Comparison of GAN Deep Learning Methods for Underwater Optical Image Enhancement

Hong-Gi Kim1,2, Jung-Min Seo3 and Soo Mee Kim4,5

1Graduate Student, Department of Convergence Study on the Ocean Science and Technology, Korea Maritime and Ocean University, Busan, Korea
2OST Student-in-training, Maritime ICT R&D Center, Korea Institute of Ocean Science and Technology, Busan, Korea
3Research Specialist, Maritime ICT R&D Center, Korea Institute of Ocean Science and Technology, Busan, Korea
4Senior Researcher, Maritime ICT R&D Center, Korea Institute of Ocean Science and Technology, Busan, Korea
5Professor, Dept. of Convergence Study on the Ocean Science and Technology, Korea Maritime and Ocean University, Busan, Korea

KEY WORDS: Generative adversarial networks, Image fusion, Image enhancement, Underwater optical image, Underwater image deep learning techniques

ABSTRACT: Underwater optical images face various limitations that degrade the image quality compared with optical images taken in our atmosphere. Attenuation according to the wavelength of light and reflection by very small floating objects cause low contrast, blurry clarity, and color degradation in underwater images. We constructed an image data of the Korean sea and enhanced it by learning the characteristics of underwater images using the deep learning techniques of CycleGAN (cycle-consistent adversarial network), UGAN (underwater GAN), FUnIE-GAN (fast underwater image enhancement GAN). In addition, the underwater optical image was enhanced using the image processing technique of Image Fusion. For a quantitative performance comparison, UIQM (underwater image quality measure), which evaluates the performance of the enhancement in terms of colorfulness, sharpness, and contrast, and UCIQE (underwater color image quality evaluation), which evaluates the performance in terms of chroma, luminance, and saturation were calculated. For 100 underwater images taken in Korean seas, the average UIQMs of CycleGAN, UGAN, and FUnIE-GAN were 3.91, 3.42, and 2.66, respectively, and the average UCIQEs were measured to be 29.9, 26.77, and 22.88, respectively. The average UIQM and UCIQE of Image Fusion were 3.63 and 23.59, respectively. CycleGAN and UGAN qualitatively and quantitatively improved the image quality in various underwater environments, and FUnIE-GAN had performance differences depending on the underwater environment. Image Fusion showed good performance in terms of color correction and sharpness enhancement. It is expected that this method can be used for monitoring underwater works and the autonomous operation of unmanned vehicles by improving the visibility of underwater situations more accurately.

1. Introduction

In recent years, unmanned automatic technology has been actively used in order to safely and accurately perform underwater work on behalf of humans in extreme and diversified underwater environments, and image sensing technology that visualizes underwater situations for unmanned automatic technology has become an essential element technology (Schettini and Corchs, 2010; Bharal, 2015). As shown in Fig. 1, from among underwater image sensors, images taken with underwater optical cameras have low contrast and blurred clarity due to the absorption and scattering of light attributed to water and floating matter, such as plankton and sand. The degree of deterioration in image quality varies depending on the turbidity of the water and the characteristics of the photographing camera. Moreover, unlike in an atmospheric environment, the red wavelength of light is longer than the blue or green wavelengths, which means it has more difficulties passing through the water. Therefore, optical images taken underwater tend to emphasize blue or green more than red (Mobley and Mobley, 1994).

Various algorithms have been devised to improve the hue, contrast, and clarity of underwater images. Dark channel prior (DCP) techniques using physical phenomena, in which hues are degraded differently depending on the wavelength of the light, have been applied alongside general image processing algorithms in several stages, and mathematical model-based algorithms for underwater optical imaging processes have been proposed (He et al., 2010; Li et
Comparison of GAN Deep Learning Methods for Underwater Optical Image Enhancement

2. GAN-based Underwater Optical Image Improvement Techniques

GAN consists of two networks: a generator (G) and a discriminator (D) (Fig. 2). Generators and discriminators learn competitively with each other, as generators have to generate fake images that are able to deceive discriminators, and discriminators have to determine the authenticity of fake images created by the generators (Goodfellow et al., 2014). In order to train GAN, the optimization problem is solved for the objective function \( L_{GAN}(G, D) \) as outlined in Eq. (1).

\[
\min_{G} \max_{D} L_{GAN}(G, D) = \log D + \log(1 - D(G(z)))
\]

In Eq. (1), \( z \) is a random input noise, and \( G \) and \( D \) are generator and discriminator functions, respectively. Based on Eq. (1), \( G(z) \) is trained to minimize the value of the objective function by generating fake images and thereby deceiving the \( D \) function, and the \( D \) function is trained to maximize the value of the objective function without being deceived by the fake images generated by \( G \). In this study, CycleGAN, UGAN, and FUnIE-GAN were considered as deep learning-based underwater optical image improvement techniques.

2.1 GAN Model for Improving Underwater Optical Images

CycleGAN was designed to learn a model \((G_{X\rightarrow Y}, G_{Y\rightarrow X})\) that can conduct mutual mapping of data domains \(X\) and \(Y\). CycleGAN was composed of a pair consisting of a generator and a discriminator in the two data domains, and each generator and discriminator network consisted of Resnet-9block (He et al., 2016) and PatchGAN with \(70\times70\) patch sizes, as shown in Figs. 3(a) and 3(b), respectively (Zhu et al., 2017).

The generator, Resnet-9block, repeatedly uses the Residual block 9 times. The network improves performance by adding the initial input value \( x \) one more time to the output, \( F(x) \) in order to learn additional information. PatchGAN, a discriminator, consists of a total of five convolutional layers. The first four convolutions were carried out with stride 2, the last convolution was calculated with stride 1, and a feature map of \( 16\times16 \) was finally obtained using a \( 70\times70 \) patch. In Eq. (2), the objective function of CycleGAN, \( L_{CycleGAN}(G, D) \), is defined with the GAN adversarial objective function of Eq. (1) and the cycle consistency loss function Eq. (3). In Eq. (2), \( x \) and \( y \) refer to data domains, and in Eq. (3), \( x \) and \( y \) represent data in the \( x \) and \( y \) domains, respectively. \( G_{X\rightarrow Y} \) and \( G_{Y\rightarrow X} \) are generators that are converted between data domains into \( X\rightarrow Y \) and \( Y\rightarrow X \), respectively. \( D_{X} \) and \( D_{Y} \) are discriminators that distinguish the authenticity of the fake images generated in each data domain.

Fig. 1 Light attenuation by water and floating particles causes low visibility with haziness and color cast in underwater optical images

Fig. 2 GAN model consisting of a generator (G) and a discriminator (D).
Fig. 3 Network architectures of CycleGAN: (a) generator based on Resnet-9 block and (b) discriminator based on PatchGAN.

Fig. 4 Network architecture of UGAN: (a) generator based on U-Net and (b) discriminator based on PatchGAN.
Comparison of GAN Deep Learning Methods for Underwater Optical Image Enhancement

\[ L_{\text{CycleGAN}}(G_{X \rightarrow Y}, G_{Y \rightarrow X}, D_X, D_Y) = L_{\text{GAN}}(G_{X \rightarrow Y}, D_Y, X; Y) + L_{\text{GAN}}(G_{Y \rightarrow X}, D_X, Y; X) + \lambda \cdot L_{\text{Cycle}}(G_{X \rightarrow Y}, G_{Y \rightarrow X}) \]  \hspace{1cm} (2)

\[ L_{\text{Cycle}}(G_{X \rightarrow Y}) = \|G_{Y \rightarrow X}(G_{X \rightarrow Y}(x)) - x\| + \|G_{X \rightarrow Y}(G_{Y \rightarrow X}(y)) - y\| \]  \hspace{1cm} (3)

The cycle consistency loss function optimally learns the characteristics of clean images while maintaining the characteristics of original underwater images by adding the difference between the generated and input images. CycleGAN was trained by setting the batch size to 6, the learning rate to 0.0002, and the epoch to 60.

The generator and discriminator networks of UGAN were composed of U-Net and PatchGAN, respectively, as shown in Fig. 4 (Ronneberger et al., 2015; Fabbri et al., 2018). The U-Net consisted of eight encoder convolutional layers and seven decoder convolutional layers. PatchGAN consisted of five convolutional layers and distinguished fake and real images by obtaining a feature map of 32×32 using a 70×70 patch. The UGAN objective function of Eq. (4),

\[ L_{\text{GAN}} = \min_{G} \max_{D} L_{\text{WGAN}}(G, D) + \lambda_{L_1} L_{L_1}(G) \]  \hspace{1cm} (4)

\[ L_{\text{WGAN}}(G, D) = D(f^{GT}) - D(f^{IW}) + \lambda_{GP} \]  \hspace{1cm} (5)

\[ L_{L_1} = E[\|f^{GT} - f^{IW}\|] \]  \hspace{1cm} (6)

As presented in Fig. 5, the generator of FUlle-GAN has a U-Net structure consisting of an encoder with five convolutional layers and a decoder with five convolutional layers. The discriminator is a PatchGAN consisting of five convolutional layers that provide a result feature map of 16×16 using a 70×70 patch (Islam et al., 2020). The FUlle-GAN is trained with a batch size of 4, a learning rate of 0.0003, and an epoch of 100. The objective function of FUlle-GAN in Eq. (7), \( L_{\text{FUlle-GAN}} \) includes the adversarial objective function of Eq. (1), the global similarity loss of Eq. (8) which reduces blurred characteristics, and the content loss of Eq. (9) which allows for the

![Fig. 5 Network architecture of FUlle-GAN: (a) generator based on U-Net and (b) discriminator based on PatchGAN](image-url)
generator to extract the characteristics of the ground truth at a high level.

$$L_{F_{\text{EIE-GAN}}} = \min_{G} \max_{D} L_{\text{GAN}}(G, D) + \lambda_1 L_{\text{L2}}(G) + \lambda_2 L_{\text{con}}(G)$$ (7)

$$L_{\text{L2}}(G) = E_{U,W,G,T,Z}[[|\text{GT} - G(U,W,Z)||]]$$ (8)

$$L_{\text{con}}(G) = E_{U,W,G,T,Z}[[\phi(\text{GT}) - \phi(G(U,W,Z))||]]$$ (9)

$U$: GT and $Z$ of Eqs. (8) and (9) are the distorted image, the ground truth image, and random noise, respectively. Eq. (8) improves the blurry part by using the difference in clarity between the image generated by the generator $G$ and the ground truth image. Eq. (9) uses the $\phi$ function which is learned in advance using the VGG-19 network and extracts the characteristics of the image.

The characteristics of the ground truth image and the image generated by the generator are extracted using the $\phi$ function to provide a loss value that enables the creation of an image that is as similar as possible to the target image.

2.2. Learning Data Set

For GAN-based underwater image learning improvement techniques, two learning datasets were configured with underwater images and clean images. In the first learning dataset (unpaired), there were a total of 5,202 underwater images, consisting of 1,813 images from the ImageNet database and 3,389 underwater images directly taken in Korea's waters (Geoje, Donghae, Jeju, etc.), and 6,050 clean images which had been selected from the ImageNet database (Deng et al., 2009).

First, CycleGAN was trained with unpaired datasets to generate clean images of the scenes depicted in the underwater images, and a total of 5202 second datasets (paired) were constructed (Fig. 6). CycleGAN, UGAN, and FUNIE-GAN, which were compared in this study, were trained with paired datasets. The training was carried out using Python version 3.6, Tensorflow version 2.4.1, Keras version 2.1.6, and two NVIDIA GeForce RTX 2080 Ti GPUs.

3. Image Fusion-Based Underwater Optical Image Improvement Techniques

As shown in Fig. 7, the image fusion technique consists of a white balancing step for color tone correction, a step for enhancing boundaries and contrast, and a step for the fusion of two enhanced images (Ancuti et al., 2017). As the energy of green light is higher than that of red and blue light, it penetrates relatively more deeply underwater and is more likely to reach the camera. The white balancing step of the image fusion technique uses green light with a good signal transduction to compensate for the signal attenuation of red light and blue light. Eqs. (10) and (11) compensate for the red light $I_r$ and blue light $I_b$ of each pixel in the white balancing image, respectively. The compensation is conducted using the average values of each color tone ($I_r$, $I_g$, $I_b$), the color values of one pixel ($I_r$, $I_g$, $I_b$), and the weights $\alpha$ and $\beta$.

$$WB_r = I_r(x) + \alpha(I_r - I_r)(1 - I_r(x))$$ (10)

$$WB_b = I_b(x) + \beta(I_b - I_b)(1 - I_b(x))$$ (11)

Because images photographed in different underwater environments show different hue biases, it is necessary to set weights $\alpha$ and $\beta$ of Eqs. (10) and (11) differently according to the images. In order to determine $\alpha$ and $\beta$, the values of red and blue light after the white balance were replaced with the average values of the green light, which has a relatively low attenuation, and the weight values were derived as outlined in Eqs. (12) and (13).

![Fig. 6 Unpaired and paired training datasets](image-url)
Comparison of GAN Deep Learning Methods for Underwater Optical Image Enhancement

Fig. 7 Three image processing steps of image fusion enhancement method

\[ a = \frac{1}{1 - (I_1(x))^2} \]  
\[ \beta = \frac{1}{1 - (I_2(x))^2} \]  

In the second image enhancement step, the boundary signal and the contrast were enhanced, respectively. The unsharp masking principle technique in Eq. (14) for boundary enhancement applies a Gaussian filter to extract and reinforce boundary difference information with respect to the original image. \( WB \) is an image in which any color tone bias is corrected for in the white balancing step, and \( \gamma \) and \( Gauss \) are the weight and Gaussian filter functions, respectively, which both describe the degree to which boundary difference information is applied. For enhancing the contrast, the contrast limit adaptive histogram equalization (CLAHE) method was utilized. After dividing the image into blocks of a fixed size, smoothing the histogram of each block was performed to obtain an image with enhanced contrast.

\[ UMP[ WB] = WB + \gamma \cdot WB - (Gauss \cdot WB) \]  
\[ J = \omega_1 UMP[ WB] + \omega_2 CLAHE[ WB] \]  

In the third image fusion step of Eq. (15), the boundary enhancement result image and the contrast enhancement result image of the second step are fused using weights to obtain the resulting final improved image \( J \). In Eq. (15), the weight uses \( \omega_1 \) which is a value obtained by applying a Laplacian filter to the luminance channel and \( \omega_2 \), which is a difference value between the hue channel and the luminance channel.

4. Results

Fig. 8 qualitatively shows the results of the image fusion of underwater images taken in the real sea and learned GAN techniques.

(a) Original and enhanced underwater images by (b) CycleGAN-Unpaired, (c) CycleGAN-Paired, (d) UGAN, (e) FUnIE-GAN, (f) Image fusion
CycleGAN and UGAN showed significant improvements in image quality overall in all underwater environments; however, in some cases, it was confirmed that artifacts and noise were produced (fish body part in Fig. 8(c)). Moreover, in the case of FUnIE-GAN, its performance showed a large deviation according to the underwater environment. In terms of image fusion, it was confirmed that the hue correction was improved, but that the noise in the parts affected by reflected light could not be improved.

UIQM and UCIQE were measured to quantitatively compare the underwater optical image improvement techniques considered in this study. UIQM is an index that evaluates the quality of underwater images considering the sum of the weights of colorfulness, sharpness, and contrast, and is expressed as in Eq. (16) (Panetta et al., 2015). Three weights were estimated based on the colorfulness, sharpness, and contrast of 30 images obtained from various underwater environments and using various imaging equipment in the original reference, in which UIQM was devised, and they were set to the following values: \( a_1 = 0.0282 \), \( a_2 = 0.2983 \), and \( a_3 = 0.0339 \).

\[
UIQM = a_1 UICM + a_2 UISM + a_3 UIGmM
\] (16)

UCIQE is an index calculated from the sum of the weights of the chroma, luminance, and the saturation of colors (Yang and Sowmya 2015). It is expressed as in Eq. (17), and the weights were estimated based on the chroma, luminance, and saturation of the underwater images in the original reference in which UCIQE was devised. They were set to the following values: \( b_1 = 0.4680 \), \( b_2 = 0.2745 \), and \( b_3 = 0.2576 \).

\[
UCIQE = b_1 \sigma_{\text{chr}} + b_2 \text{con}_{L, c} + b_3 \text{sat}
\] (17)

Table 1 and Fig. 9 quantitatively compare the mean values and deviations of UIQM and UCIQE of 100 underwater images from real sea areas and images improved with GAN techniques and image fusion techniques. All underwater optical image improvement techniques showed improved UIQM and UCIQE compared with underwater images from real sea areas. Among the GAN techniques, CycleGAN-UnPaired showed the best average UIQM and UCIQE. UGAN using the paired dataset and image fusion techniques showed similar performance. In the case of FUnIE-GAN, the largest deviation between UIQM and UCIQE was observed.

Table 2 compares the average calculation time of the studied underwater optical image improvement methods. In order to make a comparison under the same conditions, the size of a total of 100 underwater images used was readjusted to 256×256. Among the GAN techniques, UGAN showed the fastest computational speed with a speed similar to the image fusion techniques.

![Fig. 9 Comparative graphs of (a) mean UIQM and (b) mean UCIQE computed from 100 underwater images.](image-url)
5. Conclusions

In this paper, to improve the quality of underwater images, CycleGAN, UGAN, and FUnIE-GAN deep learning techniques and image fusion methods were applied to 100 underwater images taken in real sea areas, and the quality improvement was compared and evaluated based on UIQM and UCIQE. CycleGAN and UGAN showed excellent performance qualitatively and quantitatively in various underwater environments, but artifacts were generated in some cases. FUnIE-GAN had a very large performance variation depending on the underwater environment, and image fusion showed good performance in color tone correction and sharpness improvement. It is expected that an underwater situation can be more accurately visualized through the improvement of underwater images, which can then be used for monitoring underwater works and planning autonomous driving routes for unmanned vehicles.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Funding

This study was conducted with the support of the National Research Foundation of Korea (Development of Underwater Stereo Camera Stereoscopic Visualization Technology, NRF-2021R1A2C2006682) with funding from the Ministry of Science and ICT in 2021 and the Korea Institute of Marine Science & Technology Promotion (establishment of test evaluation ships and systems for the verification of the marine equipment performance in real sea areas) with funding from the Ministry of Science and ICT in 2021 and the Korea Institute of Marine Science & Technology Foundation of Korea (Development of Underwater Stereo Camera Stereoscopic Visualization Technology, NRF-2021R1A2C2006682).

References


Author ORCIDs

<table>
<thead>
<tr>
<th>Author name</th>
<th>ORCID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kim, Hong-Gi</td>
<td>0000-0001-7731-8429</td>
</tr>
<tr>
<td>Seo, Jung-Min</td>
<td>0000-0001-6323-7783</td>
</tr>
<tr>
<td>Kim, Soo Mee</td>
<td>0000-0001-8414-1297</td>
</tr>
</tbody>
</table>