MLFcGAN: Multilevel Feature Fusion-Based Conditional GAN for Underwater Image Color Correction

Xiaodong Liu, Zhi Gao, and Ben M. Chen

***Abstract*— Color correction for underwater images has received increasing interest, due to its critical role in facilitating available mature vision algorithms for underwater scenarios. Inspired by the stunning success of deep convolutional neural network (DCNN) techniques in many vision tasks, especially the strength in extracting features in multiple scales, we propose a deep multiscale feature fusion net based on the conditional generative adversarial network (GAN) for underwater image color correction. In our network, multiscale features are extracted first, followed by augmenting local features in each scale with global features. This design was verified to facilitate more effective and faster network learning, resulting in better performance in both color correction and detail preservation. We conducted extensive experiments and compared the results with state-of-the-art approaches quantitatively and qualitatively, showing that our method achieves significant improvements.**

***Index Terms*— Conditional generative adversarial network (cGAN), feature extraction and fusion, image enhancement, underwater image color correction.**

I. Introduction

UNDERWATER imaging has been proven valuable in numerous remote sensing applications [1], [2], where remote sensors such as sonar and light detection and ranging (LiDAR) are deployed conventionally. Due to recent advances in both hardware and algorithmic approaches, affordable and compact off-the-shelf underwater cameras are becoming popular, allowing people to easily collect images from a wide range of undersea worlds by either diverse or remotely operated submersibles. These captured underwater images and videos with color information are valuable resources for many underwater scientific remote sensing missions, such as marine biology [3] and ecological research [4].

Manuscript received July 31, 2019; revised October 2, 2019 and October 19, 2019; accepted October 24, 2019. Date of publication November 7, 2019; date of current version August 28, 2020. *(Corresponding authors: Xiaodong Liu; Zhi Gao.)*

X. Liu is with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117583 (e-mail: xiaodongliu@u.nus.edu).

Z. Gao is with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China (e-mail: gaozhinus@gmail.com).

B. M. Chen is with the Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Hong Kong, and also with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117583 (e-mail: bmchen@nus.edu.sg).

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/LGRS.2019.2950056

Compared with everyday images captured in air, underwater images typically suffer from color shift and relatively low quality due to light absorption and light scattering, posing significant challenges to available mature vision algorithms in achieving the expected performance. For underwater scenes, light absorption is wavelength dependent; the longer the wavelength, the higher the absorption rate is. Thus, the red component of light is absorbed first and the underwater images often appear bluish or greenish. Severe underwater light scattering is due to the relatively larger particles than those observed in air, resulting in decreased visibility. Moreover, such absorption and scattering effects are hard to be explicitly modeled, as they are relevant to many factors, such as water temperature, salinity, types of particles, and so on. This complex degradation makes it challenging to restore the visibility and color of underwater images [5]. Restoring underwater images with natural colors and fine details still remains an open problem [6]. Several pioneering works tackle this issue by inferencing the nondegraded images based on the image degradation model. As this inverse problem is ill-posed, prior knowledge or assumptions are introduced to obtain a solution. These include the approaches based on a dark channel prior [7] and its variants [8], methods based on haze-lines prior [9], and so on. However, the prior knowledge may fail for some underwater scenes, and all these model-based methods reported less competitive image color correction results [10].

Like many other computer vision tasks, underwater image color correction has been benefiting from the deep convolutional neural networks (DCNNs). In [10], the convolutional neural network (CNN) was trained to approximate the underwater image restoration function given the synthesized paired underwater images. Li *et al.* [11] proposed a two-stage CNN for depth estimation and color restoration. Recently, the generative adversarial networks (GANs) have achieved huge success on many tasks such as super-resolution [12] and image synthesis and translation [13]. Inspired by this, in [14], the conditional GAN (cGAN) was exploited to address the underwater image enhancement as the image-to-image translation problem. Based on [14], Yu *et al.* [15] introduced perceptual loss into the cGAN framework for underwater image color correction. Later, CycleGAN was introduced for color correction in [16] and [17]. Generally speaking, such CNN-based methods outperform aforementioned model-based methods. However, methods in [10], [11], and [14] leverage

only low-level local features with relatively shallow networks, and such low-level features captured with a limited receptive field can hardly encode the high-level semantic knowledge, resulting in noisy and imperfect color restoration results. To overcome the limitations of the available CNN-based methods, we propose to exploit high-level features in the cGAN framework for underwater image color correction. Leveraging on the high-level features, impressive improvements have been made for detection [18], segmentation [19], pose estimation [20], and so on. Different from those methods, we augment local features of each level with global features that capture the semantic information, such as the overall lighting condition and scene layout of the whole image, for underwater image color correction.

In this letter, we propose a generic multilevel features fusion-based conditional GAN (MLFcGAN) for underwater image color correction. Compared to the existing network structures, MLFcGAN extracts more scale of features, and the global features are fused with low-level features on each scale. Extensive experiments are conducted and comparisons with the state-of-the-art approaches are made quantitatively and qualitatively, showing that our design achieves significant improvements.

II. Proposed Method

Our design is based on the framework of cGAN, where the adversarial loss is beneficial to image generation compared to DCNNs with Euclidean loss [13]. Consisting of one generator G and one discriminator D, GAN is initially deployed to produce vivid images, given the noise input z. G aims to produce images to fool D and D is trained to distinguish samples from real images, where G and D are updated in an adversarial fashion. Slightly different from the basic GAN, cGAN takes conditional variables as input.

A. Generator

The generator is based on the encoder-decoder structure and the multiscale feature extraction and feature fusion unit are designed and novelly integrated, whose effectiveness is demonstrated in Section IV.

1) Global Features and Multilevel Local Feature Extraction: High-level information extracted from the image with a receptive field of the entire image is termed as the global features. The extraction of this information is prevalent in feature engineering, and can be achieved by the average pooling along the spatial dimension, as described in [21]. Unlike the average pooling, we extract the global features by gradually down-sampling the input images with convolution layers until the output has a dimension $1 \times 1 \times c_g$, where c_g represents the number of channels of global features. The benefits of gradually down-sampling compared to simple average pooling are of two-fold: First, the number of feature maps at each resolution are free to be chosen. Second, in this way, more scales of local features can be extracted simultaneously. The multiscale local features offer abstraction of the image at different resolutions, and are beneficial for generating images with fine details. (This is validated in Section IV.) As the encoder part of the generator functions as feature extraction, and to make the network deeper and easier to optimize, the residual building blocks are employed as proposed in [22].

2) Fusion of the Global and Local Feature Unit: Here, we propose the feature fusion unit to dynamically fuse the global features with local features. Suppose the local feature map f_l at scale i with dimension $h_i \times w_i \times c_i$ and the global features f_g with dimension $1 \times 1 \times c_g$. The global features are first adjusted through a 1×1 convolution layer with learnable weights for channel matching. Normally, c_g is larger than c_i, therefore this step is designed to adaptively extract the most useful information from global features for local features at scale i. The parameter settings for this convolution layer are: kernel size 1×1, stride 1, number of input channels c_g, and number of output channels c_i.

Denote F_{conv}, F_{copy}, F_{reshape}, and F_{concat} as the convolution, copy, reshape, and concatenate operation separately. The feature fusion unit follows the operations, as described in (1)–(4). The output after the convolution can be denoted as follows:

$$f_{g1} = F_{\text{conv}}(f_g, W) \tag{1}$$

where W denotes the learnable weights. Then, f_{g1} is copied in total $h_i \times w_i$ times

$$f_{g2} = F_{\text{copy}}(f_{g1}, \text{num} = h_i \times w_i) \tag{2}$$

Then, f_{g2} is reshaped into (h_i, w_i, c_i)

$$f_{g3} = F_{\text{reshape}}(f_{g2}, \text{size} = (h_i, w_i, c_i)). \tag{3}$$

Finally, the features f_l and f_{g3} with the same dimension are concatenated along the channel dimension

$$f_{\text{out}} = F_{\text{concat}}(f_l, f_{g3}). \tag{4}$$

The fused feature f_{out} is fed into the corresponding layer in the decoder with a skip connection.

Remarks: Local features and global features cover different scales of the image and convey variant knowledge of the image. Normally, the local image features represent low-level features such as edges. The global features encode the high-level information such as the overall light condition, the layout, or the type of the scene, and so on. Fusing with low and high information at different scales is beneficial to generate images with plausible natural color and better details. In addition, since the global features are a higher abstraction of local features, they could act as the regularizer to penalize the artifacts generated in the enhanced images due to mishandling in the low-resolution images. Hence, here, we fuse the global features with low-level features at each resolution, as shown in Fig. 1.

B. Discriminator

In this letter, PatchGAN is adopted as the discriminator [13]. PatchGAN is designed to identify if each $N \times N$ patch in the image is real or generated by G, and the overall decision is achieved by averaging the authenticity of all patches. In this case, the generated image will only be considered as real, when nearly all image patches are generated with good and less blurring details to be considered with high probability to

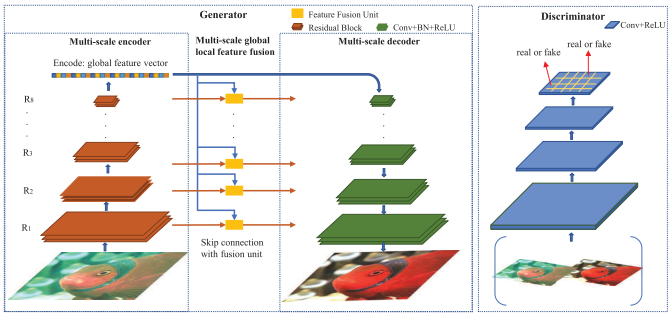


Fig. 1. Overview of the network structure. (Left) Generator which consists of the encoder, decoder, and multiscale local and GF with skip connection. (Right) PatchGAN-based discriminator.

be real. Considering the network is trained in an adversarial way, the generator is pushed to focus more on high frequency (details in the image) and generate better image details so as to fool the discriminator. On the other hand, compared to the whole-image discriminator, less convolutional layers are needed for PatchGAN. For more details, please see [13].

C. Objective Function

Considering the problem of exploding or vanishing gradient during training the original GAN [15], [23], many improvements and modifications have been made to stabilize the training such as the LSGAN [24], Wasserstein GAN [25], and Wasserstein GAN with gradient penalty (WGAN-GP), and the WGAN-GP shows the best performance for image generation [26]. In this letter, the WGAN-GP [26] loss is adopted and modified into conditional setting as the adversarial loss

$$\mathcal{L}_{c\text{WGAN-GP}} = E_{x,y}[D(x,y)] - E_x[D(x,G(x))] + \lambda E_{\hat{x}}[(||_{\hat{x}}D(\hat{x})||_2 - 1)^2] \quad (5)$$

where x and y are the original raw image and the ground-truth underwater image (with good color balance and details), respectively, $\hat{x}$ are the samples along the lines between the generated images $G(x)$ and y, and λ stands for the weight factor.

The adversarial loss measures the Wasserstein distance from the distribution perspective between the distribution of the generated images and that of the ground-truth images. The traditional loss such as the L_2 or L_1 loss measures the distance from the pixel perspective, and it has been demonstrated to be helpful to combine the adversarial loss with traditional distance loss for image-to-image translation tasks [27]. As reported in [13] and [28], the L_1 loss is likely to give less blurring results than the L_2 loss, therefore the L_1 loss is introduced

$$\mathcal{L}_{L_1}(G) = E_{x,y}[||y - G(x)||_1]. \quad (6)$$

The overall objective function $\mathcal{L}^*$ is (λ_1 is the weight factor)

$$\mathcal{L}^* = \min_G \max_D \mathcal{L}_{c\text{WGAN-GP}}(G,D) + \lambda_1 \mathcal{L}_{L_1}(G). \quad (7)$$

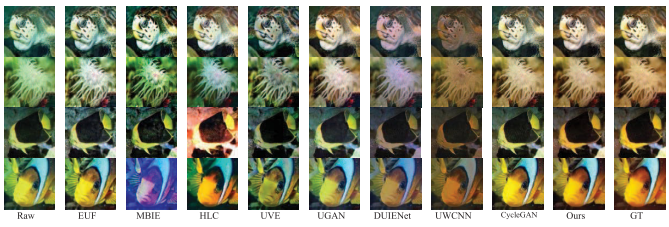


Fig. 2. Color correction comparisons of *val* images with EUF [29], MBIE [30], HLC [9], UVE [31], UGAN [14] and DUIENet [32], UWCNN [33], and CycleGAN [34]. Better 4× zoomed-in view.

III. Experiments and Analysis

A. Preparation

1) Data: Different from other computer vision tasks, there is limited publicly available data set for underwater images as the challenge to acquire the ground-truth. The data set used for training was recently proposed in [14]. It contains 6128 image pairs with ground truth (nondistorted underwater images) and distorted underwater images. We randomly select 6000 image pairs as the training set, and the remaining images are used for validation. In addition, to further evaluate the generalization ability of our method, 86 real world underwater images are collected from the Internet with various scenes. All the images are resized to 256×256.

2) Training Settings: In our experiment, we set $\lambda = 10$ and $\lambda_1 = 10$. The values are selected based on the basic hyperparameter tuning. We apply the Adam solver, with the learning rate $= 0.0002$, $\beta_1 = 0.5$, and $\beta_2 = 0.999$. The batch size is 1 and the network is trained for 50 epochs.

3) Methods for Comparison: Comparisons are made with the following state-of-the-art methods, which are published in top conferences or journals recently. 1) **Model-based methods**: EUF [29], MBIE [30], HLC [9], and UVE [31], and they are tested with the code provided by their authors. 2) **Learning-based methods**: UGAN [14] and DUIENet [32], UWCNN [33], and CycleGAN [34]. UGAN [14] and CycleGAN [34] are retrained on the data set from scratch with the recommended parameter settings in their letters to achieve the best enhancement results. DUIENet [32] and UWCNN [33] are tested with the pretrained model provided by the authors.

4) Evaluation Metrics: The evaluation is based on both the validation images and real-world underwater images. For the validation images set where the ground truth is available, the peak signal to noise ratio (PSNR) and the structural similarity index (SSIM) are adopted for quantitative comparisons.

B. Experimental Results

1) Evaluation of the Validation Image Set: The average PSNR and SSIM were calculated and shown in Table I. Fig. 2 demonstrates that our method can achieve the best image restoration performance, which is rather close to the ground truth. Quantitatively, as can be seen from Table I, our method achieves the best PSNR and SSIM and outperforms other methods with a large margin, demonstrating the superior learning ability of our structure.

2) Evaluation of Real World Underwater Images: Visual comparisons are presented in Fig. 3. UVE [31] and HLC [9] perform poorly to remove the color cast, and images processed

TABLE I

PSNR AND SSIM EVALUATION OF VALIDATION IMAGES

	EUF	MBIE	HLC	UVE	UGAN	DUIENet	UWCNN	CycleGAN	Ours
PSNR	16.36	14.67	15.72	15.58	18.58	19.6279	15.2351	23.0224	**23.42**
SSIM	0.5527	0.4244	0.5249	0.4716	0.5851	0.6431	0.6133	0.7893	**0.8158**

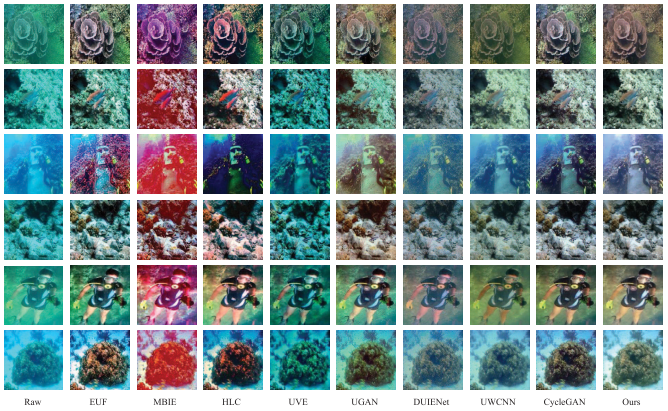


Fig. 3. Color correction result comparisons of *real* underwater images with EUF [29], MBIE [30], HLC [9], UVE [31], UGAN [14] and DUIENet [32], and UWCNN [33] and CycleGAN [34]. Better zoomed-in view.

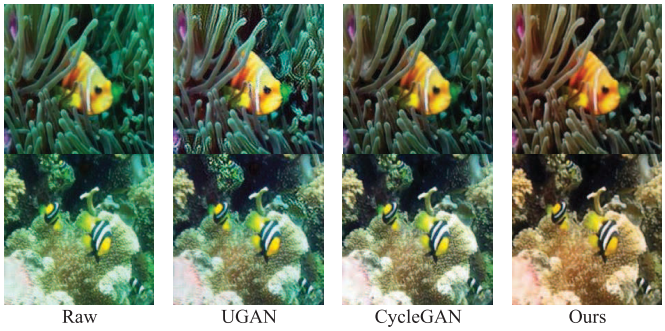


Fig. 4. Further comparisons with UGAN [14] and CycleGAN [34]. UGAN and CycleGAN may fail for deep green hue scenarios and generate more blurry results than ours.

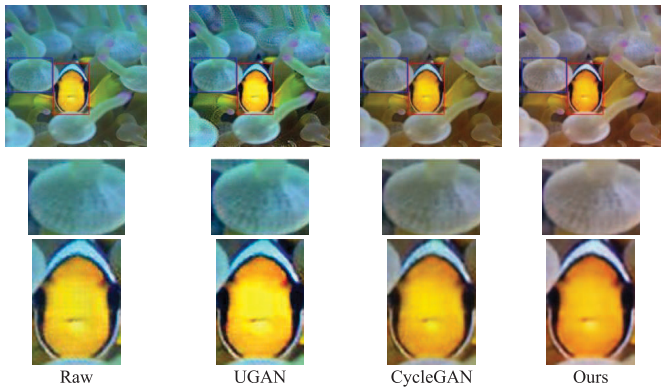


Fig. 5. Zoomed-in view comparisons of the image patches against UGAN [14] and CycleGAN [34], where they may generate more blurry results, and our method could restore natural color and produce smooth results without loss of details.

by these methods remain bluish or greenish. MBIE [30], on the other hand, can generate images with good color saturation, but it introduces unwanted superfluous red hue. Similarly, EUF [29] introduces unwanted artifacts and noise. The poor performance of the model-based methods may be because the prior knowledge or the assumed parameters may not hold for some underwater scenes. For DUIENet [32], it is likely to introduce unexpected gray hue (such as the images in the second and fourth rows) or red hue (such as the images in the first and fifth rows) with low brightness, as shown in Fig. 3. For UWCNN [33], similarly, the enhanced results appear dim with low intensity, as shown in Figs. 2 and 3. As UGAN and CycleGAN are retrained on the unified data set, we made further comparisons in Figs. 4 and 5. As for UGAN [14], the color correction performance is generally acceptable, but it is likely to generate noisy patches in the texture-less areas and along the boundary of the images. For CycleGAN [34], it may fail for some cases (such as the deep green scenarios) and retain the green hue with blurry details, as shown in Figs. 4 and 5. On the other hand, our method can achieve good color correction results with smooth details. As discussed earlier, only based on the limited scales of local features, it is likely to produce noisy patches. On the other hand, the proposed multiscale feature extraction and fusion strategy can help generate images with better preservation of details.

IV. GENERATOR STRUCTURE COMPARISONS

To evaluate the effectiveness of the proposed multiscale feature extraction and global feature design as well as the proposed generator structure, we make a comparison against the following generator model.

1) Generator used in [16] and [17] (Two scales of low-level features, which we term it as G-2 here for simplicity);
2) The basic U-Net as used in [14];
3) U-Net implemented with residual blocks (RBs) in the encoder;
4) U-Net augmented with global features fusion (GF);
5) Our generator without the global feature fusion: denoted Ours-GF;
6) Our proposed generator structure.

TABLE II

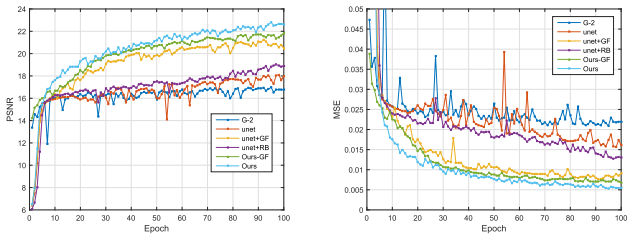
COMPARISON AND EVALUATION OF DIFFERENT GENERATOR STRUCTURES

	G-2	U-Net	UNet+RB	UNet+GF	Ours-GF	Ours
PSNR	16.78	18.11	18.85	20.46	21.82	**22.60**
MSE	0.022	0.016	0.013	0.0091	0.0068	**0.0055**

Here, we construct a small training set with 512 paired underwater images. The training is conducted with a batch size of 8 for 100 epochs. The objective for training is to minimize the mean square error (MSE). The PNSR and MSE results are reported in Table II, and the training process is shown in Fig. 6.

From Table II, we can have the following findings.

1) The number of scales of feature matters. The comparison between G-2 (2 scales of features), U-Net (5 scales of features), and Ours-GF (8 scales of features) shows



(a) (b)

Fig. 6. Comparisons for generator structures. Better zoomed-in view. (a) PSNR. (b) MSE.

that multiscale features are beneficial for image-to-image translation learning.

2) RBs in the encoder improves the learning ability.
3) Global feature fusion works as well. Overall, our proposed structure can achieve the highest PSNR and the smallest MSE, which demonstrate good learning ability and the effectiveness of the proposed generator structure.

V. Conclusion

In this letter, we propose a generic MLFcGAN under the framework of conditional GAN for underwater image color correction. Extensive experimental results demonstrate that by embedding the high-level information with low-level knowledge at multiple scales, MLFcGAN possesses better learning ability. Our method can effectively restore the underwater images' color with fine details and alleviate the unwanted artifacts, which outperform the state of the art both subjectively and objectively. Furthermore, as feature aggregation is fundamental in solving computer vision tasks via deep learning, our strategy could also be explored for other computer vision topics such as segmentation and salient object detection.

References

[1] S. Matteoli, G. Corsini, M. Diani, G. Cecchi, and G. Toci, "Automated underwater object recognition by means of fluorescence LIDAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 375–393, Jan. 2015.

[2] R. Fandos, A. M. Zoubir, and K. Siantidis, "Unified design of a feature-based ADAC system for mine hunting using synthetic aperture sonar," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2413–2426, May 2014.

[3] M. Ludvigsen, B. Sortland, G. Johnsen, and H. Singh, "Applications of geo-referenced underwater photo mosaics in marine biology and archaeology," *Oceanography*, vol. 20, no. 4, pp. 140–149, 2007.

[4] N. J. C. Strachan, "Recognition of fish species by colour and shape," *Image Vis. Comput.*, vol. 11, no. 1, pp. 2–10, 1993.

[5] H. Lu, Y. Li, Y. Zhang, M. Chen, S. Serikawa, and H. Kim, "Underwater optical image processing: A comprehensive review," *Mobile Netw. Appl.*, vol. 22, no. 6, pp. 1204–1211, 2017.

[6] M. Shortis and E. H. D. Abdo, "A review of underwater stereo-image measurement for marine biology and ecology applications," in *Oceanography and Marine Biology*. Boca Raton, FL, USA: CRC Press, 2016, pp. 269–304.

[7] 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.

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

[9] D. Berman, D. Levy, S. Avidan, and T. Treibitz, "Underwater single image color restoration using haze-lines and a new quantitative dataset," Nov. 2018, *arXiv:1811.01343*. [Online]. Available: https://arxiv.org/abs/1811.01343

[10] Y. Wang, J. Zhang, Y. Cao, and Z. Wang, "A deep CNN method for underwater image enhancement," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 1382–1386.

[11] 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. 2018.

[12] K. Jiang, Z. Wang, P. Yi, G. Wang, T. Lu, and J. Jiang, "Edge-enhanced GAN for remote sensing image superresolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5799–5812, Aug. 2019.

[13] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," Nov. 2018, *arXiv:1611.07004*. [Online]. Available: https://arxiv.org/abs/1611.07004

[14] C. Fabbri, M. J. Islam, and J. Sattar, "Enhancing underwater imagery using generative adversarial networks," Jan. 2018, *arXiv:1801.04011*. [Online]. Available: https://arxiv.org/abs/1801.04011

[15] X. Yu, Y. Qu, and M. Hong, "Underwater-GAN: Underwater image restoration via conditional generative adversarial network," in *Proc. Int. Conf. Pattern Recognit.* Cham, Switzerland: Springer, 2018, pp. 66–75.

[16] C. Li, J. Guo, and C. Guo, "Emerging from water: Underwater image color correction based on weakly supervised color transfer," *IEEE Signal Process. Lett.*, vol. 25, no. 3, pp. 323–327, Mar. 2018.

[17] J. Lu, N. Li, S. Zhang, Z. Yu, H. Zheng, and B. Zheng, "Multi-scale adversarial network for underwater image restoration," *Opt. Laser Technol.*, vol. 110, no. 1, pp. 105–113, 2019.

[18] J. Dai, K. He, Y. Li, S. Ren, and J. Sun, "Instance-sensitive fully convolutional networks," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 534–549.

[19] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[20] W. Yang, W. Ouyang, H. Li, and X. Wang, "End-to-end learning of deformable mixture of parts and deep convolutional neural networks for human pose estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 3073–3082.

[21] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[23] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[24] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2813–2821.

[25] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," Jan. 2017, *arXiv:1701.07875*. [Online]. Available: https://arxiv.org/abs/1701.07875

[26] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of wasserstein GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5767–5777.

[27] D. Pathak, P. Krähenbühl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2536–2544.

[28] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for neural networks for image processing," Nov. 2015, *arXiv:1511.08861*. [Online]. Available: https://arxiv.org/abs/1511.08861

[29] C. Ancuti, C. O. Ancuti, T. Haber, and P. Bekaert, "Enhancing underwater images and videos by fusion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 81–88.

[30] Y. Cho, J. Jeong, and A. Kim, "Model-assisted multiband fusion for single image enhancement and applications to robot vision," *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 2822–2829, Oct. 2018.

[31] Y. Cho and A. Kim, "Visibility enhancement for underwater visual SLAM based on underwater light scattering model," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May/Jun. 2017, pp. 710–717.

[32] C. Li *et al.*, "An underwater image enhancement benchmark dataset and beyond," Jan. 2019, *arXiv:1901.05495*. [Online]. Available: https://arxiv.org/abs/1901.05495

[33] C. Li, S. Anwar, and F. Porikli, "Underwater scene prior inspired deep underwater image and video enhancement," *Pattern Recognit.*, vol. 98, Feb. 2020, Art. no. 107038.

[34] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2223–2232.