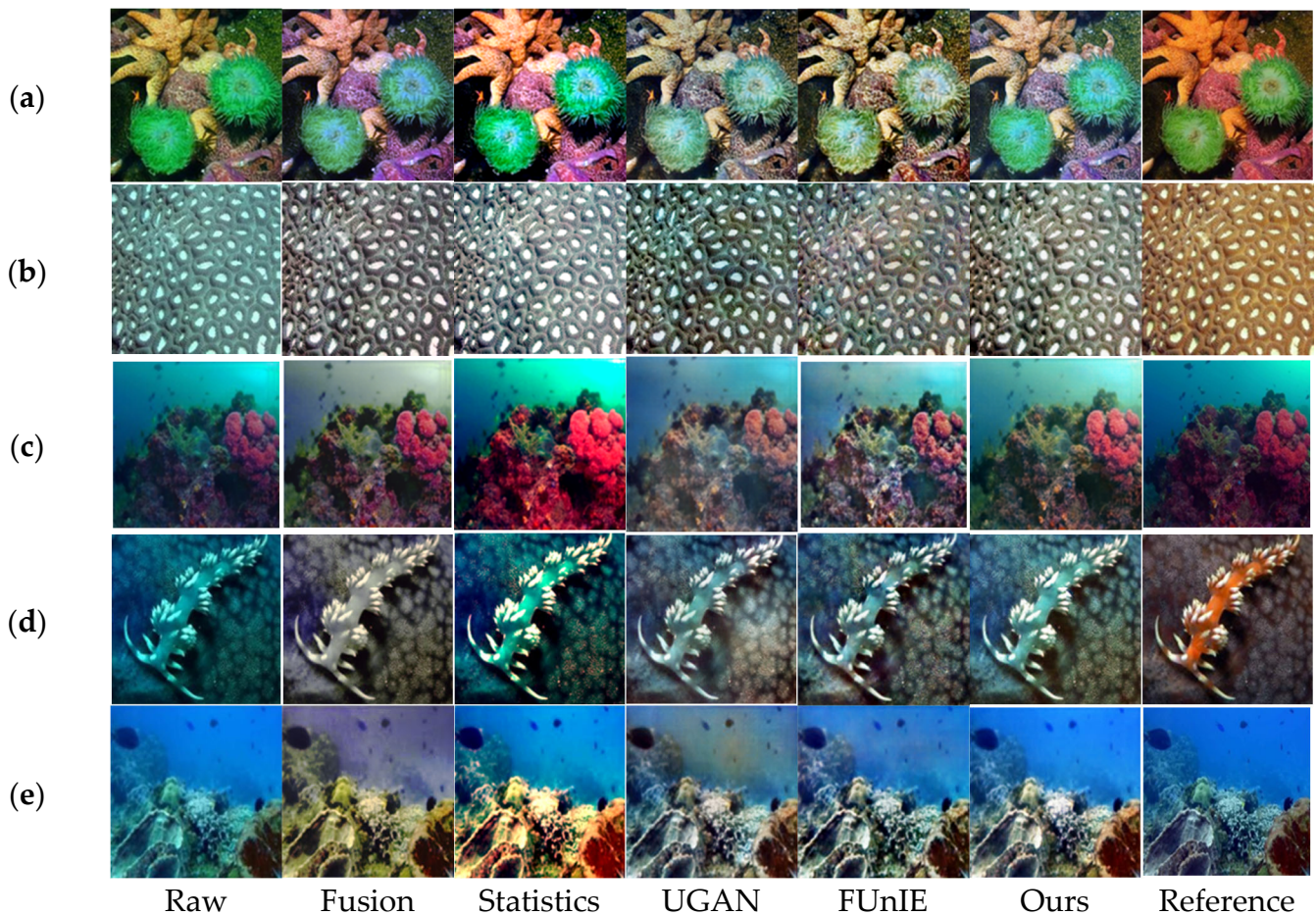


Figure 8. Test results of our method and other comparative methods on the EUVP dataset.

5. Conclusions

We present a general and effective model for underwater image enhancement in this paper. Our model filters the noise in the images and effectively retains the important feature information during the reconstruction process to generate high-quality enhanced images. We construct a new multi-loss function weighted objective function and train our model on a real-world underwater image dataset, UIEB, and validate the effectiveness and generalization of our method on two different datasets, UIEB and EUVP. The comparison experiments show that our method outperforms all the compared methods, the ablation experiments show that the loss function proposed in this paper outperforms a single loss function, and finally, the generalization of our method is verified by testing it on different datasets. In our future work, we will focus on improving the performance of the network to make it faster, lighter, and stronger.

Author Contributions: Funding acquisition, H.M.; writing—original draft, P.T.; writing—review and editing, L.L., Y.X., M.L. and Z.J. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Shanghai Aerospace Science and Technology Innovation Fund under Grant No. SAST2019-048, and the Cross-Media Intelligent Technology Project of Beijing National Research Center for Information Science and Technology (BNRist) under Grant No. BNR2019TD01022.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jobson, D.J.; Rahman, Z.; Woodell, G.A.A. Multiscale Retinex for Bridging the Gap Between Color Images and the Human Observation of Scenes. *IEEE Trans. Image Process.* **1997**, *6*, 965–976. [[CrossRef](#)] [[PubMed](#)]
2. Pizer, S.M.; Johnston, R.E.; Ericksen, J.P.; Yankaskas, B.C.; Muller, K.E. Contrast-limited Adaptive Histogram Equalization: Speed and Effectiveness. In Proceedings of the First Conference on Visualization in Biomedical Computing, Atlanta, GA, USA, 22–25 May 1990; pp. 337–345.
3. Ancuti, C.O.; Ancuti, C.; De Vleeschouwer, C.; Bekaert, P. Color Balance and Fusion for Underwater Image Enhancement. *IEEE Trans. Image Process.* **2018**, *27*, 379–393. [[CrossRef](#)] [[PubMed](#)]
4. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
5. Isola, P.; Zhu, J.; Zhou, T.; Efros, A.A. Image-to-Image Translation with Conditional Adversarial Networks. *IEEE Conf. Comput. Vis. Pattern Recognit.* **2017**, *2017*, 1125–1134.
6. Fan, Y.S.; Niu, L.H.; Liu, T. Multi-Branch Gated Fusion Network: A Method That Provides Higher-Quality Images for the USV Perception System in Maritime Hazy Condition. *J. Mar. Sci. Eng.* **2022**, *10*, 1839. [[CrossRef](#)]
7. Li, C.Y.; Guo, C.L.; Ren, W.Q.; Cong, R.M.; Hou, J.H.; Kwong, S.; Tao, D.C. An Underwater Image Enhancement Benchmark Dataset and Beyond. *IEEE Trans. Image Process.* **2019**, *29*, 4376–4389. [[CrossRef](#)]
8. Fabbri, C.; Islam, M.J.; Sattar, J. Enhancing Underwater Imagery Using Generative Adversarial Networks. *IEEE Int. Conf. Robot. Autom.* **2018**, *2018*, 7159–7165.
9. Islam, M.J.; Xia, Y.; Sattar, J. Fast Underwater Image Enhancement for Improved Visual Perception. *IEEE Robot. Autom. Lett.* **2020**, *5*, 3227–3234. [[CrossRef](#)]
10. Li, C.Y.; Anwar, S.; Porikli, F. Underwater Scene Prior Inspired Deep Underwater Image and Video Enhancement. *Pattern Recognit.* **2020**, *98*, 107038. [[CrossRef](#)]
11. Wang, Y.D.; Guo, J.C.; Gao, H.; Yue, H.H. UIEC²-Net: CNN-based Underwater Image Enhancement Using two Color Space. *Signal Process. Image Commun.* **2021**, *96*, 116250. [[CrossRef](#)]
12. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M. Attention U-Net: Learning Where to Look for the Pancreas. *arXiv* **2018**, arXiv:1804.03999.
13. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M. Generative Adversarial Networks. *Commun. ACM* **2020**, *63*, 139–144. [[CrossRef](#)]
14. Mirza, M.; Osindero, S. Conditional Generative Adversarial Nets. *arXiv* **2014**, arXiv:1411.1784.
15. Gatys, L.A.; Ecker, A.S.; Bethge, M. Image Style Transfer Using Convolutional Neural Networks. *IEEE Conf. Comput. Vis. Pattern Recognit.* **2016**, *2016*, 2414–2423.
16. Jin, Y.H.; Zhang, J.K.; Li, M.J. Towards the Automatic Anime Characters Creation with Generative Adversarial Networks. *arXiv* **2017**, arXiv:1708.05509.
17. Karras, T.; Aila, T.; Laine, S.; Lehtinen, J. Progressive Growing of GANs for Improved Quality, Stability, and Variation. *arXiv* **2018**, arXiv:1710.10196.
18. Reed, S.; Akata, Z.; Yan, X.C.; Logeswaran, L. Generative Adversarial Text to Image Synthesis. *Int. Conf. Mach. Learn.* **2016**, *48*, 1060–1069.
19. Liu, R.S.; Ma, L.; Zhang, J.A.; Fan, X.; Luo, Z.X. Retinex-inspired Unrolling with Cooperative Prior Architecture Search for Low-light Image Enhancement. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 10556–10565.
20. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Med. Image Comput. Comput. Assist. Interv.* **2015**, *8*, 234–241.
21. Johnson, J.; Alahi, A.; Li, F.F. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In Proceedings of the European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 11–14 October 2016; pp. 694–711.
22. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
23. Song, W.; Wang, Y.; Huang, D.; Liotta, A.; Perra, C. Enhancement of Underwater Images with Statistical Model of Background Light and Optimization of Transmission Map. *IEEE Trans. Broadcast.* **2020**, *66*, 153–169. [[CrossRef](#)]
24. Chen, Z.; He, Z.; Lu, Z.M. DEA-Net: Single image dehazing based on detail-enhanced convolution and content-guided attention. *arXiv* **2023**, arXiv:2301.04805.
25. Yang, M.; Sowmya, A. An Underwater Color Image Quality Evaluation Metric. *IEEE Trans. Image Process.* **2015**, *24*, 6062–6071. [[CrossRef](#)] [[PubMed](#)]
26. Panetta, K.; Gao, C.; Agaian, S. Human-Visual-System-Inspired Underwater Image Quality Measures. *IEEE J. Ocean. Eng.* **2016**, *41*, 541–551. [[CrossRef](#)]
27. Huang, X.; Belongie, S. Arbitrary Style Transfer in Real-Time with Adaptive Instance Normalization. *IEEE Int. Conf. Comput. Vis.* **2017**, *2017*, 1510–1519.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.