Underwater Image Enhancement

Arkadip Bhattacharya[1], Arunangshu Biswas[2], Pranjal Gain[3], and Dr. Abul Hasnat[+]

[+]Guide, Assistant Professor
[1,2,3]Government College of Engineering and Textile Technology, Berhampore

August 13, 2021

Abstract

In this report, we present an approach for real-time underwater image enhancement using Cycle Generative Adversarial Network, or *CycleGan*. The objective is to train the model to learn the mapping between the underwater image and the enhanced image using a training set of aligned image pairs. For the training we used the **EUVP** (Enhancing Underwater Visual Perception) a large scale dataset of a paired and an unpaired collection of underwater images (of 'poor' and 'good' quality) that are captured using seven different cameras over various visibility conditions during oceanic explorations and human-robot collaborative experiments. For our case, we have only supplied the images in an unpaired manner during the training for better generalisation of the mapping. The model and associated training pipelines are available at https://github.com/darkmatter18/Underwater-image-enhancement.

1   Introduction

Underwater optical imaging (OPI) is a challenging field in computer vision research. Due to the medium, scattering always causes a blurring effect in underwater photography. The wavelength absorption usually causes a colour reduction in the captured images as well. In the shallow ocean, sunshine causes flickering effects. This causes the captured images with strong highlights. Whereas, the usage of artificial lighting for underwater photography causes vignetting in captured images due to the non-uniform nature of the lighting. Common issues with underwater images, such as light attenuation, scattering, non-uniform lighting, shadows, colour shading, suspended particles, or the abundance of marine life, can be overcome via underwater optical image processing.

Visually-guided AUVs (Autonomous Underwater Vehicles) and ROVs (Remotely Operated Vehicles) are widely used in important applications such as the monitoring of marine species migration and coral reefs [41], inspection of submarine cables and wreckage [5], underwater scene analysis, seabed mapping, human-robot collaboration [26], and more. One major operational challenge for these underwater robots is that despite using high-end cameras, visual sensing is often greatly affected by poor visibility, light refraction, absorption, and scattering [34, 52, 26]. These optical artifacts trigger non-linear distortions in the captured images, which severely affect the performance of vision based tasks such as tracking, detection and classification, segmentation, and visual servoing. Fast and accurate image enhancement techniques can alleviate these problems by restoring the perceptual and statistical qualities [15, 52] of the distorted images in real-time.

This problem can be broadly classified as image-to-image translation [27], converting an image from one representation of a given scene, x, to another, y. Years of research in computer vision, image processing, computational photography, and graphics have produced powerful translation systems in the supervised setting, where example image pairs $\{x_i, y_i\}_{i=1}^{N}$ are available (Figure 1, left), e.g., [14, 23, 27, 28, 30, 33, 40, 45, 47, 51]. However, obtaining paired training data can be difficult and expensive at times.

To tackle this issue, we present an approach of using Cycle Generative Adversarial Network **(CycleGan)** for enhancement of the underwater photographs in real-time. The objective is for the model to learn about the various said effects present in an underwater photograph and generate a corresponding enhanced image, which is free of them.
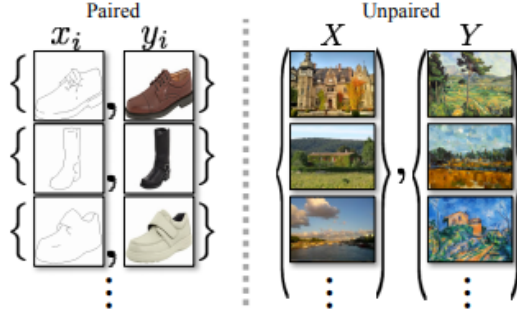
Figure 1: Paired training data (left) consists of training examples $\{x_i, y_i\}_{i=1}^N$, where the correspondence between x_i and y_i exists [27], We instead consider unpaired training data (right), consisting of a source set $\{x_i\}_{i=1}^N (x_i \epsilon X)$ and a target set $\{y_j\}_{j=1}^M (y_j \epsilon Y)$, with no information provided as to which x_i matches which y_j.

CycleGan is ideal for this task as it is specifically designed to be able to learn translation between domains without paired input-output examples (Figure 1, right).

2 Related Work

2.1 Automatic Image Enhancement

Automatic image enhancement is a well-studied problem in the domains of computer vision, robotics, and signal processing. Classical approaches use hand-crafted filters to enforce local color constancy and improve contrast/lightness rendition [44]. Additionally, prior knowledge or statistical assumptions about a scene (e.g., haze-lines, dark channel prior [4], etc.) are often utilized for global enhancements such as image deblurring, dehazing [21], etc. Over the last decade, single image enhancement has made remarkable progress due to the advent of deep learning and the availability of large-scale datasets. The contemporary deep CNN-based models provide state-of-the-art performance for problems such as image colorization [51], color/contrast adjustment [10], dehazing [8], etc. These models learn a sequence of non-linear filters from paired training data, which provide much better performance compared to using handcrafted filters.

Moreover, the GAN-based models [19] have shown great success for style-transfer and image-to-image translation problems [27]. They employ a two-player min-max game where the 'generator' tries to fool the 'discriminator' by generating fake images that appear to be sampled from the real distribution. Simultaneously, the discriminator tries to get better at discarding fake images and eventually (in equilibrium) the generator learns to model the underlying distribution. Although such adversarial training can be unstable, several tricks and choices of loss functions are proposed in the literature to mitigate that. For instance, Wasserstein GAN [3] improves the training stability by using the earth-mover distance to measure the distance between the data distribution and the model distribution. Energy-based GANs [53] also improve training stability by modeling the discriminator as an energy function, whereas the LeastSquared GAN [35] addresses the vanishing gradients problem by adopting a least-square loss function for the discriminator. On the other hand, conditional GANs [36] allow constraining the generator to produce samples that follow a pattern or belong to a specific class, which is particularly useful to learn a pixel-to-pixel (Pix2Pix) mapping [27] between an arbitrary input domain (e.g., distorted images) and the desired output domain (e.g., enhanced images).

A major limitation of the above-mentioned models is that they require paired training data, which may not be available or can be difficult to acquire for many practical applications. The two-way GANs (e.g., CycleGAN [54], DualGAN [49], etc.) solve this problem by using a 'cycleconsistency loss' that allows learning the mutual mappings between two domains from unpaired data. Such models have been effectively used for unpaired learning of perceptual image enhancement [9] as well. Furthermore, Ignatov et al. [25] showed that additional loss-terms for preserving the high-level feature-based content improve the quality of image enhancement using GANs.

2.2 Improving Underwater Visual Perception

Traditional physics-based methods use the atmospheric dehazing model to estimate the transmission and ambient light in a scene to recover true pixel intensities [11, 7]. Another class of methods design a series of bilateral and trilateral filters to reduce noise and improve global contrast [34, 52]. In recent work, Akkaynaket al. [1] proposed a revised imaging model that accounts for the unique distortions pertaining to underwater light propagation; this contributes to a more accurate color reconstruction and overall a better approximation to the ill-posed underwater image enhancement problem. Nevertheless, these methods require scene depth (or multiple images) and optical waterbody measurements as prior.

On the other hand, several single image enhancement models based on deep adversarial [16, 50, 31] and residual learning [32] have reported inspiring results of late. However, they typically use only synthetically distorted images for paired training, which often limit their generalization performance. The extent of large-scale unpaired training on naturally distorted underwater images have not been explored in the literature. Moreover, most existing models fail to ensure fast inference on single-board robotic platforms, which limits their applicability for improving real-time visual perception. We attempt to address these aspects in this paper.
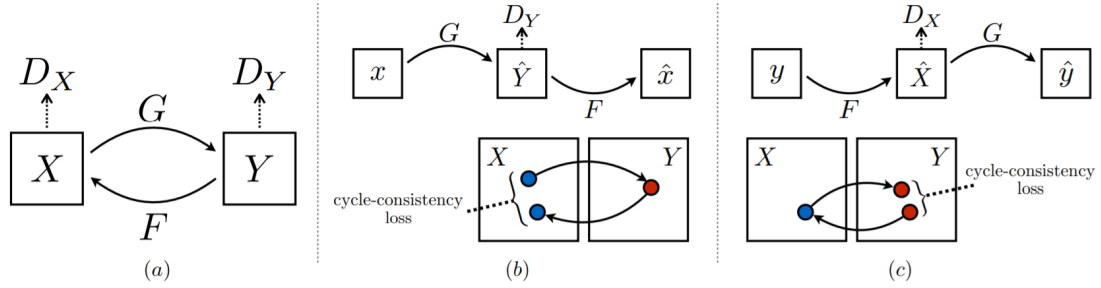
3 Proposed Model

3.1 CycleGan Architecture

Figure 2: (a) Our model contains two mapping functions $G : X \rightarrow Y$ and $F : Y \rightarrow X$, and associated adversarial discriminators D_Y and D_X. D_Y encourages G to translate X into outputs indistinguishable from domain Y, and vice versa for DX and F. To further regularize the mappings, we introduce two cycle consistency losses that capture the intuition that if we translate from one domain to the other and back again we should arrive at where we started: (b) forward cycle-consistency loss: $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$, and (c) backward cycle-consistency loss: $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$

3.2 Generator

We adopt the architecture for our generative networks from DCGAN [38]. The network contains 3 convolution and several residual blocks [22], two fractionally-strided convolutions with stride $\frac{1}{2}$, and one convolution that maps features to RGB. We use 6 blocks for 128 × 128 images and 9 blocks for 256×256 and higher-resolution training images. Similar to [28].

We use instance normalization [43] and used Leaky ReLU [48] as activation layer. For the last layer, we used hyperbolic tangent as the activation function [39].

3.3 Discriminator

The discriminator networks are based on convolutional neural network (CNN)[2]. It uses 9 layers of convolutional layer to extract deep features from the image and classify between 2 sets of images.

4 Objective Function Formulation

Our goal is to learn mapping functions between two domains X and Y given training samples $x_i{}_{i=1}^N$ where $x_i \in X$ and $y_j{}_{j=1}^N$ where $y_j \in Y$. We denote the data distribution as $x \sim pdata(x)$ and $y \sim pdata(y)$. As illustrated in Figure 2 (a), our model includes two mappings $G : X \rightarrow Y$ and $F : Y \rightarrow X$. In addition, we introduce two adversarial discriminators D_X and D_Y, where D_X aims to distinguish between images x and translated images $F(y)$; in the same way, D_Y aims to discriminate between y and $G(x)$. Our objective contains two types of terms: adversarial losses [18] for matching the distribution of generated images to the data distribution in the target domain; and cycle consistency losses to prevent the learned mappings G and F from contradicting each other.

4.1 Adversarial Loss

A standard conditional GAN-based model learns a mapping $G : X, Z \rightarrow Y$, where $X(Y)$ represents the source(desired) domain, and Z denotes random noise. The conditional adversarial loss function[36] is expressed as:

$$\mathcal{L}_{GAN}N(G, D_Y, X, Y) = \mathbb{E}_{X,Y}[\log D_Y(Y)] + \mathbb{E}_{X,Y}[\log 1 - D(X, G(X, Z))] \tag{1}$$

Here, the generator G tries to minimize $\mathcal{L}_{cGAN}$ while the discriminator D_Y tries to maximize it.

4.2 Cycle Consistency Loss

Adversarial training can, in theory, learn mappings G and F that produce outputs identically distributed as target domains Y and X respectively (strictly speaking, this requires G and F to be stochastic functions) [17]. However, with large enough capacity, a network can map the same set of input images to any random permutation of images in the target domain, where any of the learned mappings can induce an output distribution that matches the target distribution. Thus, adversarial losses alone cannot guarantee that the learned function can map an individual input x_i to a desired output y_i. To further reduce the space of possible mapping functions, we argue that the learned mapping functions should be cycle-consistent functions should be cycle-consistent: as shown in Figure 2 (b), for each image x from domain X, the image translation cycle should be able to bring x back to the original image, i.e., $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$

$$\mathcal{L}_{cyc}(G, F) = \mathbb{E}_{x \sim p_{data}(x)}[\|F(G(x)) - x\|] + \mathbb{E}_{y \sim p_{data}(y)}[\|F(G(y)) - y\|] \tag{2}$$

4.3 Full Objective

Our Full objective is:

$$\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{GAN}(G, D_Y, X, Y) + \mathcal{L}_{GAN}(F, D_X, Y, X) + \lambda\mathcal{L}_{cyc}(G, F) \tag{3}$$

where λ controls the relative importance of the two objectives. We aim to solve:

$$G^*, F^* = arg \min_{G,F} \max_{D_X,D_Y} \mathcal{L}(G, F, D_X, D_Y) \tag{4}$$

5 Training

To minimize the objective function 4, we used Adam optimizer [29] with a batch size of 8. All the networks are trained from ground up with initial learning rate of 0.0002. We keep the same learning rate for first 100 epochs, then linearly decay the learning rate to 0 over the next 100 epochs.

6 EUVP Dataset

The EUVP dataset contains a large collection of paired and unpaired underwater images of poor and good perceptual quality. We used seven different cameras, which include multiple GoPros [20], Aqua AUV's uEye cameras [13], low-light USB cameras [6], and Trident ROV's HD camera [42], to capture images for the dataset. The data was collected during oceanic explorations and human-robot cooperative experiments in different locations under various visibility conditions. Additionally, images extracted from a few publicly available $YouTube^{TM}$ videos are included in the dataset. The images are carefully selected to accommodate a wide range of natural variability (e.g., scenes, waterbody types, lighting conditions, etc.) in the data.

The unpaired data is prepared, i.e., good and poor quality images are separated based on visual inspection by six human participants. They inspected several image properties (e.g., color, contrast, and sharpness) and considered whether the scene is visually interpretable, i.e., foreground/objects are identifiable. Hence, the unpaired training endorses the modeling of human perceptual preferences of underwater image quality. On the other hand, the paired data is prepared by following a procedure suggested in [16]. Specifically, a CycleGAN [54]-based model is trained on our unpaired data to learn the domain transformation between the good and poor quality images. Subsequently, the good quality images are distorted by the learned model to generate respective pairs; we also augment a set of underwater images from the ImageNet dataset [12] and from $Flickr^{TM}$.

(a) Paired instances: ground truth images and their respective distorted pairs are shown on the top and bottom row, respectively.

(b) Unpaired instances: good and poor quality images are shown on the top and bottom row (in no particular order), respectively.
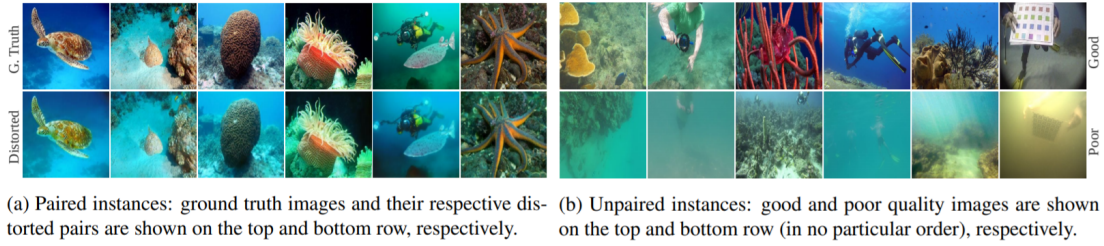
Figure 3: A few sample images from the EUVP dataset are shown.

There are over 12K paired and 8K unpaired instances in the EUVP dataset; a few samples are provided in Fig. 3. It is to be noted that our focus is to facilitate perceptual image enhancement for boosting robotic scene understanding, not to model the underwater optical degradation process for image restoration, which requires scene depth and waterbody properties.

7 Experimental Results

7.1 Qualitative Evaluations

We first qualitatively analyse the output images generated by the model compared to their respective ground truths. As figure 4 shows, the bluish hue which is characteristic of underwater images is greatly reduced. Additionally, the model tries to recover the original colour of the objects in the photographs. If trained further the model would be able to learn about the natural colours of the underwater objects and will be able to recreate it much better. Since we only provided unpaired images, the model identified the common artefacts that are present in the A type images but is absent on B type images.
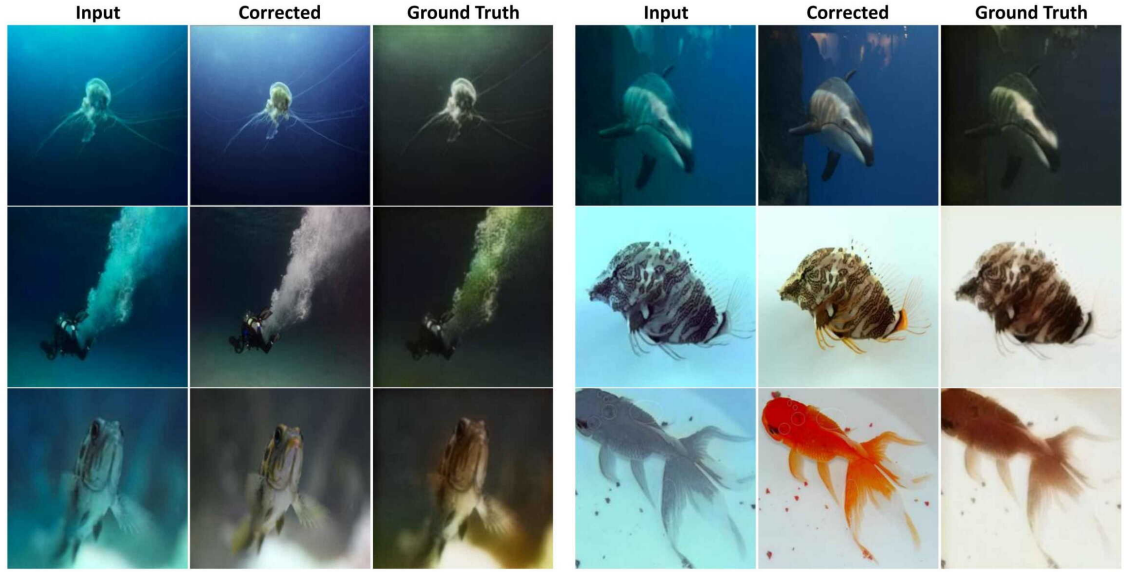
Figure 4: The bluish hue is rectified and the model tries to regenerate the original colours of the objects.

7.2 Quantitative Evaluation

We consider four standard metrics [25, 9, 24] named Peak Signal-to-Noise Ratio (PSNR), Structural Similarity (SSIM), Entrophy and Underwater Image Quality Metric (UIQM) in order to quantitatively compare CycleGan enhanced images with their respective ground truths.

7.2.1 Peak Signal-to-Noise Ratio (PSNR)

The PSNR approximates the reconstruction quality of a generated image x compared to its ground truth y based on their Mean Squared Error (MSE) as follows:

$$PSNR(x,y) = 10\log_{10}[255^2/MSE(x,y)] \tag{5}$$

7.2.2 Structural similarity index measurement (SSIM)

SSIM [46] compares the image patches based on three properties: luminance, contrast, and structure. It is defined as follows:

$$SSIM(x,y) = (\frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1})(\frac{2\sigma_{xy} + c_2}{\sigma_x^2 + \sigma_y^2 + c_2}) \tag{6}$$

In equation 6, μ_x (μ_y) denotes the mean, and σ_x^2 (σ_y^2) is the variance of x (y); whereas σ_{xy} denotes the cross correlation between x and y. Additionally, $c_1 = (255 \times 0.01)^2$ and $c_2 = (255 \times 0.03)^2$ are constants that ensure numeric stability.

7.2.3 Entropy

Entropy is a statistical measure of randomness that can be used to characterize the texture of the input image.

The Shannon entropy is defined as:

$$S = -\sum(p_k \times \log(p_k)) \tag{7}$$

In equation 7, p_k is the frequency/probability of pixels of value k.

7.2.4 Underwater Image Quality Metric (UIQM)

The *underwater image quality measure (UIQM)* comprises three attribute measures, namely, the *underwater image colorfulness measure (UICM)*, the *underwater image sharpness measure (UISM)*, and the *underwater image contrast measure (UIConM)*. Details of the new measures are presented in this section.

Underwater Image Colorfulness Measure (UICM)   The UICM deals with the color casting problem. As the depth of the water increases, colors attenuate one by one depending on their wavelength. The color red disappears first due to it possessing the shortest wavelength. As a result, underwater images usually demonstrate a bluish or greenish appearance. Furthermore, limited lighting conditions also causes severe color desaturation in underwater images.

To measure the UICM, we use the two opponent color components related with chrominance, RG and YB as shown in

$$RG = R - G \tag{8}$$

$$YB = \frac{R+G}{2} - B \tag{9}$$

Underwater images usually suffer from heavy noise. Therefore, instead of using the regular statistical values, the asymmetric alpha-trimmed statistical values are used for measuring underwater image colorfulness. The mean is defined by:

$$\mu_{\alpha,RG} = \frac{1}{K - T_{\alpha L} - T_{\alpha R}} \sum_{i=T_{\alpha L}+1}^{K-T_{\alpha R}} Intensity_{RG,i} \tag{10}$$

The second-order statistic variance σ^2 in:

$$\sigma_{\alpha,RG}^2 = \frac{1}{N} \sum_{p=1}^{N} (Intensity_{RG,p} - \mu_{\alpha,RG})^2 \tag{11}$$

The overall colourfulness metric used for measuring underwater image colourfulness is demonstrated as

$$UICM = -0.0268\sqrt{\mu_{\alpha,RG}^2 + \mu_{\alpha,YB}^2} + 0.1586\sqrt{\sigma_{\alpha,RG}^2 + \sigma_{\alpha,YB}^2} \tag{12}$$

Underwater Image Sharpness Measure (UISM)   To measure the sharpness on edges, the Sobel edge detector is first applied on each RGB color component. The resultant edge map is then multiplied with the original image to get the grayscale edge map. By doing this, only the pixels on the edges from the original underwater image are preserved. It is known that the enhancement measure estimation (EME) measure is suitable for images with uniform background and shown non-periodic patterns

The UISM is formulated as shown in

$$UISM = \sum_{c=1}^{3} \lambda_c EME(greyscale, edge_c) \tag{13}$$

$$EME = \frac{2}{k_1 k_2} \sum_{l=1}^{k1} \sum_{k=1}^{k2} \log(\frac{I_{max,k,l}}{I_{min,k,l}}) \tag{14}$$

Underwater Image Contrast Measure (UIConM)   Contrast has been shown to correspond to underwater visual performance such as stereoscopic acuity [37]. For underwater images, contrast degradation is usually caused by backward scattering

$$UIConM = \log AMEE(Intensity) \tag{15}$$

The $\log AMEE$ in

$$\log AMEE = \frac{1}{k_1 k_2} \bigotimes \sum_{l=1}^{k_1} \sum_{k=1}^{k_2} \frac{I_{max,k,l} \Theta I_{min,k,l}}{I_{max,k,l} \bigoplus I_{min,k,l}} \times \log(\frac{I_{max,k,l} \Theta I_{min,k,l}}{I_{max,k,l} \bigoplus I_{min,k,l}}) \tag{16}$$

The overall equation for the **UIQM** is

$$UIQM = c_1 \times UICM + c_2 \times UISM + c_3 \times UIConM \tag{17}$$

A generic coefficient set $c_1 = 0.0282$, $c_2 = 0.2953$, and $c_3 = 3.5753$ is used as suggested by [37]
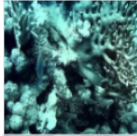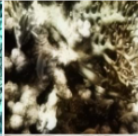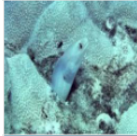
7.2.5   Metrics
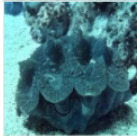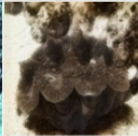
Input	Corrected	PSNR	SSIM	Entropy Input	Entropy Corrected	UIQM Input	UIQM Corrected
		20.72530771	0.765697359	6.99245163	6.886215304	-0.3075320674	-0.08049992083
		19.22043369	0.7609550976	7.429374905	7.163665066	-0.3144234176	-0.2271329756
		16.73832512	0.7570294401	7.382421505	7.174155117	0.1718645517	0.09861603291
		20.50234263	0.7358438048	6.07705762	6.868067892	-0.4030989758	-0.3179808391
		16.81580671	0.734311788	7.431492596	7.748927059	0.2206967794	0.4571219123

Figure 5: Metrics for top 5 enhanced images.

8   Conclusion

In this project we present a simple yet efficient GAN-based model for **underwater image enhancement**. With the help of *Cycle Generative Adversarial Network (CycleGAN)* the underwater enhancing model can generally and effectively transform underwater images with low visible quality into clean images with richer details and higher visible quality. Our study shows the effectiveness of each module of the proposed model. The proposed model formulates a perceptual loss function by evaluating image quality based on its global color, content, local texture and style information. We performed extensive qualitative and quantitative evaluations. The qualitative and quantitative evaluation results have shown that the proposed model performs as well and often better than the original quality images. We believe that our project demonstrates the potential of CycleGAN for improving the translation quality of underwater images.

References

[1] Derya Akkaynak and Tali Treibitz. A revised underwater image formation model. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6723–6732, 2018.

[2] Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi. Understanding of a convolutional neural network. In *2017 International Conference on Engineering and Technology (ICET)*, pages 1–6, 2017.

[3] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein gan, 2017.

[4] Dana Berman, Deborah Levy, Shai Avidan, and Tali Treibitz. Underwater single image color restoration using haze-lines and a new quantitative dataset, 2019.

[5] B. Bingham, B. Foley, H. Singh, R. Camilli, K. Delaporta, R. Eustice, A. Mallios, D. Mindell, C. Roman, and D. Sakellariou. Robotic tools for deep water archaeology: Surveying an ancient shipwreck with an autonomous underwater vehicle. *J. Field Robotics*, 27:702–717, 2010.

[6] BlueRobotics. Low-light hd usb camera. https://www.bluerobotics.com/, 2016. accessed: 3-15-2019.

[7] Mitch Bryson, Matthew Johnson-Roberson, Oscar Pizarro, and Stefan Williams. True color correction of autonomous underwater vehicle imagery. *Journal of Field Robotics*, 33, 09 2015.

[8] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11):5187–5198, 2016.

[9] Yu-Sheng Chen, Yu-Ching Wang, Man-Hsin Kao, and Yung-Yu Chuang. Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6306–6314, 2018.

[10] Zezhou Cheng, Qingxiong Yang, and Bin Sheng. Deep colorization, 2016.

[11] Younggun Cho, Jinyong Jeong, and Ayoung Kim. Model-assisted multiband fusion for single image enhancement and applications to robot vision. *IEEE Robotics and Automation Letters*, 3(4):2822–2829, 2018.

[12] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009.

[13] Gregory Dudek, Philippe Giguère, Chris Prahacs, Shane Saunderson, Junaed Sattar, L. Abril Torres-Méndez, Michael Jenkin, Andrew German, Andrew Hogue, Arlene Ripsman, James Zacher, Evangelos Milios, Hui Liu, Pifu Zhang, Martin Buehler, and Christina Georgiades. Aqua: An amphibious autonomous robot. *Computer*, 40:46 – 53, 02 2007.

[14] David Eigen and Rob Fergus. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture, 2015.

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

[16] Cameron Fabbri, Md Jahidul Islam, and Junaed Sattar. Enhancing underwater imagery using generative adversarial networks, 2018.

[17] Ian Goodfellow. Nips 2016 tutorial: Generative adversarial networks, 2017.

[18] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014.

[19] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014.

[20] GoPro Hero 5 GoPro. https://gopro.com/, 2016. accessed: 3-15-2019.

[21] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2341–2353, 2011.

[22] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.

[23] Aaron Hertzmann, Charles E. Jacobs, Nuria Oliver, Brian Curless, and David H. Salesin. Image analogies. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '01, page 327–340, New York, NY, USA, 2001. Association for Computing Machinery.

[24] Alain Horé and Djemel Ziou. Image quality metrics: Psnr vs. ssim. In *2010 20th International Conference on Pattern Recognition*, pages 2366–2369, 2010.

[25] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. Dslr-quality photos on mobile devices with deep convolutional networks, 2017.

[26] Md Jahidul Islam, Marc Ho, and Junaed Sattar. Understanding human motion and gestures for underwater human–robot collaboration. *Journal of Field Robotics*, 36(5):851–873, 2019.

[27] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks, 2018.

[28] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution, 2016.

[29] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.

[30] Pierre-Yves Laffont, Zhile Ren, Xiaofeng Tao, Chao Qian, and James Hays. Transient attributes for high-level understanding and editing of outdoor scenes. *ACM Transactions on Graphics (proceedings of SIGGRAPH)*, 33(4), 2014.

[31] Jie Li, Katherine Skinner, Ryan Eustice, and Matthew Johnson-Roberson. Watergan: Unsupervised generative network to enable real-time color correction of monocular underwater images. *IEEE Robotics and Automation Letters*, PP, 02 2017.

[32] Peng Liu, Guoyu Wang, Hao Qi, Chufeng Zhang, Haiyong Zheng, and Zhibin Yu. Underwater image enhancement with a deep residual framework. *IEEE Access*, 7:94614–94629, 2019.

[33] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation, 2015.

[34] Huimin Lu, Yujie Li, and Seiichi Serikawa. Underwater image enhancement using guided trigonometric bilateral filter and fast automatic color correction. In *2013 IEEE International Conference on Image Processing*, pages 3412–3416, 2013.

[35] Xudong Mao, Qing Li, Haoran Xie, Raymond Y. K. Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks, 2017.

[36] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets, 2014.

[37] Karen Panetta, Chen Gao, and Sos Agaian. Human-visual-system-inspired underwater image quality measures. *IEEE Journal of Oceanic Engineering*, 41(3):541–551, 2016.

[38] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks, 2016.

[39] Maicon A. Sartin and Alexandre C. R. da Silva. Approximation of hyperbolic tangent activation function using hybrid methods. In *2013 8th International Workshop on Reconfigurable and Communication-Centric Systems-on-Chip (ReCoSoC)*, pages 1–6, 2013.

[40] Yichang Shih, Sylvain Paris, Frédo Durand, and William Freeman. Data-driven hallucination of different times of day from a single outdoor photo. *ACM Transactions on Graphics (TOG)*, 32, 11 2013.

[41] Florian Shkurti, Anqi Xu, Malika Meghjani, Juan Camilo Gamboa Higuera, Yogesh Girdhar, Philippe Giguère, Bir Bikram Dey, Jimmy Li, Arnold Kalmbach, Chris Prahacs, Katrine Turgeon, Ioannis Rekleitis, and Gregory Dudek. Multi-domain monitoring of marine environments using a heterogeneous robot team. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1747–1753, 2012.

[42] OpenROV. TRIDENT. https://www.openrov.com/, 2017. accessed: 3-15-2019.

[43] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization, 2017.

[44] Zia ur Rahman, Daniel J. Jobson, and Glenn A. Woodell. Retinex processing for automatic image enhancement. *Journal of Electronic Imaging*, 13(1):100 – 110, 2004.

[45] Xiaolong Wang and Abhinav Gupta. Generative image modeling using style and structure adversarial networks, 2016.

[46] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[47] Saining Xie and Zhuowen Tu. Holistically-nested edge detection, 2015.

[48] Bing Xu, Naiyan Wang, Tianqi Chen, and Mu Li. Empirical evaluation of rectified activations in convolutional network, 2015.

[49] Zili Yi, Hao Zhang, Ping Tan, and Minglun Gong. Dualgan: Unsupervised dual learning for image-to-image translation, 2018.

[50] Xiaoli Yu, Yanyun Qu, and Ming Hong. Underwater-gan: Underwater image restoration via conditional generative adversarial network. In Zhaoxiang Zhang, David Suter, Yingli Tian, Alexandra Branzan Albu, Nicolas Sidère, and Hugo Jair Escalante, editors, *Pattern Recognition and Information Forensics*, pages 66–75, Cham, 2019. Springer International Publishing.

[51] Richard Zhang, Phillip Isola, and Alexei A. Efros. Colorful image colorization, 2016.

[52] Shu Zhang, Ting Wang, Junyu Dong, and Hui Yu. Underwater image enhancement via extended multiscale retinex. *Neurocomputing*, 245:1–9, 2017.

[53] Junbo Zhao, Michael Mathieu, and Yann LeCun. Energy-based generative adversarial network, 2017.

[54] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks, 2020.