Emerging From Water: Underwater Image Color Correction Based on Weakly Supervised Color Transfer

Chongyi Li, *Student Member, IEEE*, Jichang Guo, and Chunle Guo

***Abstract*—Underwater vision suffers from severe effects due to selective attenuation and scattering when light propagates through water. Such degradation not only affects the quality of underwater images, but limits the ability of vision tasks. Different from existing methods that either ignore the wavelength dependence on the attenuation or assume a specific spectral profile, we tackle color distortion problem of underwater images from a new view. In this letter, we propose a weakly supervised color transfer method to correct color distortion. The proposed method relaxes the need for paired underwater images for training and allows the underwater images being taken in unknown locations. Inspired by cycle-consistent adversarial networks, we design a multiterm loss function including adversarial loss, cycle consistency loss, and structural similarity index measure loss, which makes the content and structure of the outputs same as the inputs, meanwhile the color is similar to the images that were taken without the water. Experiments on underwater images captured under diverse scenes show that our method produces visually pleasing results, even outperforms the state-of-the-art methods. Besides, our method can improve the performance of vision tasks.**

***Index Terms*—Color casts, color correction, color transfer, Generative Adversarial Networks (GANs), underwater image.**

I. Introduction

RECENTLY, ocean engineering and research have increasingly relied on underwater images captured by autonomous underwater vehicles and remotely operated vehicles [1]. However, underwater images usually suffer from degeneration, such as low contrast, color casts, and noise, due to wavelength-dependent light absorption and scattering as well as the effects of low-end optical imaging devices. The scattering and absorption attenuate the direct transmission and introduce surrounding scattered light. The attenuated direct transmission leads the intensity from the scene to be weaker and introduces color casts, while the surrounding scattered light causes the appearance of the scene to be washed out. Besides, the magnitude of attenuation and scattering depends on several complex factors including water temperature and salinity, and the type and quantity of particulates in the water. Serious degeneration makes it difficult to recover the appearance and color of underwater images. However, color is extremely important for underwater vision tasks and research [2]. Therefore, how to effectively approximate the real color of underwater images has become a challenging problem.

A number of methods [3]–[16] have been proposed to improve the visual quality of underwater images, ranging from hardware solutions to image dehazing and color correction methods [17], [18]. The hardware solutions [3]–[6] have shown the effectiveness, but these solutions are not applicable to dynamic acquisition. Most of single underwater image restoration methods [7]–[13] are inspired by the outdoor dehazing strategies [19]–[22]. For underwater image color correction, traditional color constancy methods (e.g., Gray World [23], Max RGB [24], White Balance [25], Shades-of-Grey [26], etc.) and their variations are usually employed. Compared with traditional methods based on statistical priors, our method bridges the gap between the color of underwater images and that of air images by learning their cross domain relations. Recently, a semisupervised learning model for underwater image color correction, named WaterGAN, has been proposed [14]. Unlike this work, our model does not need a large annotated dataset of image pairs. Besides, WaterGAN shows limitation when it is used to process the underwater images captured under unknown sites. Contrast to the image style translation methods [27], [28], our model learns the semantic color of air images and preserves the key attributes such as content and structure of the input underwater images.

In this letter, we present a novel weakly supervised model for underwater image color correction, which maps the color of underwater scenes to the color of air scenes without any explicit labels. Specifically, given an underwater image as the input, our model directly outputs an image, which has the content and structure the same as the input, but the color as if the image was taken without the water. Although the translated color might not be the "real" color of underwater images, our method removes the color casts and improves the performance of vision tasks. In fact, it is impossible that a method recovers the appearance and color of any underwater images when the scenes and light conditions are unknown. Fig. 1 shows two examples of our results.

Contributions: This letter introduces the following main contributions.

1) Compared with the existing priors/assumptions based and semisupervised methods, to our best knowledge, this is the first attempt to build a weakly supervised model for

Manuscript received September 28, 2017; revised December 29, 2017; accepted January 8, 2018. Date of publication January 11, 2018; date of current version January 26, 2018. This work was supported in part by the National Key Basic Research Program of China under Grant 2014CB340403 and in part by the National Natural Science Foundation of China under Grant 61771334. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Daniel P. K. Lun. *(Corresponding author: Chunle Guo.)*

The authors are with the School of Electrical and Information Engineering, Tianjin University, Tianjin 300011, China (e-mail: lichongyi@tju.edu.cn; jcguo@tju.edu.cn; guochunle@tju.edu.cn).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2018.2792050

Fig. 1. Two examples of our results. (a) and (c) Raw underwater images. (b) and (d) Our results.

underwater image color correction. Here, "weak supervision" means that our model relaxes the need for paired underwater images for training and allows the underwater images being taken in unknown locations.

2) We tackle underwater image color correction problem from a new angle, which learns a cross domain mapping function between underwater images and air images.
3) A multiterm loss function allows our model capturing context and semantic information, which makes the content and structure of the outputs same as the inputs, meanwhile the color is similar to the images that were taken without the water.

II. Proposed Method

Our solution is based upon very recent advances in image-to-image translation networks [29]–[31], which captures special characteristics of one image collection and figures out how these characteristics could be translated into the other image collection, all in the absence of any paired training examples. Our goal is to learn mapping functions between a source domain X (i.e., underwater) and a target domain Y (i.e., air). The inputs are unpaired training image samples $x \in X$ and $y \in Y$. Specifically, our model includes two mappings $G : X \rightarrow Y$ (forward) and $F : Y \rightarrow X$ (backward) given training data $\{x_i\}_{i=1}^{N} \in X$ and $\{y_i\}_{i=1}^{N} \in Y$. Following the CycleGAN [29], we also use two adversarial discriminators D_X and D_Y. The D_X aims to distinguish the images $\{x\}$ from the translated images $\{F(y)\}$, when the D_Y aims to distinguish the images $\{y\}$ from the translated images $\{G(x)\}$. For the generator of forward networks, the loss function includes three terms: first, adversarial loss is to match the distribution of generated images with the distribution in the target domain; second, cycle consistency loss is to prevent the learned mappings G and F from contradicting each other; third, structural similarity index measure (SSIM) loss [32] is to preserve the content and structure of source images. For the generator of backward networks, the loss function also includes three terms. The framework of the proposed method is shown in Fig. 2. The losses of forward networks and vice versa for backward networks are as follows.

A. Adversarial Loss

For the mapping function $G : X \rightarrow Y$ and its discriminator D_Y, the adversarial loss can be expressed as follows:

$$L_{\text{GAN}}(G, D_Y, X, Y) = E_{y \sim p_{\text{data}}(y)}[\log D_Y(y)] + E_{x \sim p_{\text{data}}(x)}[\log(1 - D_Y(G(x)))] \quad (1)$$

where G tries to generate images $G(x)$ that look similar to images from domain Y, while D_Y aims to distinguish the translated samples $G(x)$ from the real samples y.

B. Cycle Consistency Loss

We add a cycle consistency loss to constrain the space of possible mapping functions. Cycle consistency means that the image translation cycle should be able to bring x from domain X back to the original image and bring y from domain Y back to the target image. Cycle consistency loss can be expressed as follows:

$$L_{\text{cyc}}(G, F) = E_{x \sim p_{\text{data}}(x)}[\|F(G(x)) - x\|_1] + E_{y \sim p_{\text{data}}(y)}[\|G(F(y)) - y\|_1]. \quad (2)$$

C. SSIM Loss

To preserve the content and structure of the input image, SSIM loss [33] is used. First, for pixel p, the SSIM between input image x and translated image $G(x)$ is defined as follows:

$$\text{SSIM}(p) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \cdot \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (3)$$

where p is the center pixel of an image patch, x is an image patch $\in X$ with size 13×13, y is an image patch $\in G(x)$ with size 13×13, μ_x is the mean of x, σ_x is the standard deviations of x, μ_y is the mean of y, σ_y is the standard deviations of y, σ_{xy} is the covariance of x and y. $C_1 = 0.02$ and $C_2 = 0.03$ are default in SSIM loss. In this way, the SSIM value for every pixel between the input image x and the translated image $G(x)$ is calculated. The SSIM loss can be expressed as follows:

$$L_{\text{SSIM}}(x, G(x)) = 1 - \frac{1}{N}\Sigma_{p=1}^{N}(\text{SSIM}(p)) \quad (4)$$

where N is the number of pixel in an image.

D. Total Loss

Finally, the total loss is the linear combination of the above-mentioned three losses with weights as follows:

$$L_{\text{loss}} = \lambda_1 L_{\text{GAN}}(G, D_Y, X, Y) + \lambda_2 L_{\text{cyc}}(G, F) + \lambda_3 L_{\text{SSIM}}(x, G(x)). \quad (5)$$

where weights λ_1, λ_2, and λ_3 are 1, 1, and 10, respectively, based on heuristic experiments on our training data, which makes the order of magnitude of three components equal because we expect that these three components have equally significant contributions to the final loss function. The optimization of generator G of forward networks is to minimize (5) while the optimization of discriminator D_Y is to maximize (1). In the training process, we alternatively optimize G, D_Y, F, and D_x, respectively.

E. Network Architecture and Training Details

The forward and backward networks have the same architecture. Following CycleGAN [29], we adapt the architecture [34] as our generators and use 70×70 PatchGANs [35]–[37] as our discriminators. To train our model, we collect a dataset from the Internet, which contains 3800 underwater images and 3800 air images. After that, those images are resized to 256×256 based on our limited memory and the architecture of our discriminators. Fig. 3 shows several training samples.

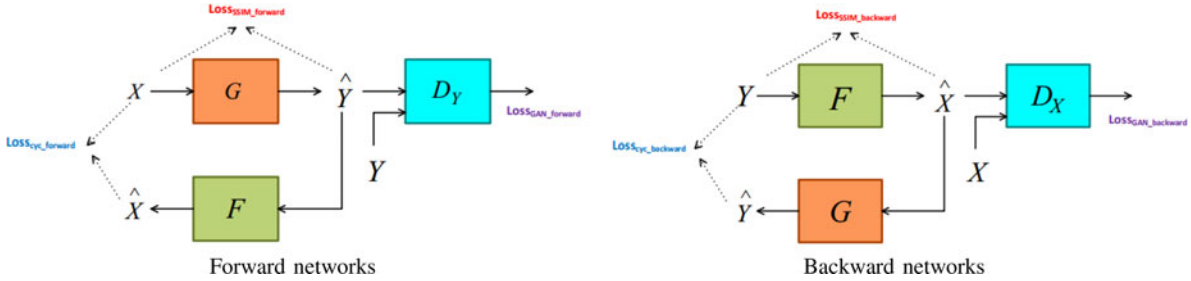

Fig. 2. Diagram of our framework. Our model contains two mapping functions $G : X \rightarrow Y$ (forward) and $F : Y \rightarrow X$ (backward), and associated adversarial discriminators D_Y and D_X.

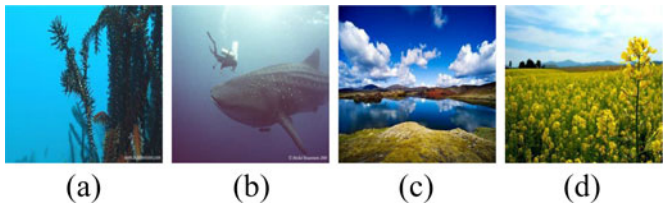

Fig. 3. Our training samples. (a) and (b) Underwater images. (c) and (d) Air images.

In addition, to stabilize our model training, we replace the negative log likelihood objective by a least square loss. Equation (1) is rewritten as follows:

$$L_{\text{GAN}}(G, D_Y, X, Y) = E_{y \sim p_{\text{data}}(y)}[(D_Y(y) - 1)^2] + E_{x \sim p_{\text{data}}(x)}[D_Y(G(x))^2]. \quad (6)$$

The discriminators D_X and D_Y were updated by previous generated images rather than the ones produced by the latest generative networks. We trained our model using ADAM [38] and set the learning rate to 0.0002 and momentum to 0.5. The batch size was set to 1. We implemented our network with the TensorFlow framework and trained it using NVIDIA TITAN X GPU. It took 15 hours to optimize our model.

III. Experiments

We compared the proposed model with several state-of-the-art methods: image-to-image transfer method (i.e., CycleGAN [29]), color constancy method (i.e., Gray Word (GW) [23]), image enhancement method (i.e., INT [39]), and underwater image restoration methods (i.e., RED [9], UWID [11], and UWIB [12]). In our experiments, the raw underwater images were captured under varying underwater scenes. For fair comparisons, we did not compare our method with the WaterGAN since it was just available for underwater images taken under designated sites. We subjectively evaluate the visual quality of the results of different methods. Then, we conduct a user study since there is no metric designed for underwater image color correction. Finally, we carry out the application test on visual tasks in order to give additional evidence.

A. Subjective Assessment

In Fig. 4, we show several results of different methods. In Fig. 4, original CycleGAN produces blurring results because it tends to translate the content and structure of air images to the underwater images. GW method introduces color casts for some images when its assumption is not available. INT method just increases the brightness of underwater images because the prior learned from natural scenes is unavailable. Three underwater images restoration methods remove the effect of haze and improve the contrast; however, the color is not well restored since the assumed optical parameters do not hold in some underwater scenes. On the contrary, our method totally removes the greenish and bluish tone as if our results were taken without the water, which leads to better visual quality.

B. User Study

To perform a visual comparison in an objective way, we randomly selected 30 underwater images from our collected dataset for user study. The results of different methods were randomly displayed on the screen and compared with the corresponding raw underwater images. After that, we invited 10 participants who had experience with image processing to score results. There was no time limitation for each participant. Moreover, the participants did not know which results were produced by our method. The scores ranged from 1 (worst) to 8 (best). As baseline, we set the scores of raw underwater images to 3. We expected the good result has high contrast and visibility, abundant details, especially the color as if the image was taken without the water. On the contrary, the bad result has low visibility, overenhanced regions, serious artifacts and noise, and inauthentic color. We first present in Table I the average visual quality scores of the images shown in Fig. 4.

In Table I, our results receive the best scores, which indicates that, from a visual perspective, our method produces much better results. The underwater image restoration methods also achieve good scores. The average visual quality scores of the selected 30 underwater images are 3, 2, 3, 3.3, 3.7, 4.6, 4, and 6.3 for raws, CycleGAN, GW, INT, RED, UWID, UWBI, and ours, respectively. More results and scores are available at https://li-chongyi.github.io/homepage.github.io/proj_Emerging_water.html. Those results provide a realistic feedback that our method generates visually pleasing results.

C. Application Test

To further demonstrate the effectiveness of the proposed method, we carry out the application test including saliency detection [40] and keypoint matching [41]. In Figs. 5 and 6, we present the results of the application test. For the limited space, we just present several examples of the application test. The proposed method unveils more details and corrects the color of the raw underwater images. Moreover, our results achieve better saliency detection performance and more matching points. Application test gives additional evidence of the effectiveness of our method.

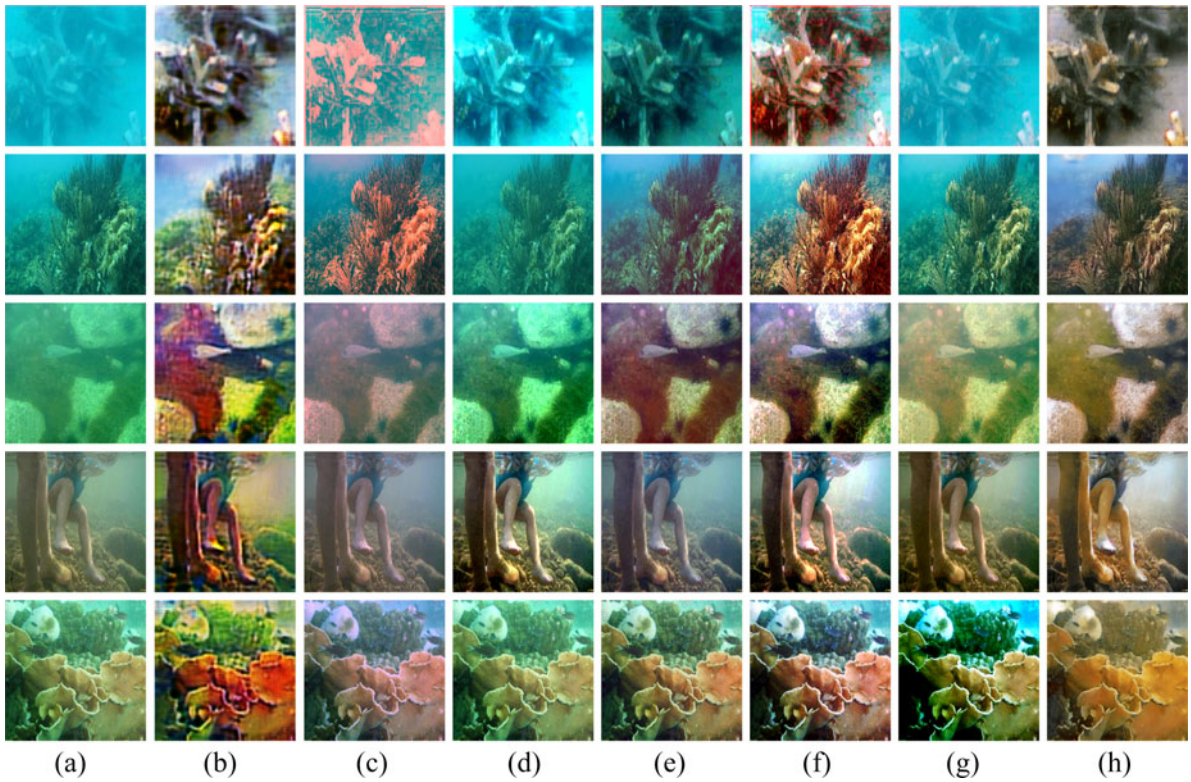

Fig. 4. Visual quality comparisons on varying underwater scenes. From top to bottom are image I1–I5. (a) Raws. (b) CycleGAN. (c) GW. (d) INT. (e) RED. (f) UWID. (g) UWIB. (h) Ours.

TABLE I
THE AVERAGE SCORES OF THE IMAGES PRESENTED IN FIG. 4

Method	I1	I2	I3	I4	I5
Raws	3	3	3	3	3
CycleGAN	3.3	2.5	2.8	2.9	3.5
GW	1.4	1.7	3.8	3.5	3
INT	3.1	3.1	3.5	4.9	3
RED	6.4	3.8	5.3	3.5	4.5
UWID	3.1	4.5	5.9	5.4	5.9
UWBI	3.5	4.1	5.5	4.5	3.5
Ours	**7.5**	**7.5**	**6.2**	**6.2**	**6.6**

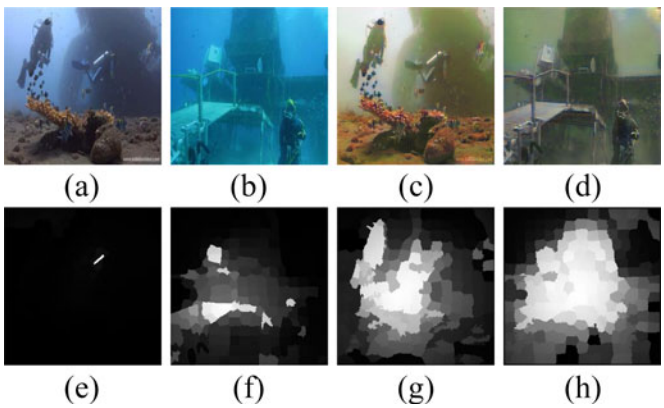

Fig. 5. Saliency detection test. (a) and (b) Raw underwater images. (c) and (d) Our results. (e) and (f) Saliency maps of (a) and (b). (g) and (h) Saliency maps of (c) and (d).

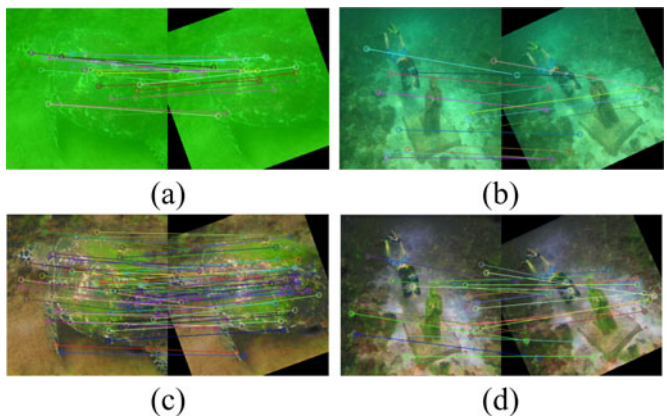

Fig. 6. Keypoint matching test. (a) and (b) Keypoint matching maps of raw underwater images. (c) and (d) Keypoint matching maps of our results.

IV. CONCLUSION

In this letter, we presented a novel method for underwater image color correction. Based on the learned cross domain relations, the proposed method can remove color distortion by weakly supervised model. It is the first attempt that corrects the color casts of underwater images by weakly supervised learning. Furthermore, our method can be used as a guide for subsequent research of the learning-based underwater image color correction. Experiments including the subjective assessment, user study, and application test demonstrated the effectiveness of the proposed method.

REFERENCES

[1] M. Bryson, M. Johnson, O. Pizarro, and S. B. Williams, "True color correction of autonomous underwater vehicle imagery," *J. Field Robot.*, vol. 33, no. 6, pp. 853–874, 2016.

[2] R. Gibson, R. Atkinson, and J. Gordon, "A review of underwater stereo-image measurement for marine biology and ecology applications," *Oceanogr. Marine Biol.: An Annu. Rev.*, vol. 47, pp. 257–292, 2016.

[3] D. He and G. Seet, "Divergent-beam lidar imaging in turbid water," *Opt. Lasers Eng.*, vol. 41, no. 1, pp. 217–231, 2004.

[4] Y. Schechner and N. Karpel, "Clear underwater vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2004, pp. 536–543.

[5] Y. Schechner and N. Karpel, "Recovery of underwater visibility and structure by polarization analysis," *IEEE J. Oceanic Eng.*, vol. 30, no. 3, pp. 570–587, Jul. 2005.

[6] T. Treibitz and Y. Schechner, "Active polarization descattering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 3, pp. 385–399, Mar. 2009.

[7] N. Carlevaris Bianco, A. Mohan, and R. Eustice, "Initial results in underwater single image dehazing," in *Proc. IEEE Conf. Oceans*, 2010, pp. 1–8.

[8] J. Chiang and Y. Chen, "Underwater image enhancement by wavelength compensation and dehazing," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1756–1769, Apr. 2012.

[9] A. Galdran, D. Pardo, and A. Picn, "Automatic red-channel underwater image restoration," *J. Vis. Commun. Image R.*, vol. 26, pp. 132–145, 2015.

[10] P. Drews-Jr, E. R. Nascimento, S. Botelho, and M. F. M. Campos, "Underwater depth estimation and image restoration based on single images," *IEEE Comput. Graph. Appl.*, vol. 36, no. 2, pp. 24–35, Mar./Apr. 2016.

[11] C. Y. Li, J. C. Guo, R. M. Cong, Y. W. Pang, and B. Wang, "Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior," *IEEE Trans. Image Process.*, vol. 25, no. 12, pp. 5664–5677, Dec. 2016.

[12] Y. Peng and P. Cosman, "Underwater image restoration based on image blurriness and light absorption," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1579–1594, Apr. 2017.

[13] S. Emberton, L. Chittka, and A. Cavallaro, "Underwater image and video dehazing with pure haze region segmentation," *Computer Vision and Image Understanding*, 2017, pp. 1–12.

[14] 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 Robot. Autom. Lett.*, vol. 3, no. 1, pp. 387–394, Jan. 2017.

[15] S. Zhang, T. Wang, J. Dong, and H. Yu, "Underwater image enhancement via extended multiscale retinex," *Neurocomputing*, vol. 245, pp. 1–9, 2017.

[16] J. Ahn, S. Yasukawa, T. Sonoda, T. Ura, and K. Ishii, "Enhancement of deep-sea floor images obtained by an underwater vehicle and its evaluation by crab recognition," *J. Marine Sci. Technol.*, vol. 22, no. 4, pp. 758–770, 2017.

[17] R. Schettini and S. Corchs, "Underwater image processing: State of the art of restoration and image enhancement methods," *EURASIP J. Adv. Signal Process.*, 2010, vol. 2010, Art. no. 746052.

[18] T. Haware and P. Gumble, "A review on underwater image scene enhancement and restoration using image processing," *Inter. J. Innovative Res. Elect., Electron., Instrum. Control Eng.*, vol. 5, no. 9, pp. 28–31, 2017.

[19] K. He, J. Sun, and X. Tang, "Single Image Haze Removal Using Dark Channel Prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.

[20] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3522–3533, Nov. 2015.

[21] K. Tang, J. Yang, and J. Wang, "Investigating haze-relevant features in a learning framework for image dehazing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 2995–3002.

[22] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "DehazeNet: An end-to-end system for single image haze removal," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5187–5198, Nov. 2016.

[23] G. Buchsbaum, "A spatial processor model for object colour perception," *J. Franklin Inst.*, vol. 310, no. 1, pp. 1–26, 1980.

[24] E. Land, "The retinex theory of color vision," *Sci. Amer.*, vol. 237, no. 6, pp. 108–128, 1977.

[25] M. Ebner, "Color Constancy," 1st ed. Hoboken, NJ, USA: Wiley, 2007.

[26] G. Finlayson and E. Trezzi, "Shades of gray and colour constancy," in *Proc. Color Imag.*, 2004, pp. 37–41.

[27] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neur. Inform. Process. Sys.*, 2014, pp. 2672–2680.

[28] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," arXiv preprint arXiv:1511.06434, 2015.

[29] Y. Zhu, T. Park, P. Isola, and A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vision (ICCV)*, 2017, pp. 2242–2251.

[30] T. Kim, M. Cha, H. Kim, J. Lee, and J. Kim, "Learning to discover cross-domain relations with generative adversarial networks," in *Proc. Int. Conf. Machine Learn. (ICML)*, 2017, pp. 1857–1865.

[31] M. Liu, T. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," in *Proc. Adv. Neural Inform. Process. Syst. (NIPS)*, 2017, pp. 700–708.

[32] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[33] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Trans. Computational Imaging*, vol. 3, no. 1, pp. 47–57, 2017.

[34] J. Johnson, A. Alahi, and L. Fei-Fei., "Perceptual losses for real-time style transfer and super-resolution" in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 694–711.

[35] P. Isola, J. Zhu, T. Zhou, and A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vision Pattern Recog. (CVPR)*, 2017, pp. 5967–5976.

[36] C. Ledig *et al.*, "Photo-realistic single image superresolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vision Pattern Recog. (CVPR)*, 2017, pp. 105–114.

[37] C. Li and M. Wand, "Precomputed real-time texture synthesis with Markovian generative adversarial networks," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 702–716.

[38] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv:1412.6980, 2017.

[39] Y. Gong and F. Sbalzarini, "A natural-scene gradient distribution prior and its application in light-microscopy image processing," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 1, pp. 99–114, Feb. 2016.

[40] H. Peng, B. Li, H. Ling, W. Hu, W. Xiong, and S. J. Maybank, "Salient object detection via structured matrix decomposition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 818–832, Apr. 2017.

[41] H. Bay, T. Tuytelaars, and L. Van, "Surf: Speeded up robust features," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 404–417.