

Article

# FE-GAN: Fast and Efficient Underwater Image Enhancement Model Based on Conditional GAN

Jie Han <sup>1</sup>, Jian Zhou <sup>1,2,\*</sup>, Lin Wang <sup>1</sup>, Yu Wang <sup>1</sup> and Zhongjun Ding <sup>3</sup><sup>1</sup> School of Information Engineering, Wuhan University of Technology, Wuhan 430070, China<sup>2</sup> Hubei Key Laboratory of Broadband Wireless Communication and Sensor Networks, Wuhan 430070, China<sup>3</sup> National Deep Sea Center, Qingdao 266237, China

\* Correspondence: jianzhou@whut.edu.cn

**Abstract:** The processing of underwater images can vastly ease the difficulty of underwater robots' tasks and promote ocean exploration development. This paper proposes a fast and efficient underwater image enhancement model based on conditional GAN with good generalization ability using aggregation strategies and concatenate operations to take full advantage of the limited hierarchical features. A sequential network can avoid frequently visiting additional nodes, which is beneficial for speeding up inference and reducing memory consumption. Through the structural re-parameterization approach, we design a dual residual block (DRB) and accordingly construct a hierarchical attention encoder (HAE), which can extract sufficient feature and texture information from different levels of an image, and with 11.52% promotion in GFLOPs. Extensive experiments were carried out on real and artificially synthesized benchmark underwater image datasets, and qualitative and quantitative comparisons with state-of-the-art methods were implemented. The results show that our model produces better images, and has good generalization ability and real-time performance, which is more conducive to the practical application of underwater robot tasks.

**Keywords:** underwater image enhancement; generative adversarial networks; real-time application



**Citation:** Han, J.; Zhou, J.; Wang, L.; Wang, Y.; Ding, Z. FE-GAN: Fast and Efficient Underwater Image Enhancement Model Based on Conditional GAN. *Electronics* **2023**, *12*, 1227. <https://doi.org/10.3390/electronics12051227>

Academic Editor: Yu-Chen Hu

Received: 2 February 2023

Revised: 17 February 2023

Accepted: 28 February 2023

Published: 4 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/>).

## 1. Introduction

With the research and application of AUVs (autonomous underwater vehicles) and ROVs (remote operated vehicles), ocean exploration has achieved many breakthrough results. Over the last several years, many AUVs and ROVs have been applied to ship hull inspection, underwater target detection and tracking [1,2], pipeline leak detection [3], underwater cable inspection [4], and mineral prospecting. The preeminent underwater image contains a wealth of semantic and construction information, which can help the AUV to complete tasks, such as classification and recognition, efficiently and accurately [5]. However, due to the nonlinear attenuation of light caused by underwater particles, it is difficult to obtain high-resolution underwater images despite the use of high-pixel cameras. Therefore, image enhancement algorithms that can effectively enhance image perception and statistical quality have become an urgent need. Its application can vastly promote the development of ocean exploration.

Natural light is absorbed and scattered when propagating in seawater. Usually, red light with the longest wavelength is absorbed the fastest, and the propagation distance is the shortest. The green and blue light with a shorter wavelength will travel farther [6] but will be affected by the scattering and refraction of particles in the sea during the diffusion process. Based on this situation, images taken in shallow waters with natural light usually show a blue-green color. In the deeper sea (over 60 m), there is no natural light at all [7]. Limited by the characteristics of natural light spreading, underwater images have problems, such as color distortion, deviation, and uneven brightness [8]. We use image enhancement technology to meet these challenges. Specifically, the raw image needs to be defogged and

deblurred to make the color hue consistent with the ground truth, thereby improving the image quality and highlighting useful information.

Many scholars have carried out in-depth research on the scattering phenomenon of light propagating in the medium. Through the inversion of this process, the distorted images (fogging, blurring, color unevenness, etc.) captured are operated to obtain the clear images as the desired output [9]. Some research groups have applied part of these models to underwater image-processing tasks and have made some progress. However, the attenuation parameter changes nonlinearly during the propagation process, so it is tricky to estimate [10]. Specific parameters set in the model limit it only to be used in given scenarios, resulting in insufficient robustness, so they do not satisfy the expected performance. The methods based on deep learning, especially the rapid development of CNN and GAN, provide scholars with another research idea. Some underwater image enhancement models based on CNN and GAN train paired and unpaired images, extract local and global image features, and optimize transmission maps. These models complete the image enhancement process by generating the target image output by the generator of the network, which improves the image contrast and color. Significant results have been achieved in enhancement. Nevertheless, since real underwater image datasets are difficult to obtain directly, many models are trained on artificially synthesized datasets, which usually cannot cope with different types and have insufficient generalization ability. Considering the requirements of underwater equipment for computing power and real-time processing, there is still room for a model with high efficiency and strong generalization performance.

This article proposed an underwater image enhancement model FE-GAN (fast and efficient generative adversarial network) to solve these problems. A hierarchical attention encoder (HAE) can extract deeper features and texture information, while preserving the overall structure of the image. Different loss functions based on texture and content are combined with weights to constrain the generator and discriminator. Through the learning of paired images, FE-GAN achieved end-to-end underwater image enhancement, which effectively improved the image quality. Results on different datasets prove that the model also has good generalization ability. The main contributions of this paper are as follows:

- We present a hierarchical attention encoder (HAE) to fully extract texture detail information, and a dual residual block (DRB) can more efficiently utilize residual learning to accelerate network inference.
- Through structural re-parameterization, we equate complex modules to simple convolutional layers, which accelerates the model during inference while maintaining a good enhancement effect.
- Based on HAE and DRB, we construct a fast and efficient underwater image enhancement network. Experiments on different datasets show that the enhanced image can achieve higher PSNR and SSIM values, and the mAP value also achieved significant results in the object detection task.

## 2. Related Work

### 2.1. Physical-Based Methods

As a crucial processing technology in the field of computer vision, image enhancement can purposefully emphasize the holistic or partial characteristics of an image. It can also expand the difference between the features of different objects in the image, improve the image quality, enrich the amount of information, and strengthen the recognition effect. The early underwater imaging model was presented by Ref. [11], where it is supposed that the scattering coefficient of each color channel remains constant within the camera's sensitivity range. Later, the initial model was deeply simplified [12,13], thinking that the attenuation coefficients in all color channels are consistent, and applying the simplified model to underwater image restoration tasks. Inspired by these models, some researchers assigned a value to the attenuation coefficient of each color channel, respectively, in the model. Berman et al. [14] proposed a method to set different spectral profiles for relevant water types. Spier et al. [15] adopted an approach for estimating the medium properties

using only images of backscattered light from the system's light sources. Their experimental results show that the refined model can achieve a better effect.

Derya Akkaynak et al. [9] proposed the latest underwater imaging model; they attributed the scattering of light to the direct transmission attenuation and the backscatter attenuation. They are guided by scene reflectance and the spectrum of ambient light, which is advantageous to color reconstruction and image enhancement. Combined with this new model, the author of [16] proposed an underwater scene depth estimation method based on image blur and light absorption, which can be used in image formation models to restore and enhance underwater images. The refined imaging model will cause a vast increase in the parameters that need to be estimated, and the computational complexity will also increase sharply. Moreover, the physical-based method model is too singular to adapt to a variety of complex underwater scenes. Thus, the generalization and real-time ability of the application are greatly restricted.

## 2.2. Learning-Based Methods

In recent years, deep learning gradually occupied a leading position in the field of computer vision with its high plasticity and universality. Inspired by this trend, some scholars proposed to use the computing power of convolutional neural networks to calculate the parameters that need to be estimated in the physical imaging model [16]. Then the deep learning methods and physical models were combined for underwater image enhancement research. Liu et al. [17] proposed an integration method based on the Akkaynak–Trebitz physical model [9]. Specifically, the physical model guides network learning and the network model design for components and coefficients estimation. Liu et al. [18] introduced VDSR (very-deep super-resolution reconstruction) into underwater resolution applications and proposed an underwater ResNet model for enhancement tasks. This model made good progress in automatic color enhancement, dehazing, and contrast adjustment. Qi et al. [19] proposed UICoE-Net, which introduced correlation feature matching units to provide rich complementary information for a mutual enhancement between images in the same scene.

The emergence of the GAN (generative adversarial network) opened up another path for image enhancement issues. However, the training process of GAN is usually unstable. Refs. [20–22] proposed different loss functions to constrain the discriminator, which effectively alleviates this situation. The conditional GAN proposed by the author of [23] is suitable for image-to-image translation tasks and achieves the mission of using images as conditions and generating corresponding output images.

For the existing synthetic and real underwater image datasets, many GAN-based methods have been proven to have achieved good results in underwater image enhancement. Cycle-GAN [24] and Dual-GAN [25] learn the mutual mapping between the two domains through the “cycle consistency loss” from unpaired data, ensuring that the network can complete the image generation task through unpaired image datasets. IPMGAN [17] proposed a new physical model integration network framework for underwater image enhancement. Through network training, the parameters and coefficients of the image degradation model are learned to reconstruct clear underwater images. FUnIE-GAN [26] proposed a real-time underwater image enhancement model based on fully convolutional conditional GAN, and formulated a multi-modal objective function. Experiments proved that the model achieved good results in both image enhancement and underwater human posture estimation. Zhou et al. [27] proposed a domain-adaptive learning framework, embedding a domain adaptive mechanism to eliminate the domain gap. UW-GAN [28] processes the single input image with a coarse-level network, then concatenates the result with the input image, and sends it to a fine-level network for a generation.

The application scenarios of most existing models are still very restricted, and it is rare to achieve good results in both real and synthetic underwater image datasets. Considering that image enhancement can be applied to the actual scene of underwater robots in the future, real-time performance is an indispensable part of model testing. We will explain

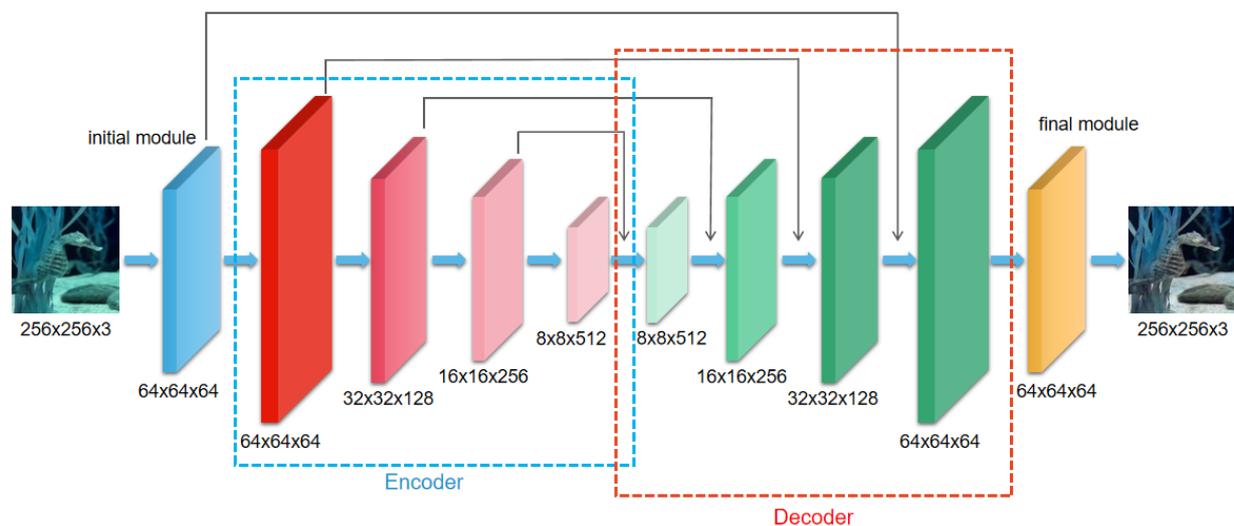
the results of our model in terms of generalization ability and real-time testing in the following section.

### 3. Proposed Model

#### 3.1. Overall Architecture

In image-related tasks, the generator of GAN receives a random noise  $z$ , then generates an image  $G(z)$ ; the discriminator accepts  $G(z)$  to determine if the image is authentic ( $D(x) = 1$ ) or not ( $D(x) = 0$ ) [29]. Therefore, in the training process of GAN, the goal of generator  $G$  is to generate images as real as possible to deceive discriminator  $D$ , and the target of  $D$  is to distinguish whether the images generated by  $G$  are real or not. In the optimal state,  $G$  generates a picture  $G(z)$  that is sufficient to “make it fake”, that is,  $D(G(z)) = 0.5$  [30]. When the model converges, the images, both generated and real, have the same distribution.

As shown in Figure 1, we proposed a generative network  $G$ , which bypasses the output of each encoder  $e_i$  to the input of its mirror decoder  $d_i$ . This aggregated connection can fuse limited hierarchical feature information, making the decoder use fewer parameters to ensure the efficiency of the network. The encoder is preceded by an initial module which consists of a convolutional layer with a kernel size of  $4 \times 4$ , and a spatial maximum pooling in a  $3 \times 3$  area with a stride of 2. This initial module can effectively remove some redundant information by reducing the dimension, maintaining the invariance of features, and also ensuring the integrity of the generated image. The decoder is followed by a final module, which performs zero padding and upsampling operation on the  $64 \times 64$  images twice. After that, the final output image is transferred to the discriminator along with the ground truth image. The Markov discriminator [23] is adopted in our model.

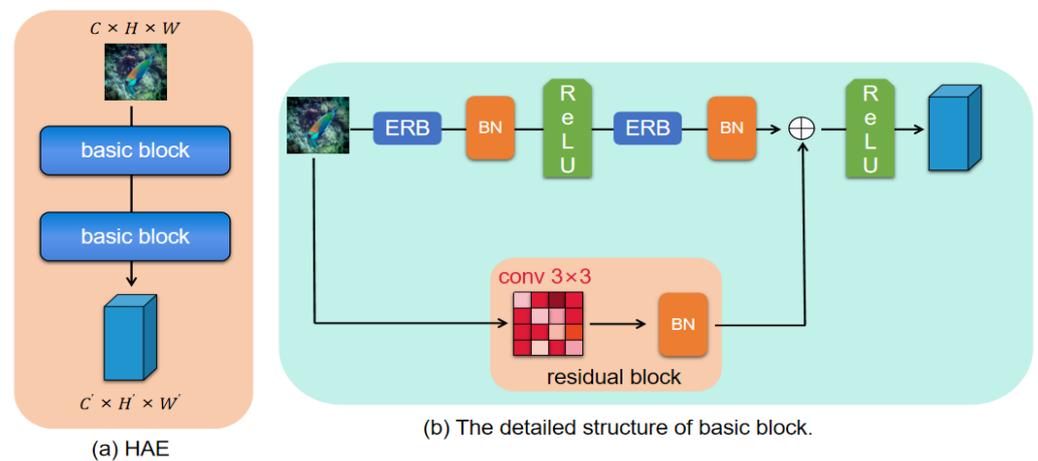


**Figure 1.** Generator of the FE-GAN.

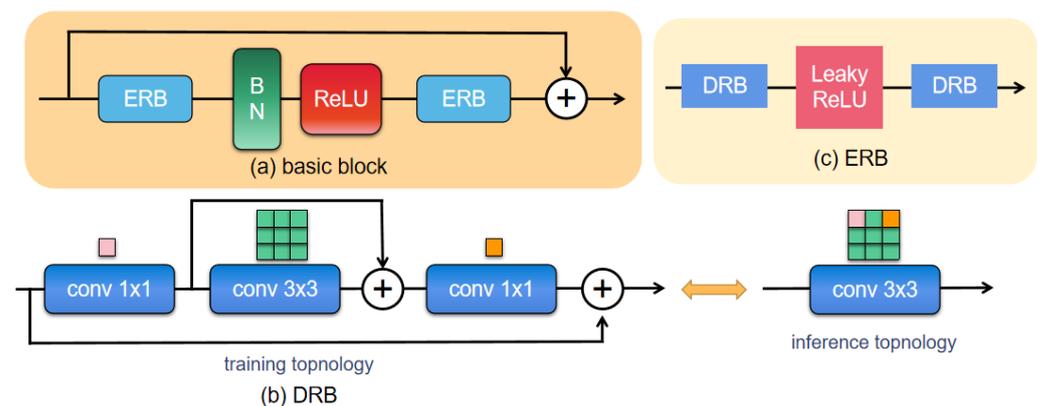
##### 3.1.1. Generator

The generator adopts the information multi-distillation module method to fuse the information of the encoder and its mirror decoder, improve the feature representation via the attention mechanism, and aggregate the hierarchical features. The details of the hierarchical attention encoder (HAE) are shown in Figure 2. Each encoder contains two basic modules composed of the ERB and residual block. To construct the relationship between local signals, we stack the ERB and ReLU operator. Batch normalization (BN) can capture the global context during training while incorporating it into convolutions during inference [31]. Although the dense residual block will help the generator to capture global information in the spatial dimension and adaptively focus on discriminative information, the skip connection introduces additional memory consumption and slows down inference. The enhanced residual block (ERB) [31,32] based on structural re-parameterization can

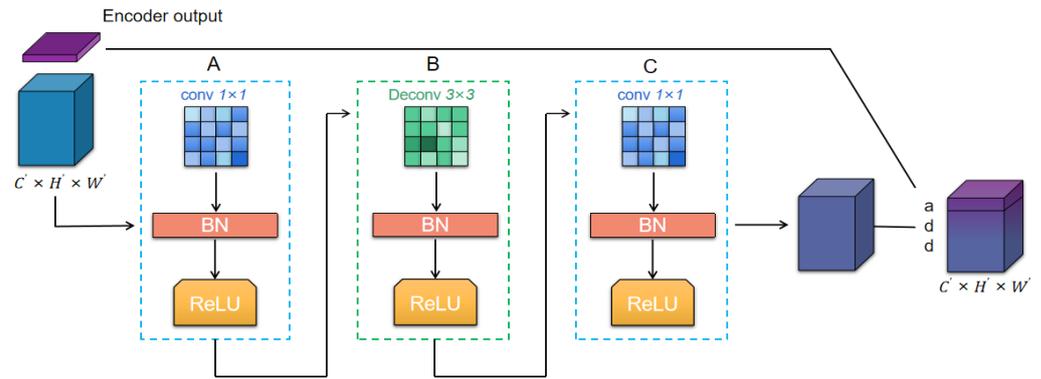
effectively solve this problem. As shown in Figure 3, ERB is composed of two dual residual blocks (DRB) in series, which are applied for deep feature learning. The residual module can provide a more stable gradient during the backpropagation of the training and avoid gradient disappearance. During inference, we equate DRB to a  $3 \times 3$  convolution layer to reduce memory consumption and accelerate the model. Ultimately, the encoder learns 512 feature maps of size  $8 \times 8$ . The decoder ( $d_1, d_2, d_3$ , and  $d_4$ ) uses a convolution and deconvolution structure in series, as shown in Figure 4. This Conv-Deconv module has been proven to be effective in super-resolution tasks [33], which can minimize the potential artifacts generated during the upsampling process. Meanwhile, to reduce the computational complexity and increase the weight of discriminant information, the number of channels in  $B$  is reduced to  $1/4$  of the original. This multi-stage convolutional connection facilitates gradient propagation and assists in training deeper networks. Features from multiple levels can be integrated using cascading mechanisms at local and global levels. The hierarchical pyramid structure can reconstruct residuals of different resolutions, and the local and global cascading mechanism also contributes to boosting expressive ability.



**Figure 2.** The overall structure of the HAE. (a) The connection structure of the HAE, which consists of two basic blocks in series. (b) The specific connection of each network layer in basic block.



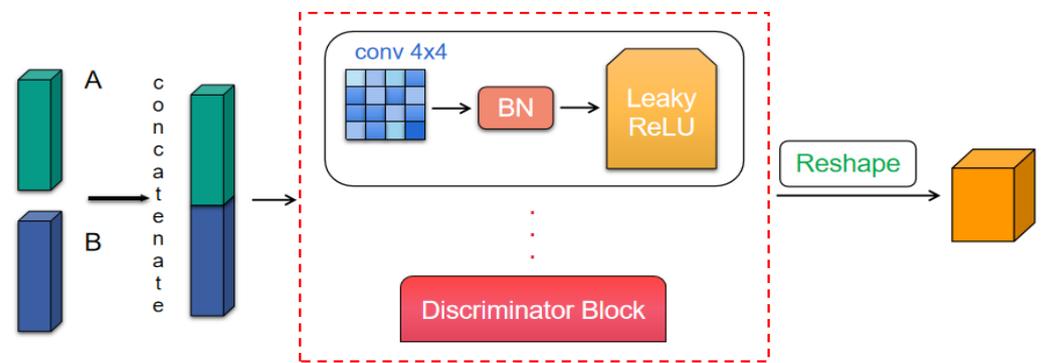
**Figure 3.** (a) The structure of the basic block in the encoder. (b) Structure of DRB, which consists of three convolutional layers, equivalent to a  $3 \times 3$  convolutional layer during inference; (c) ERB, which consists of two DRBs.



**Figure 4.** The overall structure of the decoder is composed of the connection of conv and deconv structures. The processing result is added with the output of the mirrored encoder.

3.1.2. Discriminator

For the discriminator, we use a Markov discriminator [23], which is similar to that in the literature [26]. This discriminator can distinguish the true and false of the  $N \times N$  block in the image, and the  $N$  here can be much smaller than the real size of the image so that the network has fewer parameters, runs faster, and can be applied to images of any size. Moreover, the three consecutive discriminator blocks maintain a certain high resolution and high detail for ultra-high resolution and image clarity in style transfer. The details of the discriminator are shown in Figure 5, and the input is set to  $256 \times 256 \times 6$ . There are a total of four modules, each of which performs convolution operations with a kernel size of  $4 \times 4$  and a stride of 2 on the image, followed by the BN and LeakyReLU layers. The final output is  $16 \times 16 \times 1$ , representing the average effective response of the discriminator.



**Figure 5.** Discriminator, each discriminator block consists of conv, BN, LeakyReLU in series, where A is the original image and B is the result produced by the generator.

3.2. Objective Function Formulation

The conditional generative adversarial network introduces additional auxiliary information and can learn the mapping  $G : \{X, Z\} \rightarrow Y$ . The objective function expresses as

$$L_{cGAN} = E_{X,Y}[\log D(Y)] + E_{X,Y}[\log(1 - D(X, G(X, Z)))] \tag{1}$$

where the generator  $G$  tries to minimize  $L_{cGAN}$ , and the discriminator  $D$  tries to maximize it [34]. Therefore, a game state is formed, which is

$$GS = \operatorname{argmin}_G \operatorname{max}_D L_{cGAN}(G, D) \tag{2}$$

To further improve the quality of the generated image, we introduce the pixel-level and image-level loss functions into the objective function formulation.

**Pixel-level:** Existing research shows that the  $L_1$  loss can efficiently capture high-frequency information in images. On the one hand, measure the pixel value between the real and generated image, ignoring the difference between the adjacent pixel values; on the other hand, the introduction of blur can be avoided, making  $G$  focus more on the color distortion of the image, and produce more realistic results. The water medium absorbs different wavelengths of light disparity, and the suspended particles have the effect of absorbing and scattering light, so most of the underwater images appear blue-green. Therefore, to make the enhanced image have a similar color style to the natural image, we add the  $L_1$  loss to the objective function:

$$L_1 = \frac{1}{N} \sum_{i=1}^N |Y_i - G(X_i)| \quad (3)$$

where  $Y$  represents the real image, and  $G(X)$  represents the image generated by  $G$ .

**Image level:** Inspired by [35,36], according to the activation function generated by the ReLU layer of the pre-trained VGG-19 network, we define the content loss at the image level. This loss ignores the differences between per-pixels, encourages them to have similar feature representations in terms of content and perceived quality, and ensures that the image generated by the  $G$  and the real image have similar content. The formula for this loss function is as follows:

$$L_{con}(G) = E_{X,Y,Z} [\|\Phi(Y) - \Phi(G(X,Z))\|_2] \quad (4)$$

The model we proposed uses paired image training, and an objective function is constructed for this purpose to guide  $G$  to generate an enhanced image that is highly similar to the real image in color and overall content.  $D$  filters out and discards images with different local textures and styles. Specifically, our model uses the following objective function for paired image training:

$$L_{total} = \operatorname{argmin}_G \max_D L_{cGAN} + \lambda_1 L_1 + \lambda_2 L_{con} \quad (5)$$

where  $\lambda_1 = 0.7$ ,  $\lambda_2 = 0.3$  respectively represent pixel-level and content-level loss function factors obtained based on experience.

## 4. Experiment

### 4.1. Dataset

In this section, we chose a relatively complete set of real and artificial synthetic underwater images to test the enhancement effect of the proposed model.

#### 4.1.1. EUVP Dataset

In this dataset, part of the images are collected by seven different camera equipment; the other part comes from images captured in YouTube videos. It was proposed by Ref. [26], which contains three different categories: underwater\_dark (5550 pairs), underwater\_imagenet (3700 pairs), underwater\_scenes (2185 pairs). The ground truth used for reference is generated by the trained CycleGAN [24]. Here, we use 3 different classifications of image pairs for training, another 515 images, and their corresponding reference images for testing.

#### 4.1.2. ImageNet Dataset

This dataset uses the images with good brightness and visibility collected from Imagenet as ground truth. Similar to the EUVP dataset, using the trained CycleGAN [24] to degrade the ground truth images to generate distorted images, a total of 6128 pairs of images makes this dataset. In this experiment, we randomly selected 5500 pairs of images for training and the remaining 628 pairs for testing.

#### 4.1.3. Mixed Dataset

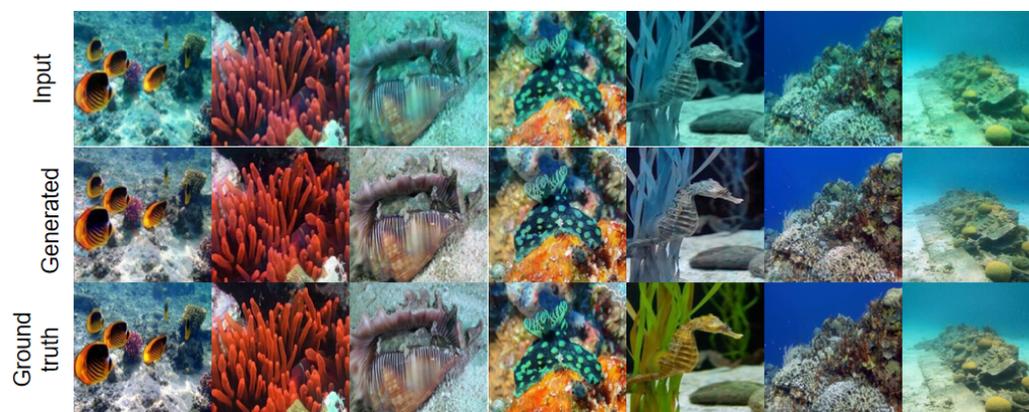
Due to the lack of real underwater images, Silberman et al. in [37] proposed a synthetic underwater image dataset NYU-v2. The distorted images are generated based on the underwater image model and image synthesis algorithm. This dataset has generated the ocean (I, IA, IB, II, III, where Type I is the clearest, Type III is the most turbid) and coastal (1, 3, 5, 7, 9, where Type 1 is the clearest and Type 9 is the most turbid) types of images. In addition, UIEBD is a dataset composed of 890 real underwater images with their corresponding reference images [38]. In this experiment, we randomly selected 1250 pairs of images from the NYU-v2 dataset, and 800 pairs from the UIEBD dataset, then mixed them for training. The remaining 90 pairs of images from UIEBD were used as the test set Test-R90. There are 60 real underwater images without reference in UIEBD, which are severely distorted and difficult to restore. We also named them Test-C60 for this experiment.

#### 4.2. Implement Details

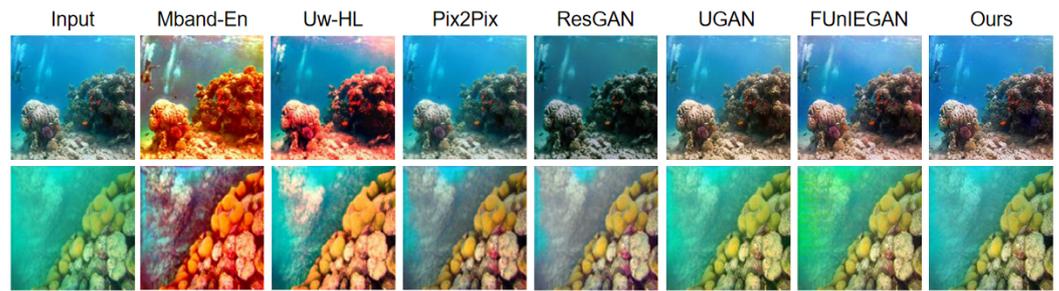
We used Pytorch 1.8.0 to implement the FE-GAN model. Due to the limitation of memory, all pictures were resized to  $256 \times 256$  before training, which can also preserve global information better. For each dataset, 200 epochs with a batch\_size of 8 were trained on an NVIDIA GeForce RTX 2060 graphics card. In the training process, we used Adam as the optimizer. Now, we will analyze the experimental results in both qualitative and quantitative aspects.

#### 4.3. Qualitative Evaluation

We apply the FE-GAN model to real and artificially synthesized underwater image datasets, process paired and unpaired distorted images, and compare them with the corresponding ground truth images. Figure 6 shows some examples of the FE-GAN model in the EUVP dataset. It can be seen that our model effectively improves the contrast of the distorted image and makes the overall color more vivid. Moreover, it also improves the sharpness, so the image appears clearer. Here, we selected several state-of-the-art models. Among them, pix2pix [23], ResGAN [39], UGAN [40], FUnIE-GAN [26] are models based on the learning method, Mband-En [41] and Uw-HL [42] are physical-based models. As can be seen from Figure 7, the images generated by the physical-based models (Mband-En and Uw-HL) have over-saturated brightness and over-exposure, simultaneously, and the whole image presents red and yellow hues. The background of the image generated by ResGAN is too dark, and the contrast is not distinct; UGAN and FUnIE-GAN are not thorough enough to deal with the blue-green hues of the distorted image, so the background is still light green.



**Figure 6.** Several generated results of the EUVP dataset.

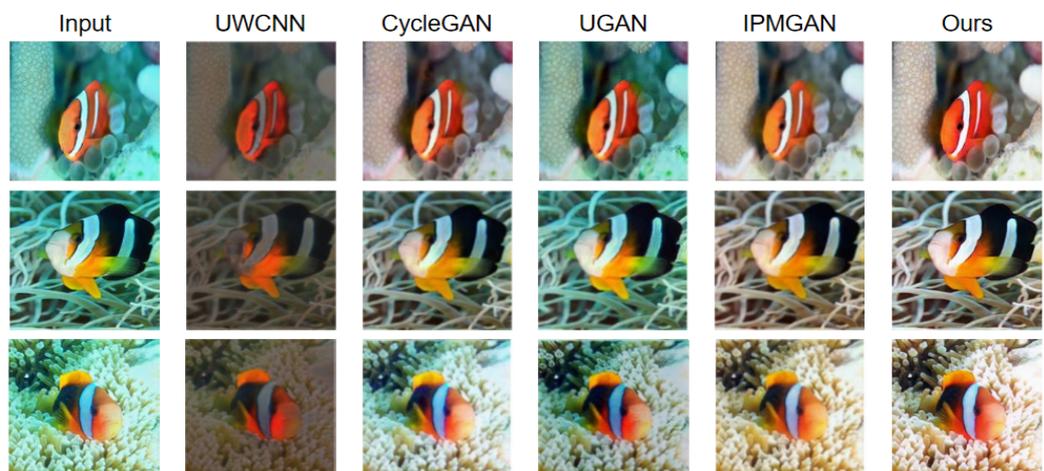


**Figure 7.** Comparison results with Mband-En, Uw-HL, Pix2Pix, ResGAN, UGAN, FUnIEGAN on the EUVP dataset.

For the ImageNet dataset, we randomly selected 628 pairs of real underwater images for testing. Figure 8 shows a part of the test results. Through the processing of our FE-GAN model, the generated image effectively repairs the overall blue-green hue, and the entire brightness of the image is improved. The enhanced image effectively restores the red light, which is absorbed fastest, better distinguishes the background and foreground, and the overall image is clearer. We also selected several state-of-the-art learning-based models, UWCNN [43], CycleGAN [24], UGAN [40], IPMGAN [17], and compared them with our experimental results. As can be seen from Figure 9, the image generated by UWCNN is dark overall, but some areas are too bright to highlight the pivotal areas in the image. CycleGAN and UGAN have poor repair effects on green hues, and obvious traces of the original image can still be seen in the background. The sharpening effect of the image generated by IPMGAN is not good, and the holistic image is pale yellow and not clear enough.



**Figure 8.** Generated results of the ImageNet test set.



**Figure 9.** Comparison results with UWCNN, CycleGAN, UGAN, IPMGAN on the ImageNet dataset.

To further verify the generalization ability of FE-GAN, we selected 990 images from the artificially synthesized dataset for testing and compared them with the corresponding ground truth images. Figure 10 shows the effect of partially generated images. The overall hue of the image generated by FE-GAN is close to the ground truth. At the same time, the repair effect of some overexposed areas in the distorted image is also better. We can see more texture information from the restored image. Considering the complex layout of the images in the dataset, the holistic effect of the enhanced image can meet the practical applications. However, it has to be mentioned that whether it is a real or artificial dataset, our model does not perform well for images with insufficient light. It cannot restore the brightness and details of the image very well, and the enhanced image is still relatively dark.



**Figure 10.** Generated results of the synthetic dataset.

#### 4.4. Quantitative Evaluation

To quantitatively analyze the enhancement effect of the FE-GAN model on the paired underwater image, we choose PSNR (peak signal-to-noise ratio) and SSIM (structural similarity) as reference indicators. PSNR is an index used in the image field to measure the quality of reconstructed images, which is defined by taking the logarithm of MSE (mean squared error). Given a generated image  $I$  of size  $m \times n$  and its corresponding ground truth image  $K$ , their MSE is expressed as

$$MSE = \frac{1}{mn} \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} [I(i, j) - K(i, j)]^2 \quad (6)$$

Since we resized the image before the experiment, the values of  $m$  and  $n$  are both 256. The definition of PSNR is as follows:

$$PSNR = 10 \log_{10} \frac{MAX_I^2}{MSE} \quad (7)$$

where  $MAX_I$  represents the maximum possible pixel value of the image, where each pixel is represented by 8 in binary, so  $MAX_I = 2^8 - 1 = 255$ .

SSIM is a metric used to measure the similarity of images, and it can also be used to judge the quality of images after compression. It is mainly composed of three parts: luminance, contrast, and structure contrast. These are expressed as follows:

$$l(x, y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \quad (8)$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \quad (9)$$

$$s(x, y) = \frac{\sigma_{xy} + c_3}{\sigma_x \sigma_y + c_3} \quad (10)$$

where, the  $\mu_x, \mu_y$  represent the mean value of  $x, y$ ;  $\sigma_x^2, \sigma_y^2$  represent the variance of  $x, y$ ; and  $\sigma_{xy}$  represent the covariance of  $x, y$ .  $c_1 = (k_1 L)^2, c_2 = (k_2 L)^2$  are two constants, and  $L$  represents the range of pixel values [0, 255]. The definition of SSIM is as follows:

$$SSIM = \frac{(2\mu_x \mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (11)$$

UIQM is a non-referenced underwater image quality evaluation metric based on the human visual system excitation, mainly for the degradation mechanism and imaging characteristics of underwater images. Using UICM (color measurement index), UISM (sharpness measurement index), UIConM (contrast measurement index) as the evaluation basis. UIQM is expressed as a linear combination of these three indexes. The larger the value, the better the color balance, clarity, and color of the image. UIQM expresses as follows:

$$UIQM = c_1 UICM + c_2 UISM + c_3 UIConM \quad (12)$$

where  $c_1 = 0.0282, c_2 = 0.2953, c_3 = 3.5753$  [44].

Table 1 shows the quantitative comparison results of our model and state-of-the-art models on the EUVP dataset. We chose 1k paired images and tested PSNR, SSIM, and UIQM, respectively. As can be seen from the table, our model achieved better results in the two measurement indicators of PSNR and UIQM. SSIM achieves second place after FUnIE-GAN among the learning-based methods. In general, after comparing with ground truth images, the generated images of our model are better.

**Table 1.** Quantitative comparison results with the state-of-the-art methods in the EUVP dataset.

Model	PSNR	SSIM	UIQM
Mband-EN [41]	12.11	0.4565	2.28
Uw-HL [42]	18.85	0.7722	2.62
Pix2Pix [23]	20.27	0.7081	2.65
ResGAN [39]	14.75	0.4685	2.62
UGAN [40]	19.59	0.6685	2.72
FUnIE-GAN [26]	21.92	0.8876	2.78
MFFN [45]	24.73	0.8456	2.32
FE-GAN	26.83	0.8779	2.87

In the ImageNet dataset, we randomly selected 5500 pairs of images for training and the remaining 628 pairs for testing. Table 2 shows the test results of our model with UWCNN [43], CycleGAN [24], UGAN [40], FUnIE-GAN [26], IPMGAN [17] on the ImageNet dataset test set. Similarly, the PSNR value of the image generated by our model reached the highest results.

**Table 2.** Quantitative comparison results with the state-of-the-art methods in the ImageNet dataset.

Model	PSNR	SSIM
UWCNN [43]	15.4190	0.6127
CyCleGAN [24]	22.3160	0.7464
UGAN [40]	18.6562	0.5702
FUnIE-GAN [26]	24.3204	0.8193
IPMGAN [17]	23.5439	0.8142
FE-GAN	24.5396	0.8106

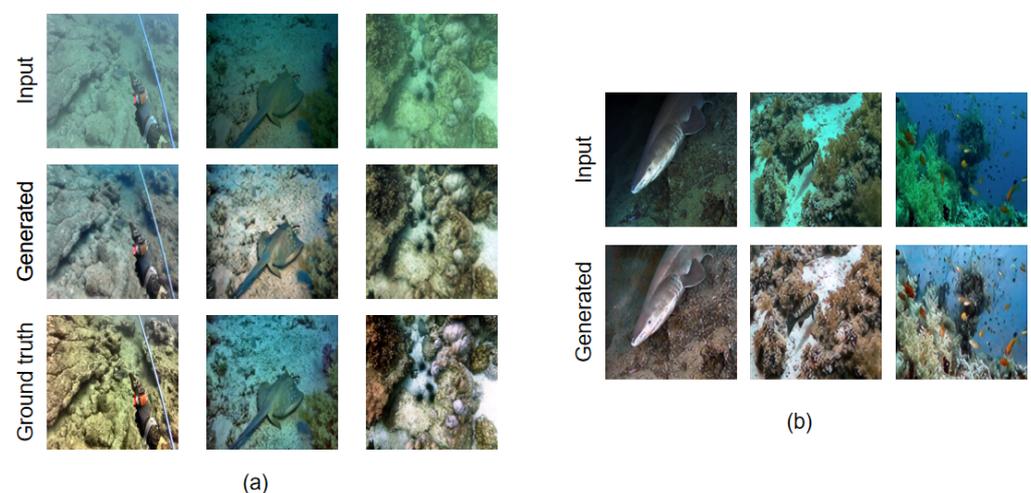
For the Mixed dataset, we selected Test-R90 (90 paired images) and Test-C60 (60 unpaired images) as the test sets of paired and unpaired images respectively and compared them with the same methods in qualitative evaluation. Both of these test sets are from the UIEBD dataset, which is more challenging. These images were taken in a poor light environment, and the overall number of this dataset is small, which brings a certain degree of difficulty to training. Here, we selected UCycleGAN [46], WaterNet [38], UWCNN [43], Unet-U [47], Ucolor [48], and FUnIE-GAN [26] models for comparison. Table 3 shows the PSNR test results of Test-R90. Our model achieved the best value. Figure 11a shows some of the test results. The image generated by our model can improve the clarity and brightness of the image while retaining the overall layout, effectively improving the yellow-green hue problem. Table 4 shows the test results of Test-C60. Since there is no reference image in this dataset, we only chose UIQM as a comparison indicator. Similarly, our model achieved better results. Figure 11b shows part of the generated image of Test-C60. The overall brightness of the image is significantly improved, which will be helpful for the recognition of underwater images and resource exploration.

**Table 3.** PSNR test results of Test-R90. FE-GAN achieved the highest value.

Model	PSNR
UCycleGAN [46]	16.61
Water-Net [38]	19.81
UWCNN [43]	16.69
Unet-U [47]	18.14
Ucolor [48]	20.63
FUnIE-GAN [26]	20.27
FE-GAN	20.68

**Table 4.** UIQM test results of Test-C60. FE-GAN achieved the highest value.

Model	UIQM
UCycleGAN [46]	0.91
Water-Net [38]	0.97
UWCNN [43]	0.84
Unet-U [47]	0.94
Ucolor [48]	0.88
FUnIE-GAN [26]	0.98
Framework [27]	0.99
MDP Framework [49]	0.97
FE-GAN	1.01



**Figure 11.** Generated results of Test-R90 (a), Test-C60 (b).

The application of underwater image enhancement technology to underwater detection equipment is an important research direction. Several aspects should be taken into consideration, such as FLOPs, number of parameters, and inference time during deploying on resource-limited devices. We chose fps as a metric to measure inference time, which expresses as

$$fps = \frac{framNum}{elapsedTime} \quad (13)$$

Table 5 shows the comparison results among FE-GAN, UGAN, and the more lightweight network FUnIE-GAN. The amount of parameters of the FE-GAN model is much fewer than that of UGAN. However, since the FE-GAN encoder is deeper than the FUnIE-GAN network, the amount of parameters is slightly larger than FUnIE-GAN. Furthermore, due to the structural re-parameterization module's ability to speed up network inference and reduce memory consumption, our model has the lowest FLOPs, which means that the real-time performance is better. The fps indicators of FE-GAN and FUnIE-GAN both exceed 200. In summary, the FE-GAN model can meet the current real-time requirements of underwater detection.

**Table 5.** Real-time comparison results with UGAN and FUnIE-GAN.

Model	FLOPs (G)	Paramaters (M)	fps (Hz)
UGAN [40]	18.143	54.4	93.1
FUnIE-GAN [26]	7.834	7.02	217.9
FE-GAN	2.84	11.476	204.1

For AUVs and ROVs, during underwater exploration activities, the purpose of improving the image quality is to improve the accuracy of tasks such as object detection and classification. We chose the pre-trained YOLOv5 as the object detection model and tested the images before and after enhancement on the EUVP dataset. Part of the test results is shown in Figure 12.



**Figure 12.** Object detection results on EUVP dataset.

The first line is the unprocessed original distorted images, and the second line is the FE-GAN processed images. Compared with the original distorted image, the processed image has a more natural tone and increased brightness, so the target in the image is clearer and easier to identify. The results in the second, fifth, and last columns show that the fuzzy target can be detected in the processed image. Other examples show that the recognition error of the processed image is alleviated.

In addition, we downloaded the Aquarium Combined dataset, then trained and tested this dataset on the same hardware environment as the FE-GAN enhancement experiment. The object detection test was performed before and after the FE-GAN processing. Here, we chose YOLOv5 as the object detector. Figure 13 shows part of the test results. The overall blue tone of the original image is obvious, and the processed image effectively repairs this

problem, making the image closer to the ground truth situation. Figures 14–16 are the precision curve and confusion matrix trained on YOLOv5 for this dataset.

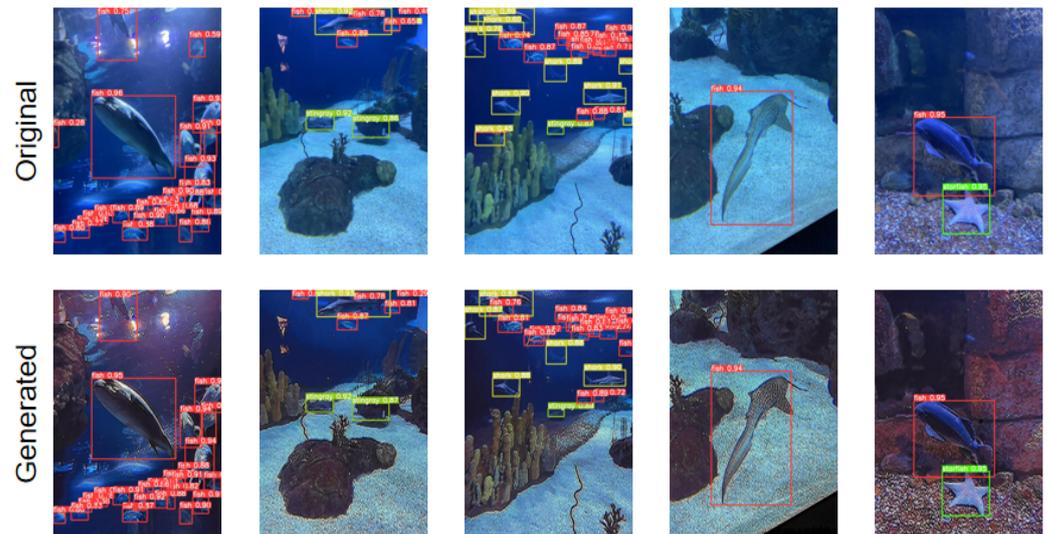


Figure 13. Object detection results on Aquarium Combined dataset.

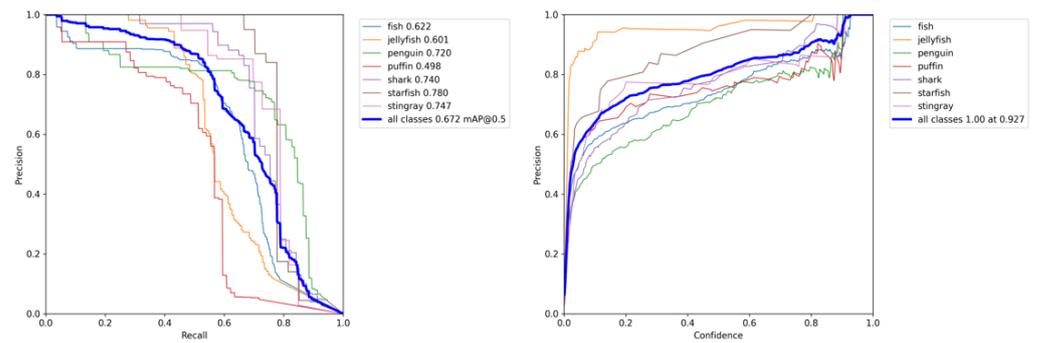


Figure 14. PR-curve and R-curve.

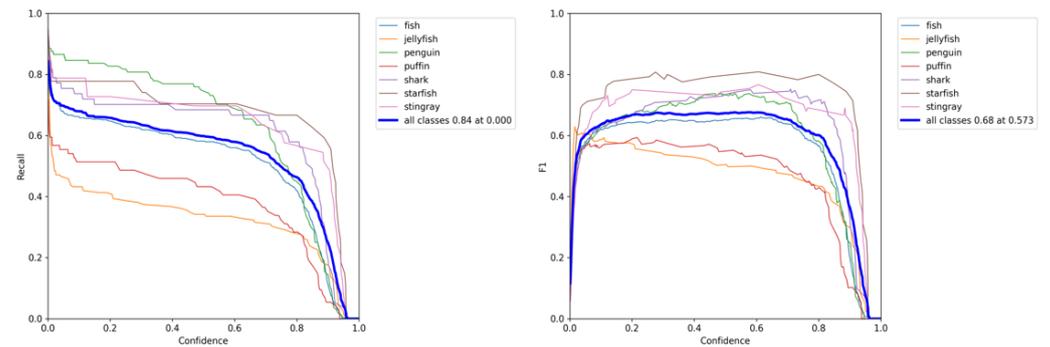


Figure 15. F1-curve and P-curve.



Figure 16. Confusion matrix on Aquarium Combined dataset; the darker the color, the higher the detection accuracy.

In the Aquarium Combined dataset, there are seven types of targets to be detected: fish, jellyfish, penguin, puffin, shark, starfish, and stingray. Here we used mAP (mean average precision) as a reference metric. Table 6 shows FUnIE-GAN, UGAN, Pix2Pix, and FE-GAN in the above different types of mAP, and the average mAP of all classes. FE-GAN obtained the highest mAP value in the five categories of detection, and the average mAP of all classes also reached 0.672, which is the highest value among the above models.

Table 6. The mAP comparison results of object detection with the state-of-the-art methods in the Aquarium Combined dataset.

Model	Fish	Jellyfish	Penguin	Puffin	Shark	Starfish	Stingray	All Classes
FUnIE-GAN [26]	0.587	0.631	0.664	0.406	0.657	0.799	0.674	0.631
UGAN [40]	0.599	0.588	0.683	0.411	0.709	0.738	0.707	0.634
Pix2Pix [23]	0.580	0.587	0.662	0.416	0.697	0.775	0.652	0.624
FE-GAN	0.622	0.601	0.720	0.498	0.740	0.780	0.747	0.672

In the case of insufficient natural light, the image obtained with the artificial light source itself is extremely distorted. Although the brightness and details of the image enhanced by FE-GAN were restored partially, there is still a large gap from the image style under natural light, which is also the focus of future research.

### 5. Ablation Study

We conducted feature fusion experiments between the encoder and decoder utilizing concatenate and aggregation, respectively. Isola et al. [23] used concatenate in the connection of encoder and decoder in U-Net, achieving encouraging results in the task of generating images of conditional GAN. In [26,50], concatenate is further proven to have better results in underwater image enhancement tasks. The aggregation method increases the amount of information describing the characteristics of the image, which means only the amount of information in one dimension is increasing [51]. In this case, Equations (14) and (15) ( $X_i, Y_i$  represent the two input channels respectively) show that the aggregation method can greatly reduce the amount of computation, which is very beneficial for FE-GAN. We used the concatenate and aggregation methods to carry out information fusion experiments between the encoder and decoder. Table 7 shows the impact of the two

methods on the inference speed with 44.31% promotion in GFLOPs and 3.19% reduction in number of parameters, and fps also increased 19.3%. To further verify the effectiveness of this choice, we tested the comparison network on the test set of the ImageNet dataset (628 paired images) and Test-R90 (90 paired images) of the Mixed dataset. The experimental results are shown in Tables 8 and 9 below.

$$Z_{concat} = \sum_{i=1}^c (X_i * K_i) + \sum_{i=1}^c (Y_i * K_i) \quad (14)$$

$$Z_{add} = \sum_{i=1}^c [(X_i + Y_i) * K_i] \quad (15)$$

Here, we also chose PSNR and SSIM as the evaluation indicators that regard aggregation and concatenate as the connection mode between the encoder and the decoder. Under the same experimental conditions, the test results using the aggregation operation method perform better in both PSNR and SSIM values. It should be pointed out that because the training set and test set of the Mixed dataset are relatively small, the experimental gap here is not very large.

**Table 7.** Experimental results of information fusion on model inference speed using concatenate and aggregation.

Fusion Method	GFLOPs	Parameters	fps
add	2.84	11.476	204.1
concatenate	5.1	11.854	171.0

**Table 8.** Test results of the ImageNet dataset. Using aggregation and concatenate connection between encoder and decoder.

ImageNet	Add	Concatenate
PSNR	24.5396	22.8863
SSIM	0.8106	0.7178

**Table 9.** Test results of the Mixed dataset. Using aggregation and concatenate connection between encoder and decoder.

Test-R90	Add	Concatenate
PSNR	20.6837	20.2470
SSIM	0.8074	0.6804

In recent years, many learning-based methods used  $L_1$ ,  $L_2$ , and *smooth*  $L_1$  loss to design the objective function. All of these loss functions have been proven to achieve significant results in different tasks. We used these three different loss functions to design the objective function of the entire model from the pixel level. Similarly, we conducted experiments on two different datasets, and the results are shown in Tables 10 and 11. *Smooth*  $L_1$  loss solves the problem that the derivative is not unique at 0 and optimizes the convergence ability of the model, and it avoids the shortcomings of  $L_2$  loss increasing with the square error. For low-level tasks, such as image enhancement, the stable gradient of  $L_1$  loss is more conducive to the convergence of the model and the optimization of training. *Smooth*  $L_1$  loss is more suitable for tasks such as object detection.

**Table 10.** Results of different loss functions under the ImageNet dataset.

ImageNet	$L_1$ loss	$L_2$ loss	smooth $L_1$ loss
PSNR	24.5396	24.1444	24.4511
SSIM	0.8106	0.7249	0.7095

**Table 11.** Results of different loss functions under the Mixed dataset.

Test-R90	$L_1$ loss	$L_2$ loss	smooth $L_1$ loss
PSNR	20.6837	19.6374	20.3014
SSIM	0.8074	0.6411	0.7859

Structural reparameterization is used in our encoder to speed up inference. As shown in Figure 3, we equate DRB to a  $3 \times 3$  convolutional layer during inference, for which we test the performance of the model on different datasets, respectively. Table 12 lists the results of the comparative experiments. The method using structural reparameterization performs better on two datasets, and its inference speed does increase, with 11.52% promotion in GFLOPs and 0.78% reduction in number of parameters.

**Table 12.** FE-GAN (without-re) represents the model used in the inference stage. The metrics under the dataset represent PSNR and SSIM, respectively.

-	FE-GAN (Training)	FE-GAN (Inference)
EUVP	25.78/0.7568	26.83/0.8779
ImageNet	24.67/0.7538	24.54/0.8106
Test-R90	20.31/0.7196	20.68/0.8074
GFLOPs	3.21	2.84
parameters(M)	11.566	11.476

## 6. Conclusions

In this paper, we proposed an underwater image enhancement model based on a conditional generative adversarial network. Structural reparameterization methods improved the ability of the model to extract features while also speeding up inference. The color, brightness, and contrast of the generated image were distinctly improved. Simultaneously, our model conducted qualitative and quantitative analysis experiments on real underwater images and artificial synthetic image datasets respectively, which effectively demonstrates the generalization ability of the model. Compared with the state-of-the-art methods, our model achieved better results. Nevertheless, our model does not perform well in enhancing darker images, especially in recovering details and textures, which means that it is still challenging in deeper waters, where artificial light sources are needed. Next, we will try to optimize more network modules with structural reparameterization to improve the enhancement effect of the model on images with insufficient brightness, and focus on the practical application in underwater object detection and scene analysis.

**Author Contributions:** Conceptualization, J.H. and J.Z.; methodology, J.H. and J.Z.; software, J.H.; validation, J.H.; resources, J.Z. and Z.D.; data curation, L.W. and Y.W.; writing—original draft preparation, J.H. and J.Z.; writing—review and editing, J.H. and J.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Key R&D plan of Shandong Province (2020JMRH0101), National Deep Sea Center.

**Data Availability Statement:** The publicly available dataset used in this research can be obtained through the following link: <https://irvlab.cs.umn.edu/resources/euvp-dataset> (accessed on 1 February 2023).

**Acknowledgments:** The authors would like to thank the Key R&D plan of Shandong Province (2020JMRH0101), National Deep Sea Center.

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

## References

1. Zeng, L.; Sun, B.; Zhu, D. Underwater target detection based on Faster R-CNN and adversarial occlusion network. *Eng. Appl. Artif. Intell.* **2021**, *100*, 104190. [[CrossRef](#)]
2. Zhang, L.; Li, C.; Sun, H. Object detection/tracking toward underwater photographs by remotely operated vehicles (ROVs). *Future Gener. Comput. Syst.* **2022**, *126*, 163–168. [[CrossRef](#)]
3. Zhang, H.; Zhang, S.; Wang, Y.; Liu, Y.; Yang, Y.; Zhou, T.; Bian, H. Subsea pipeline leak inspection by autonomous underwater vehicle. *Appl. Ocean Res.* **2021**, *107*, 102321. [[CrossRef](#)]
4. Fatan, M.; Daliri, M.R.; Shahri, A.M. Underwater cable detection in the images using edge classification based on texture information. *Measurement* **2016**, *91*, 309–317. [[CrossRef](#)]
5. Li, Y.; Lu, H.; Zhang, L.; Li, J.; Serikawa, S. Real-time visualization system for deep-sea surveying. *Math. Probl. Eng.* **2014**, *2014*. [[CrossRef](#)]
6. Jaffe, J.S. Underwater optical imaging: The past, the present, and the prospects. *IEEE J. Ocean. Eng.* **2014**, *40*, 683–700. [[CrossRef](#)]
7. Li, H.; Zhuang, P. DewaterNet: A fusion adversarial real underwater image enhancement network. *Signal Process. Image Commun.* **2021**, *95*, 116248. [[CrossRef](#)]
8. Zhang, H.; Sun, L.; Wu, L.; Gu, K. DuGAN: An effective framework for underwater image enhancement. *IET Image Process.* **2021**, *15*, 2010–2019. [[CrossRef](#)]
9. Akkaynak, D.; Treibitz, T. A revised underwater image formation model. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6723–6732.
10. Han, M.; Lyu, Z.; Qiu, T.; Xu, M. A review on intelligence dehazing and color restoration for underwater images. *IEEE Trans. Syst. Man Cybern. Syst.* **2018**, *50*, 1820–1832. [[CrossRef](#)]
11. Nayar, S.K.; Narasimhan, S.G. Vision in bad weather. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999; Volume 2; pp. 820–827.
12. Peng, Y.T.; Zhao, X.; Cosman, P.C. Single underwater image enhancement using depth estimation based on blurriness. In Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada, 27–30 September 2015; pp. 4952–4956.
13. Lu, H.; Li, Y.; Zhang, L.; Serikawa, S. Contrast enhancement for images in turbid water. *JOSA A* **2015**, *32*, 886–893. [[CrossRef](#)]
14. Berman, D.; Treibitz, T.; Avidan, S. Diving into haze-lines: Color restoration of underwater images. In Proceedings of the Proc. British Machine Vision Conference (BMVC), London, UK, 4–7 September 2017; Volume 1.
15. Spier, O.; Treibitz, T.; Gilboa, G. In situ target-less calibration of turbid media. In Proceedings of the 2017 IEEE International Conference on Computational Photography (ICCP), Stanford, CA, USA, 12–14 May 2017; pp. 1–9.
16. Akkaynak, D.; Treibitz, T. Sea-thru: A method for removing water from underwater images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 1682–1691.
17. Liu, X.; Gao, Z.; Chen, B.M. IPMGAN: Integrating physical model and generative adversarial network for underwater image enhancement. *Neurocomputing* **2021**, *453*, 538–551. [[CrossRef](#)]
18. Liu, P.; Wang, G.; Qi, H.; Zhang, C.; Zheng, H.; Yu, Z. Underwater image enhancement with a deep residual framework. *IEEE Access* **2019**, *7*, 94614–94629. [[CrossRef](#)]
19. Qi, Q.; Zhang, Y.; Tian, F.; Wu, Q.J.; Li, K.; Luan, X.; Song, D. Underwater image co-enhancement with correlation feature matching and joint learning. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *32*, 1133–1147. [[CrossRef](#)]
20. Mao, X.; Li, Q.; Xie, H.; Lau, R.Y.; Wang, Z.; Paul Smolley, S. Least squares generative adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2794–2802.
21. Zhao, J.; Mathieu, M.; LeCun, Y. Energy-based generative adversarial network. *arXiv* **2016**, arXiv:1609.03126.
22. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein generative adversarial networks. In Proceedings of the International Conference on Machine Learning PMLR, Sydney, Australia, 7–9 August 2017; pp. 214–223.
23. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.
24. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
25. Yi, Z.; Zhang, H.; Tan, P.; Gong, M. Dualgan: Unsupervised dual learning for image-to-image translation. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2849–2857.
26. Islam, M.J.; Xia, Y.; Sattar, J. Fast underwater image enhancement for improved visual perception. *IEEE Robot. Autom. Lett.* **2020**, *5*, 3227–3234. [[CrossRef](#)]
27. Zhou, Y.; Yan, K.; Li, X. Underwater image enhancement via physical-feedback adversarial transfer learning. *IEEE J. Ocean. Eng.* **2021**, *47*, 76–87. [[CrossRef](#)]

28. Hambarde, P.; Murala, S.; Dhall, A. UW-GAN: Single-image depth estimation and image enhancement for underwater images. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–12. [[CrossRef](#)]
29. Aggarwal, A.; Mittal, M.; Battineni, G. Generative adversarial network: An overview of theory and applications. *Int. J. Inf. Manag. Data Insights* **2021**, *1*, 100004. [[CrossRef](#)]
30. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *arXiv* **2014**, arXiv:1406.2661.
31. Du, Z.; Liu, D.; Liu, J.; Tang, J.; Wu, G.; Fu, L. Fast and Memory-Efficient Network Towards Efficient Image Super-Resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 21 June 2022; pp. 853–862.
32. Ding, X.; Zhang, X.; Ma, N.; Han, J.; Ding, G.; Sun, J. Repvgg: Making vgg-style convnets great again. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13733–13742.
33. Galteri, L.; Seidenari, L.; Bertini, M.; Del Bimbo, A. Deep generative adversarial compression artifact removal. In Proceedings of the International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 4826–4835.
34. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv* **2014**, arXiv:1411.1784.
35. Ignatov, A.; Kobyshev, N.; Timofte, R.; Vanhoey, K.; Van Gool, L. Dslr-quality photos on mobile devices with deep convolutional networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 3277–3285.
36. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual losses for real-time style transfer and super-resolution. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 694–711.
37. Silberman, N.; Hoiem, D.; Kohli, P.; Fergus, R. Indoor segmentation and support inference from rgbd images. *ECCV (5)* **2012**, *7576*, 746–760.
38. Li, C.; Guo, C.; Ren, W.; Cong, R.; Hou, J.; Kwong, S.; Tao, D. An underwater image enhancement benchmark dataset and beyond. *IEEE Trans. Image Process.* **2019**, *29*, 4376–4389. [[CrossRef](#)] [[PubMed](#)]
39. Li, J.; Liang, X.; Wei, Y.; Xu, T.; Feng, J.; Yan, S. Perceptual generative adversarial networks for small object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1222–1230.
40. Fabbri, C.; Islam, M.J.; Sattar, J. Enhancing underwater imagery using generative adversarial networks. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, Australia, 21–25 May 2018; pp. 7159–7165.
41. Cho, Y.; Jeong, J.; Kim, A. Model-assisted multiband fusion for single image enhancement and applications to robot vision. *IEEE Robot. Autom. Lett.* **2018**, *3*, 2822–2829.
42. Berman, D.; Levy, D.; Avidan, S.; Treibitz, T. Underwater single image color restoration using haze-lines and a new quantitative dataset. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**. [[CrossRef](#)] [[PubMed](#)]
43. Li, C.; Anwar, S.; Porikli, F. Underwater scene prior inspired deep underwater image and video enhancement. *Pattern Recognit.* **2020**, *98*, 107038. [[CrossRef](#)]
44. Panetta, K.; Gao, C.; Agaian, S. Human-visual-system-inspired underwater image quality measures. *IEEE J. Ocean. Eng.* **2015**, *41*, 541–551. [[CrossRef](#)]
45. Chen, R.; Cai, Z.; Cao, W. MFFN: An underwater sensing scene image enhancement method based on multiscale feature fusion network. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–12. [[CrossRef](#)]
46. Li, C.; Guo, J.; Guo, C. Emerging from water: Underwater image color correction based on weakly supervised color transfer. *IEEE Signal Process. Lett.* **2018**, *25*, 323–327. [[CrossRef](#)]
47. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
48. Li, C.; Anwar, S.; Hou, J.; Cong, R.; Guo, C.; Ren, W. Underwater Image Enhancement via Medium Transmission-Guided Multi-Color Space Embedding. *IEEE Trans. Image Process.* **2021**, *30*, 4985–5000. [[CrossRef](#)]
49. Sun, S.; Wang, H.; Zhang, H.; Li, M.; Xiang, M.; Luo, C.; Ren, P. Underwater image enhancement with reinforcement learning. *IEEE J. Ocean. Eng.* **2022**. [[CrossRef](#)]
50. Wang, Y.; Guo, J.; Gao, H.; Yue, H. UIEC<sup>2</sup>-Net: CNN-based underwater image enhancement using two color space. *Signal Process. Image Commun.* **2021**, *96*, 116250. [[CrossRef](#)]
51. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

**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.