

Underwater Image Dehazing via Unpaired Image-to-image Translation

Younggun Cho, Hyesu Jang, Ramavtar Malav, Gaurav Pandey, and Ayoung Kim*

Abstract: Underwater imaging has long been focused on dehazing and color correction to address severe degradation in the water medium. In this paper, we propose a learning-based image restoration method that uses Generative Adversarial Networks (GAN). For network generality and learning flexibility, we constituted unpaired image translation frameworks into image restoration. The proposed method utilizes multiple cyclic consistency losses that capture image characteristics and details of underwater images. To prepare unpaired images of clean and degraded scenes, we collected images from Flickr and filtered out false images using image characteristics. For validation, we extensively evaluated the proposed network on simulated and real underwater hazy images. Also, we tested our method on conventional computer vision algorithms, such as the level of edges and feature matching results.

Keywords: Dehazing, GAN, underwater vision.

1. INTRODUCTION

Optic cameras and sonar are widely applied for underwater perception. Sonar is a specialized sensor for underwater environments and offers strong advantages against optical imaging sensors due to its robustness to water turbidity and visible ranges. Target object detection in submerged conditions is optimal with acoustic-based sensing. Nevertheless, optic cameras have been exploited for many underwater applications. Detecting features and changes using optic images remains important, especially for target detection, inspection, and human-robot interaction. Furthermore, higher-resolution images can be exploited in the post-processing phase. For example, feature detection, depth estimation, and visual simultaneous localization and mapping (SLAM) can be performed using optical images.

The main drawback of using optical cameras in underwater occurs by water turbidity and the resulting image degradation. The water medium may deteriorate image quality in terms of detail and color. When using images for navigation, loss of image detail causes more critical issues. This is because the existing underwater visual navigation [1–3] and structure from motion [4] has been widely reported by authors who operated vehicles in clean water. In [5], in-water SLAM performance with deteriora-

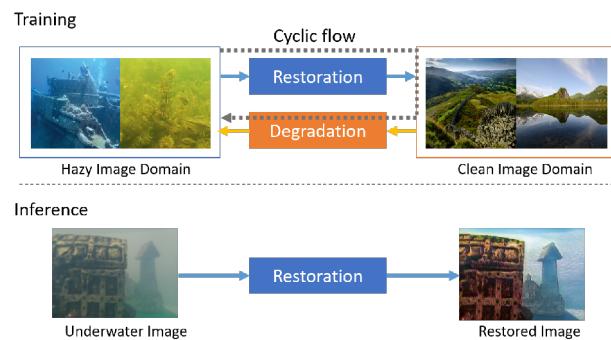


Fig. 1. Flowchart displaying training and inference flows. For training, both restoration and degradation networks are jointly trained through cyclic flow (dotted arrow) across image domains. For inference (testing), restoration network (DehazeGAN) takes underwater images and produces color-corrected and detail-restored output images.

tion of the water turbidity was reported.

The recent advent of deep learning approaches to image enhancement has facilitated substantial improvements in many vision applications. The learning-based approach is more robust to environmental variation and does not re-

Manuscript received August 20, 2019; revised December 3, 2019; accepted December 23, 2019. Recommended by Associate Editor Son-Cheol Yu under the direction of Editor-in-Chief Keum-Shik Hong. The CAMS2019 technical program committee was recommended the authors' CAMS paper (Paper No. 117, "DehazeGAN: Underwater Haze Image Restoration Using Unpaired Image-To-Image Translation") to consider for the IJCAS S.I. MRCs. This research was supported by a grant(20TSRD-B151228-02) from Urban Declining Area Regenerative Capacity-Enhancing Technology Research Program funded by Ministry of Land, Infrastructure and Transport of Korean government. We are also grateful to Australian Centre for Field Robotics (ACFR) for sharing the valuable underwater dataset used in the validation.

Younggun Cho, Hyesu Jang, and Ayoung Kim are with Department of Civil and Environmental Engineering, KAIST, Daejeon, Korea (e-mails: {yg.cho,iriter,ayoungk}@kaist.ac.kr). Ramavtar Malav is with Department of Electrical Engineering, IIT, Kanpur, India (e-mail: rmalav@iitk.ac.in). Gaurav Pandey is with Ford Motor Company, Palo Alto, CA, USA (e-mail: gpandey.iitk@gmail.com). Younggun Cho and Hyesu Jang contributed equally.

* Corresponding author.

quire prior knowledge. However, limitations could be encountered during the training phase when we need to prepare the training data for underwater dehazing. To overcome this challenge, we present a Generative Adversarial Networks (GAN) based image enhancement model for underwater optical imaging.

Our method is focused on model generalities and detail restoration. Using unpaired image-to-image translation, we tried to capture the natural characteristics of underwater images and restored them to produce color-balanced, detailed images. Fig. 1 presents the main flow of training and inference process of our method. We utilized cyclic flow (dotted arrow) when the training, and the trained restoration network was applied to image restoration. In particular, the proposed approach addresses the following attributes.

- Whereas the existing methods simulate the underwater hazy images from clean images for training, ours directly exploit the real underwater images to capture natural characteristics of image domains.
- The proposed training method leverages unpaired image sets without requiring ground truths. It is, therefore, the more appropriate underwater domain where training data acquisition is challenging.
- Our method addresses image detail restoration using multi-objective losses. This detail restoration is particularly important for underwater robotic vision approaches, such as SLAM and structure from motion.

The rest of the paper is organized as follows: Section 2 presents related works of underwater imaging; Section 3 includes a detailed description of our proposed approach; Section 4 provides training details; Section 5 contains our qualitative and quantitative results; Section 6 concludes this paper.

2. RELATED WORKS

Since a model for underwater imaging was introduced by [6], various studies have been conducted to improve underwater imaging. In early works of researches, underwater imaging methods relied on image parameter estimation of the underwater image model.

First of all, image enhancement methods for underwater image models were presented at the beginning of the research, and methods for inferring each image model parameter from single images were proposed. The key parameters of the image model are *transmission* and *ambient light*, which infer light attenuation, and added light that results from the light scattering effect in turbid media. The model-based methods used priors that indicate light transmission according to an image pixel value: well-known studies utilized red-channel attenuation [7] or wavelength compensation for each channel [8]. Berman *et al.* [9] proposed a haze-line-based image restoration method that

used color ratios in a spherical color space. However, these prior-based methods have different assumptions depending on the water-type or scene characteristics.

Unlike the aforementioned prior-based approaches, nonmodel-based approaches also have been studied. Histogram Equalization (HE) and Contrast-limited Adaptive Histogram Equalization (CLAHE) are the most widely used methods for overall image enhancement, including dehazing [10]. For underwater applications, Eustice *et al.* [11] presented using CLAHE for underwater visual SLAM. Ancuti and Ancuti [12] proposed fusion-based dehazing using Laplacian and Gaussian spaces. This method is applicable to various underwater images that use white-balanced inputs. However, distortion may occur depending on the weight in the fusion process. More recently, Cho *et al.* [13] utilized the benefits of model and fusion based methods with multi-band enhancement.

Recently, learning-based methods have been introduced. Shin *et al.* [14] proposed a method that uses a multi-scale Convolutional Neural Network (CNN), that predicts transmission and ambient light terms separately. This method uses synthetic patches for training and estimates patch-wise transmission. WaterGAN [15] used GAN to generate training data and applied CNN for depth estimation image restoration. To generate realistic underwater image models, WaterGAN utilizes GAN with a well-known haze image models, such as attenuation, bias, and vignetting. Unlike WaterGAN, UGAN and UGAN-P, proposed by [16], utilize GAN for image restoration. The authors used CycleGAN [17] to synthesize underwater images from clear images and Wasserstein GAN with Gradient Penalty (WGAN-GP) [18] for image restoration. Another GAN-based approach, Underwater-GAN [19] also utilized WGAN-GP for generator network. They used various underwater turbidity environment filters to generate a training dataset. The aforementioned learning-based methods adopted generative models for training image generation. Restoration networks were trained in a supervised manner and fitted to generated training image pairs.

3. PROPOSED METHOD

In this section, we describe an underwater image restoration model that uses GAN-based translation. GAN is composed of a generator network G and a discriminator network D . During the training process, the discriminator network learns to distinguish whether an input image is a true instance from a real image or a false instance from a generated image. However, the generator network is trained to fool the discriminator. In our case, the generator network is conditioned on an input image and translates the input image to a target domain.

We adopt cyclic architecture [17] which concatenate two GANs for unpaired image translation across two domains A and B. Since the base architecture is well-known

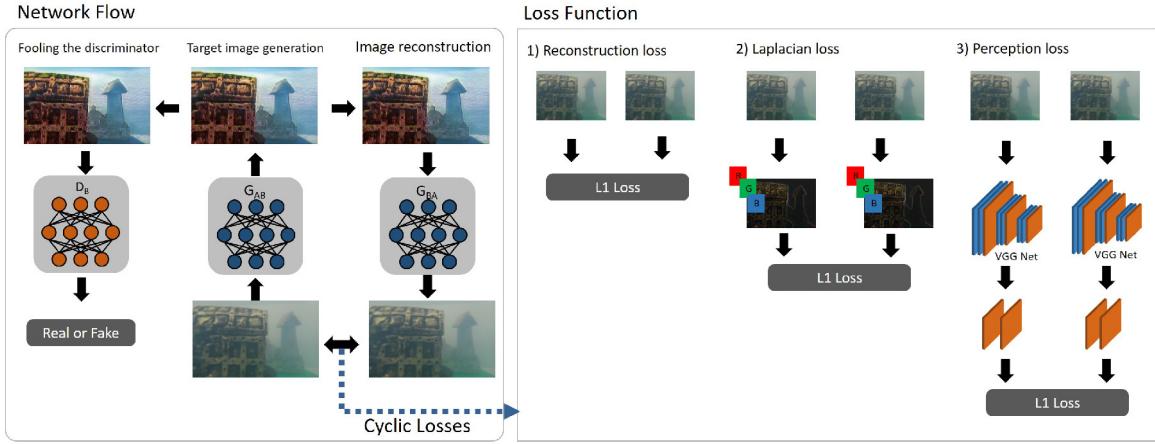


Fig. 2. Proposed networks and loss functions. Flows in gray box (left) represent network flow of cyclic architecture and discriminator. Flows in green box (right) show multiple cyclic consistency losses. Adversarial loss is not included in the figure for clear explanation.

but not performed well for image restoration, we proposed detail and texture-guided training for underwater image enhancement. Generators (G_{AB}, G_{BA}) transform images from the source domain to target domain ($G_{AB}(a) \rightarrow b$) where a and b represent an image of each domain). Discriminators (D_A, D_B) discern whether an input is real or fake from the generator. In this paper, we set domain A as degraded (hazy) images and domain B as clean (aerial) images. The basic notation and flow the networks are described as follows (Fig. 2). The network flow in Fig. 2 represents cyclic process of image domain A (hazy images). Another cyclic process for domain B (clean images) is also applied during the training procedure.

3.1. Loss function

In this section, we summarize the loss functions of the networks. As described in Fig. 2, we constructed multiple losses for the discriminators and generators. Although adversarial losses were applied to discriminators, cyclic losses were utilized for both generators with input and reconstructed images. For simple notation, we define a reconstructed image a_r as $a_r = G_{BA}(G_{AB}(a))$ and b_r as $b_r = G_{AB}(G_{BA}(b))$. The list of losses are provided in Table 1.

3.1.1 Adversarial loss

Adversarial discriminator of each domain is represented as D_A and D_B . The expression of an adversarial loss is

$$\mathcal{L}_{adv} = \mathbb{E}_A[\log D_A(a)] + \mathbb{E}_B[\log(1 - D_B(G_{AB}(b)))] \quad (1)$$

where $D(\cdot)$ aims to distinguish that an input image is from real samples or generated by $G(\cdot)$. For adversarial training, the optimal G and D are acquired by minimizing $\mathcal{L}_{adv}^G = \mathbb{E}_A[\log D_A(a)]$ and maximizing $\mathcal{L}_{adv}^D = \mathbb{E}_B[\log(1 -$

$D_B(G_{AB}(b)))]$. For stable optimization, this problem is converted to minimize both \mathcal{L}_{adv}^G and $-\mathcal{L}_{adv}^D$.

3.1.2 Cyclic consistency losses

For reliable image generation, we apply multiple cyclic consistency losses, namely reconstruction loss [17], Laplacian loss [20], and perceptual loss [21]. We construct the cyclic architecture as $a_r = G_{BA}(G_{AB}(a))$ which represents the real input image translated to the target domain and reconstructed to the original domain. The reconstruction loss is computed by comparing the real input image a and the reconstructed image a_r . The expression of the loss is represented as

$$\mathcal{L}_{rec} = \mathbb{E}_{a,a_r}[||a - a_r||_1]. \quad (2)$$

L_1 norm was used for pixel space distance.

To obtain a detail-enhanced image, we utilized Laplacian loss [20] to the networks. Laplacian loss leads to gradient similarity and prevents missing details in generated images. We model a fixed 3×3 Laplacian filter and apply the filter to each color channel. Unlike [20], we also use average pooling on color channels to capture the detail region of each color channel independently. The Laplacian image is represented as $L(a) = 1/3(L(a^R) + L(a^G) + L(a^B))$. The cyclic Laplacian loss is computed by the L_1 distance of Laplacian images of input and reconstructed images.

$$\mathcal{L}_{Lap} = \mathbb{E}_{a,a_r}[||L(a) - L(a_r)||_1]. \quad (3)$$

The final cyclic loss is perceptual loss [21]. Perceptual loss indicates deep feature consistency between input and reconstructed images. Since pixel similarity and gradient similarity capture average intensity level and gradient of local region, deep feature similarity is applied to learn texture information of local regions. We utilize a pretrained

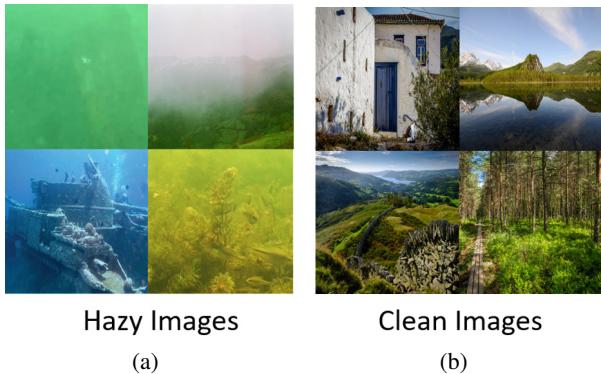


Fig. 3. Sample training images of hazy scene and clear scene.

VGG-16 [22] network as a deep feature extractor. The perceptual loss is computed by L_1 distance of `relu1-2` layer $P(\cdot)$ of VGG-16 and represented as

$$\mathcal{L}_{pert} = \mathbb{E}_{a,a_r}[||P(a) - P(a_r)||_1]. \quad (4)$$

3.1.3 Main objective

The main objective function of discriminator \mathcal{L}_D and generator \mathcal{L}_G is

$$\begin{aligned}\mathcal{L}_D &= -\mathcal{L}_{adv}, \\ \mathcal{L}_G &= \mathcal{L}_{adv} + \lambda_{rec}\mathcal{L}_{rec} + \lambda_{Lap}\mathcal{L}_{Lap} + \lambda_{pert}\mathcal{L}_{pert},\end{aligned}\tag{5}$$

where λ represent the hyper-parameter of each loss. We control λ on every losses except adversarial loss to balance the effects of losses on training. In this paper, we use $\lambda_{rec} = 10$, $\lambda_{Lap} = 20$, and $\lambda_{pert} = 10$.

3.2. Networks

Our generator networks are auto-encoders with residual blocks [23]. Auto-encoders are composed of downsampling (encoding), bottleneck (embedding), and upsampling (decoding) networks. Instead of a pooling layer for downsampling, we used a convolution layer with stride to minimize information loss during downsampling. For bottleneck part, we constructed 9 residual blocks to translate the embedded features of an original domain to a target domain.

As discriminator networks, we utilized PatchGAN [24] for learning stability. PatchGAN classifies true or fake instances of overlapping patches in input images.

4. TRAINING

The proposed network was implemented using PyTorch and trained with an NVIDIA GTX 1080ti. We employed the Adam solver [25] with $\beta_1 = 0.9$, $\beta_2 = 0.99$, and $w_{decay} = 10^{-5}$. We started the training with an initial learning rate of 10^{-4} and controlled it by a step scheduler with a step-size of 20 and $\gamma = 0.5$.

For dataset preparation, we gathered unpaired training images from Flickr. Training images were collected using several keywords, such as underwater haze, distortion and clear scene. To get reliable datasets, we filtered outlier images, which is not related to haze and clean scenes, using previous haze relevant priors and the level of color distortion. Fig. 3 shows sample images of training datasets. Hazy images and clean images are composed of underwater images with various degradations and aerial images without attenuation effects. We utilized 1000 images as training sets.

5. EXPERIMENT RESULTS

In order to evaluate the performance fairly, we verified the results with image enhancement methods based on various approaches. We compared our approach to histogram equalization, CLAHE, fusion-based method [26] & [12], and model-based method [27] & [9]. During the evaluation, we selected Underwater CNN [14] and WaterGAN [15] to analyze the performance and effectiveness of the learning-based method. Except Ancuti's method [12], we evaluated compared methods using the code provided by the original authors. For general evaluation, we constructed test sets from various datasets with different water types. We tested our method on 144 images from ACFR, Turbid [28], WaterGAN [15], and Crabster datasets¹.

5.1. Image restoration

We first evaluated the quality of image restoration. Fig. 4 and Fig. 5 depict qualitative results on simulated underwater haze images and real underwater hazy scenes. For both environments, our method revealed a promising performance against others. Overall color balance and the degree of dehazing are adequate for all dataset. Since histogram equalization stretches the color histogram to achieve maximum contrast, it shows good performance on the uniform hazy scene without color bias and illumination distortion. However, the real underwater environment is often more challenging with heavier color distortion.

5.1.1 Restoration quality

Fig. 4 represent (a) raw images without haze, (b) hazed images, and (c-l) enhanced images of compared and proposed methods. The first two rows in Fig. 4 show selected images from ACFR and bottom two rows represent test images from the TURBID dataset [28]. We simulated underwater haze condition with illumination bias for ACFR image datasets. TURBID dataset offers clear scene as ground-truth and hazy scenes.

Most of image restoration methods show limitation in haze with illumination bias except CLAHE

¹<https://www.youtube.com/watch?v=M7KR-AL1Ab0>

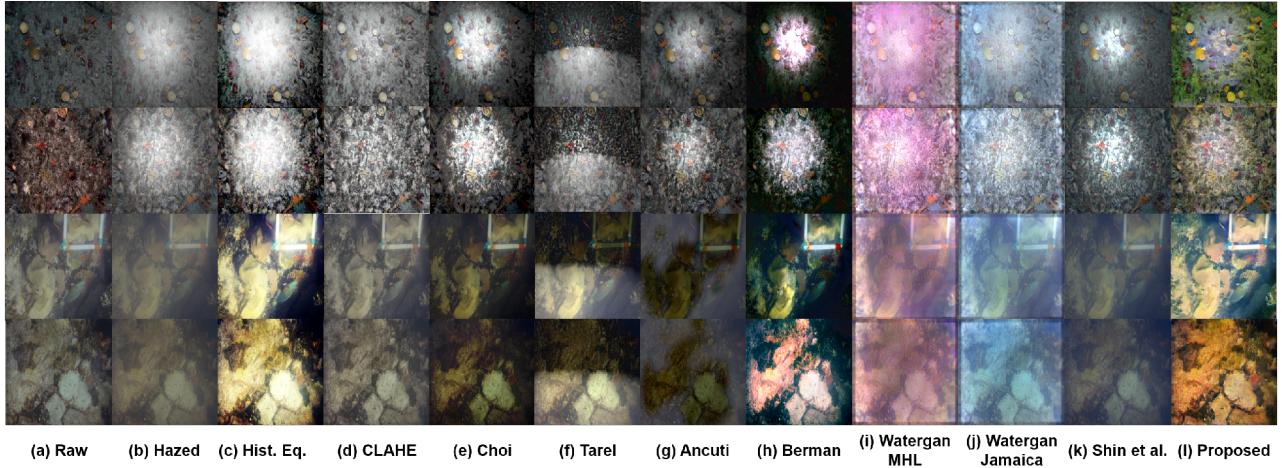


Fig. 4. Image results of simulated environment dataset. Two images from top are ACFR datasets, and the others are Turbid dataset. (a) the original clear underwater images. (b) the artificially hazed images.

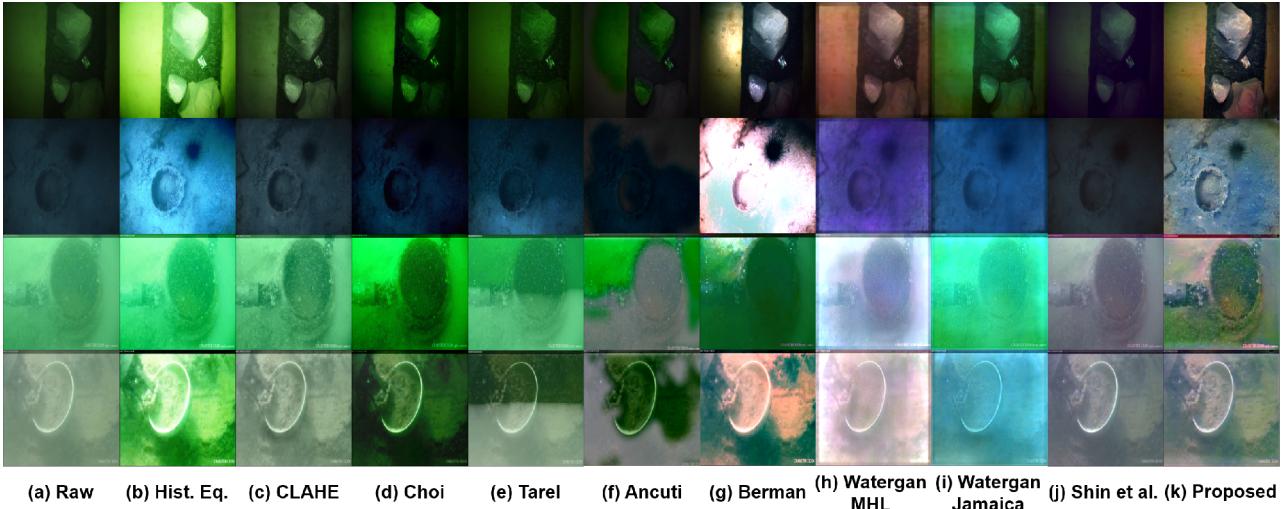


Fig. 5. Image results of real underwater environment dataset. Since real environment has no clear image, improved images are compared to raw underwater images in column (a). From the top to bottom, datasets are WaterGAN-MHL, WaterGAN-Jamaica, Crabster-Ironpot, and Crabster-Pottery.

(Contrast-limited Histogram Equalization) and the proposed method. In details, images from histogram equalization (c) and CLAHE (d) shows enhanced contrast compared to other enhancements method. However, histogram equalization has weakness under illumination bias as represented in ACFR datasets. The result images from CLAHE shows the performance dependency of parameter selection (clip-limit). While Berman's method [9] shows good result on TURBID dataset (g), however, this method also needs uniform illumination conditions. WaterGAN [15] present interesting results with different training datasets. We validated WaterGAN trained by MHL (h) or Jamaica (i) datasets. Since test images have different color characteristics, the output images have different color distortion. Shin's method (j) [14] cannot clearly improve image quality. However, the proposed method

(k) possess both image quality and light condition, color restoration also conducted vividly. The proposed method shows well-enhanced images on hazy scenes with extra degradations.

As described in Fig. 5, we also evaluated image enhancement methods on real underwater datasets. The subscription of each column represents the evaluated methods. As a result, most of the methods undergo difficulties in color restoration, while the proposed method enhanced the haze image effectively. The proposed method shows color-balanced with better visibilities on various underwater images. All results of the proposed methods are acquired by the same networks with the same weights.

For other datasets, we can find different characteristics of image enhancement according to the methodologies. Histogram-based enhancements enhance not only the

Table 1. Quantitative indices and results for simulated and real underwater environment dataset processing. SSIM and PSNR regard color intensity and distortion level of the image, CNR stands for contrast to noise ratio, and FADE measures the haze level. We could discover proposed method has high level of SSIM and PSNR CNR, and low FADE value compared to other methods. For each metric, \uparrow indicates that the higher, the better, and \downarrow indicates the lower, the better. Also, the score with bold characters represent the best score of each metric. We do not mark the best score for UIConM because all methods show similar results.

	Metric	Hist. Eq.	CLAHE	Choi	Tarel	Ancuti	Berman	WaterGAN-MHL	WaterGAN-Jamaica	Shin	Proposed
Simulated Env.	SSIM \uparrow	0.54	0.71	0.61	0.66	0.34	0.37	0.31	0.26	0.56	0.46
	PSNR \uparrow	15.11	23.86	16.36	13.36	14.38	11.88	12.07	13.94	21.94	16.61
	CNR \uparrow	217.89	269.66	211.65	179.95	216.40	141.20	215.46	218.45	155.64	210.87
	FADE \downarrow	0.54	1.05	0.65	0.59	0.66	0.53	1.06	0.97	1.28	0.47
Real Underwater Env.	CNR \uparrow	101.93	104.49	55.17	57.62	48.91	68.85	76.81	67.07	50.98	92.84
	FADE \downarrow	0.67	0.99	0.61	0.69	0.90	0.61	1.37	0.75	1.88	0.76
	UICM \uparrow	2.26	0.65	3.74	1.46	3.71	4.89	1.96	1.51	1.44	4.57
	UISM \uparrow	2.57	3.15	2.58	2.66	2.59	2.71	3.12	2.99	2.37	3.44
	UIConM \uparrow	-0.06	-0.07	-0.07	-0.07	-0.07	-0.02	-0.05	-0.06	-0.04	-0.06
	UIQM \uparrow	0.93	1.10	1.07	0.94	1.04	1.04	1.14	1.09	0.86	1.36

contrast but also the color distortion. The fusion-based method makes distortion depending on the scale in the fusion process. Berman’s method is the representative model-based method using haze-line image priors. Since hazy-line assumes that underwater images have enough parallax foreground and background region to retrieve haze-line, the method fails to estimate transmission on planar-like environments. Also, deep learning methods show the dependency on training datasets. WaterGAN utilizes MHL and Jamaica datasets independently on network training for depth estimation and color estimation. Trained weights in a limited and specific dataset cause an overfitting problem, and the restored color is confined to the training color sets. As represented in Fig. 4(i), MHL-weighted WaterGAN and Jamaica-weighted WaterGAN different color tendency on result images. Results from Shin’s method represent high-quality dehazing; however, they cannot afford to control the light condition.

5.1.2 Evaluation metric

We assigned four evaluation indicators for quantitative assessment. SSIM calculates the structural similarity between the original image and the distorted image. Higher SSIM value implies that a target image has a more similar color intensity to the original image. Since ACFR and Turbid datasets have a non-hazed true image, we calculated SSIM only for two simulation datasets. The second index is PSNR, which calculates the ratio of peak signal and noise. The third is CNR, calculating the contrast to noise ratio in the image. Since contrast in image influences object visibility, CNR is an important criterion for object recognition. The last index is FADE that measures the haze level of the image. Lower FADE value signifies a lower haze level.

Table 1 describes the value of each index. As mentioned, SSIM evaluates the color intensity and degree

of distortion. Our SSIM value is not the highest. However, we received a reliable score in overall evaluation indices. PSNR also shows an analogous tendency with SSIM. Since the simulated environment has less noise effect than the real underwater environment, CNR values in the simulated environment are mainly high. However, most of the real underwater environment has low perception quality. We verified that the proposed method is robust in all environments by comparing CNR value. We obtained the lowest FADE value among the methods except for the Crabster dataset. The periphery of the object is important for FADE index. However, the Crabster dataset is not adapted to FADE algorithm. Although containing the Crabster dataset, our average value of FADE is low. We affirm our dehazing method is robust in every underwater environment, and color restoration is more accurate in real underwater environment data. The major quality metric for underwater image measure is Underwater Image Quality Measure (UIQM) [29]. UIQM represents the image quality with respect to UICM (Colorfulness), UISM (Sharpness), and UIConM (Contrast). Since UIQM includes measure for color undistortion, our method shows the best score.

5.2. Robotic vision application

For many underwater robot applications, robust vision algorithms are needed to retrieve reliable motion and object information. In this section, we evaluated the quality of restoration through verifying meaningful edges and feature matching results. We utilized image gradient metric introduced by [30] which is robust gradient metric against image noise. Corresponding gradient images are described in Fig. 6. While histogram equalization and our method show similar and more detailed edge information compared to the original image, Berman’s method [9] results in over and undersaturated pixels due to misestimation of

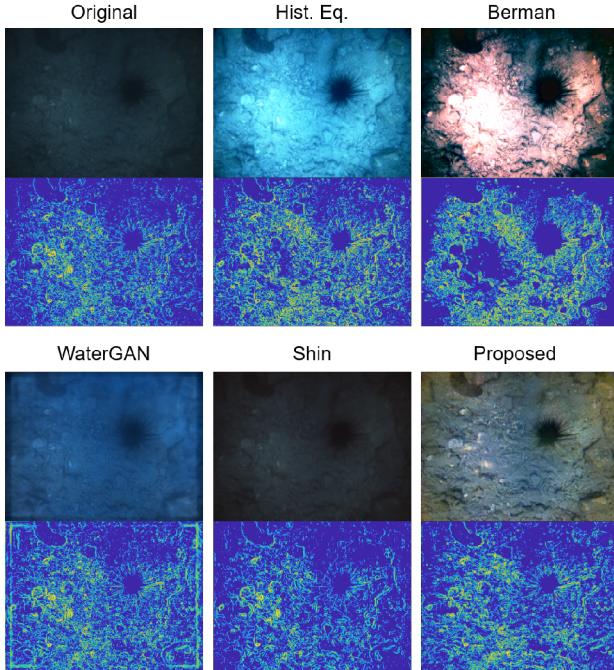


Fig. 6. Edge (gradient) evaluation. For each method, the first row is input of gradient mapping and the second row shows a corresponding edge map.

Table 2. Gradient level metric on multiple sequences. The output value of each method is the mean value of tested images. The higher level indicates the better performance.

	Original	Hist. Eq.	Berman	Shin	Water GAN	Proposed
Level	0.12	0.17	0.17	0.11	0.16	0.18

pixel transmission. WaterGAN [15] output has artifacts on the image boundary. We also verified gradient metrics over 30 sequences, the average level of each method is shown in Table 2. The higher value indicates the better performance. For motion-related evaluation, we compared the image feature quantity and quality by measuring the number of features and inliers. This evaluation represents the usefulness of image enhancement method on real robot vision applications. We chose Speeded Up Robust Features (SURF) [31] detector and descriptor.

Fig. 7 presents sample feature matching and image warping results. Since most of the methods introduced in the earlier section failed feature matching, we selected other dehazing methods from fusion-based approach [26] (Choi) and model-based approach [27] (Tarel). Both methods focus on improving dehazing quality and visible edge enhancement. For underwater images, nevertheless, Choi's and Tarel's method failed image warping due to lack of inliers. The detailed evaluation of feature matching is introduced via Table 3. We selected the image sequence

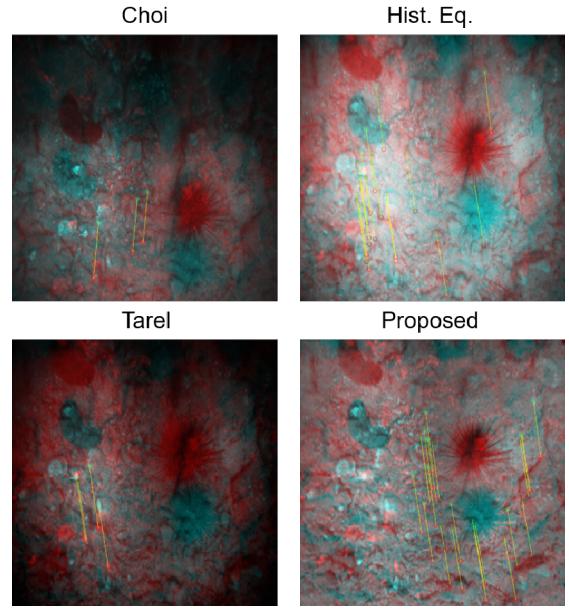


Fig. 7. Feature matching samples. Yellow lines indicate feature correspondences. Keypoints of sequential frames are represented as circle (o) and cross (+) mark.

Table 3. Feature matching results. We compared the number of features and the number of inliers. The higher level indicates the better performance.

	Hist. Eq.	Tarel	Choi	WaterGAN	Proposed
Num. Feat.	333.46	12.333	13.083	4.5417	602.17
Num. Inliers	16.5	1.3333	1.8333	0.25	17.25

with 50 images from the WaterGAN-Jamaica dataset. The higher value indicates a better quality of enhancement.

5.3. Ablation study

As in Fig. 8, we verified the effects of multiple cyclic losses. We evaluated CycleGAN, DehazeGAN with Laplacian loss, DehazeGAN with Perceptual loss, and DehazeGAN with overall losses. Since CycleGAN has pixel reconstruction loss which is common for the all setups, performance improvement is mainly discussed for multiple losses. The enlarged views in the second row imply the quality of image enhancement. When an objective function has only pixel reconstruction loss, image detail can be distorted. By adding additional constraints on the objective function, we can get a better restoration result. DehazeGAN with full losses shows sharper and well-textured output.

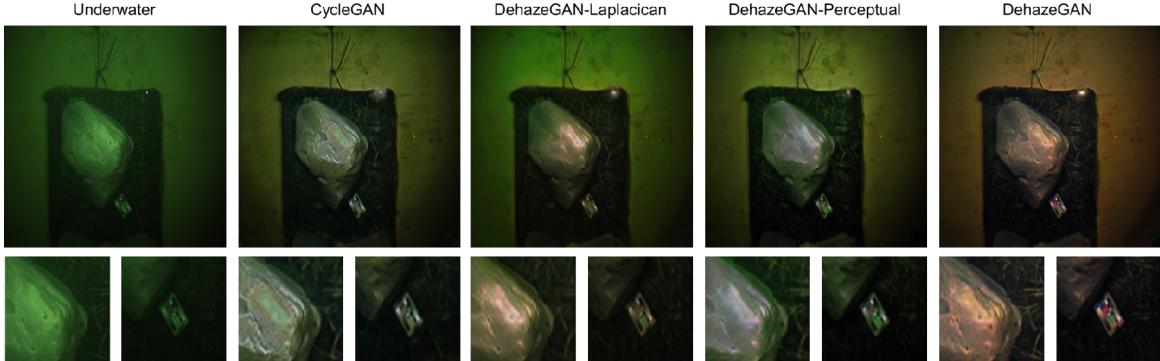


Fig. 8. Ablation study on the multiple cyclic losses. Compared to CycleGAN which only has pixel reconstruction error, we tested Laplacian, perceptual losses and the final mixture loss.

6. CONCLUSION

In this paper, we propose an underwater image restoration method using deep generative models. We utilized cyclic architecture and constructed multiple consistency losses for detail enhanced images. For evaluation, we compared against other state-of-arts underwater methods over various underwater images, reporting both qualitative and quantitative results. Because our focus is also at the underwater robotic vision by validating the proposed method in terms of the feature matching performance.

REFERENCES

- [1] F. Hover, R. Eustice, A. Kim, B. Englot, H. Johannsson, M. Kaess, and J. Leonard, “Advanced perception, navigation and planning for autonomous in-water ship hull inspection,” *International Journal of Robotics Research*, vol. 31, pp. 1445-1464, Oct. 2012.
- [2] P. Ridao, M. Carreras, D. Romagós, and R. Garcia, “Visual inspection of hydroelectric dams using an autonomous underwater vehicle,” *Journal of Field Robotics*, vol. 27, pp. 759-778, Nov. 2010.
- [3] T. Kim, J. Kim, and S.-W. Byun, “A comparison of nonlinear filter algorithms for terrain-referenced underwater navigation,” *International Journal of Control, Automation and Systems*, vol. 16, no. 6, pp. 2977-2989, 2018.
- [4] E. Galceran, R. Campos, N. Palomeras, D. Romagós, M. Carreras, and P. Ridao, “Coverage path planning with real-time replanning and surface reconstruction for inspection of three-dimensional underwater structures using autonomous underwater vehicles,” *Journal of Field Robotics*, vol. 32, no. 7, pp. 952-983, 2014.
- [5] Y. Cho and A. Kim, “Visibility enhancement for underwater visual slam based on underwater light scattering model,” *Proceedings of the IEEE International Conference on Robotics and Automation*, IEEE, 2017.
- [6] J. S. Jaffe, “Computer modeling and the design of optimal underwater imaging systems,” *IEEE Journal of Oceanic Engineering*, vol. 15, no. 2, pp. 101-111, 1990.
- [7] N. Carlevaris-Bianco, A. Mohan, and R. M. Eustice, “Initial results in underwater single image dehazing,” *Proceedings of the IEEE/MTS OCEANS Conference and Exhibition*, Seattle, pp. 1-8, IEEE, 2010.
- [8] J. Y. Chiang and Y. Chen, “Underwater image enhancement by wavelength compensation and dehazing,” *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1756-1769, 2012.
- [9] D. Berman, T. Treibitz, and S. Avidan, “Diving into haze-lines: Color restoration of underwater images,” *Proc. of the British Machine Vision Conference (BMVC)*, vol. 1, 2017.
- [10] K. Zuiderveld, “Contrast limited adaptive histogram equalization,” *Graphics Gems IV*, pp. 474-485, Academic Press Professional, Inc., 1994.
- [11] R. Eustice, O. Pizarro, H. Singh, and J. Howland, “UWIT: Underwater image toolbox for optical image processing and mosaicking in matlab,” *Proceedings of the International Symposium on Underwater Technology*, pp. 141-145, IEEE, 2002.
- [12] C. O. Ancuti and C. Ancuti, “Single image dehazing by multi-scale fusion,” *IEEE Transactions on Image Processing*, vol. 22, no. 8, pp. 3271-3282, 2013.
- [13] Y. Cho, J. Jeong, and A. Kim, “Model assisted multi-band fusion for single image enhancement and applications to robot vision,” *IEEE Robotics and Automation Letters (RA-L) (with IROS)*, vol. 3, no. 4, pp. 2822-2829, 2018.
- [14] Y.-S. Shin, Y. Cho, G. Pandey, and A. Kim, “Estimation of ambient light and transmission map with common convolutional architecture,” *Proceedings of the IEEE/MTS OCEANS Conference and Exhibition*, (Monterey, CA), pp. 1-7, Sep. 2016.
- [15] J. Li, K. A. Skinner, R. M. Eustice, and M. Johnson-Roberson, “Watergan: unsupervised generative network to enable real-time color correction of monocular underwater images,” *IEEE Robotics and Automation Letters*, vol. 3, no. 1, pp. 387-394, 2017.
- [16] C. Fabbri, M. J. Islam, and J. Sattar, “Enhancing underwater imagery using generative adversarial networks,” *Proc. of IEEE International Conference on Robotics and Automation (ICRA)*, pp. 7159-7165, IEEE, 2018.

- [17] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” *Proc. of the IEEE International Conference on Computer Vision*, pp. 2223-2232, 2017.
- [18] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, “Improved training of wasserstein gans,” *Advances in Neural Information Processing Systems*, pp. 5767-5777, 2017.
- [19] X. Yu, Y. Qu, and M. Hong, “Underwater-gan: Underwater image restoration via conditional generative adversarial network,” *Proc. of International Conference on Pattern Recognition*, pp. 66-75, Springer, 2018.
- [20] S. Li, X. Xu, L. Nie, and T.-S. Chua, “Laplacian-steered neural style transfer,” *Proceedings of the 25th ACM international conference on Multimedia*, pp. 1716-1724, ACM, 2017.
- [21] J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” *Proc. of European Conference on Computer Vision*, pp. 694-711, Springer, 2016.
- [22] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
- [23] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778, 2016.
- [24] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1125-1134, 2017.
- [25] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [26] L. K. Choi, J. You, and A. C. Bovik, “Referenceless prediction of perceptual fog density and perceptual image defogging,” *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3888-3901, 2015.
- [27] D. A. Nicolas Hautiere, Jean-Philippe Tarel and E. Du-mont, “Blind contrast enhancement assessment by gradient ratioing at visible edges,” *Image Analysis and Stereology*, vol. 27, no. 2, pp. 87-95, 2011.
- [28] A. Duarte, F. Codevilla, J. D. O. Gaya, and S. S. Botelho, “A dataset to evaluate underwater image restoration meth-ods,” *Proceedings of the IEEE/MTS OCEANS Conference and Exhibition*, Shanghai, pp. 1-6, IEEE, 2016.
- [29] K. Panetta, C. Gao, and S. Agaian, “Human-visual-system-inspired underwater image quality measures,” *IEEE Journal of Oceanic Engineering*, vol. 41, no. 3, pp. 541-551, 2015.
- [30] I. Shim, J. Lee, and I. S. Kweon, “Auto-adjusting camera exposure for outdoor robotics using gradient information,” *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1011-1017, IEEE, 2014.
- [31] H. Bay, T. Tuytelaars, and L. Van Gool, “Surf: Speeded up robust features,” *Proc. of European Conference on Computer Vision*, pp. 404-417, Springer, 2006.



Younggun Cho received his B.S. degrees in Electrical Engineering from Inha University, Incheon, Korea, in 2013, and an M.S. degree in Electrical Engineering and a Ph.D. degree in Civil and Environmental Engineering with dual degree in robotics program from Korea Advanced Institute of Science & Technology (KAIST), Daejeon, Korea, in 2015 and 2020, respectively. Currently, he is the Postdoctoral researcher in Intelligent Robotic Autonomy and Perception (IRAP) Lab. His research interests include robust sensing in turbid environments, and learning-based simultaneous localization and mapping, navigation.



Hyesu Jang received his B.S. degree in civil & environmental engineering from Korea Advanced Institute of Science & Technology (KAIST), Daejeon, Korea, in 2018. Currently, he is an M.S. student in the department of civil & environmental engineering, KAIST. His research interests include marine robotics, computer vision, and simultaneous localization and mapping.



Ramavtar Malav completed his Bachelor’s and Master’s degrees in 2017 from IIT Kanpur Electrical Department under supervision of Dr. S. R. Sahoo and Dr. G. Pandey. He is currently working as a Research and Development Engineer at IPsoft Inc. India. He has been previously part of IRAP Lab, KAIST Daejeon as a summer intern. His research interests include robotics computer vision and machine learning.



Gaurav Pandey is a Technical Expert in the Controls and Automated systems department of Ford Motor Company, and is working on developing localization algorithm for highly automated vehicles. He is also working on developing robust robotic solutions for the package delivery problem. Prior to Ford, Dr. Pandey was an Assistant Professor at the Electrical engineering department of Indian Institute of Technology (IIT) Kanpur in India. At IIT Kanpur, he was part of two research groups in the electrical engineering department: Control and Automation & Signal Processing and Communication. His current research focus is on visual perception for autonomous vehicles and mobile robots using tools from computer vision, machine learning and information theory. He received his B-Tech from IIT Roorkee in 2006 and his Ph.D. from University of Michigan, Ann Arbor in December 2013.



Ayoung Kim received her B.S. and M.S. degrees in mechanical engineering from Seoul National University, Seoul, Korea, in 2005 and 2007, respectively, and an M.S. degree in electrical engineering and a Ph.D. degree in mechanical engineering from the University of Michigan (UM), Ann Arbor, in 2011 and 2012, respectively.

Currently, she is an associate professor in the department of civil and environmental engineering with joint affiliation at KI robotics, KAIST. Her research interests include visual simultaneous localization and mapping, navigation.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.