CA-GAN: Class-Condition Attention GAN for Underwater Image Enhancement

JING WANG1,3, PING LI2, JIANHUA DENG3, YONGZHAO DU1,4, JIAFU ZHUANG5, PEIDONG LIANG6, AND PEIZHONG LIU1,3,4

1 College of Engineering, Huaqiao University, Quanzhou 362021, China
2 College of Information Science and Engineering, Huaqiao University, Xiamen 361021, China
3 ZhongfangHongye Information Technology Corporation, Quanzhou 362021, China
4 Department of Clinical Medicine, Quanzhou Medical College, Quanzhou 362000, China
5 Quanzhou Institute of Equipment Manufacturing, Haixi Institutes, Chinese Academy of Sciences (CAS), Quanzhou 350001, China
6 Fujian (Quanzhou) HIT Engineering and Technology Institute, Quanzhou 362000, China

Corresponding author: Ping Li (pingping_1213@126.com)

This work was supported in part by the Promotion Program for Young and Middle-aged Teacher in Science and Technology Research of Huaqiao University under Grant ZQN-PY518 and 509, in part by the National Natural Science Foundation of China under Grant 61605048, in part by Fujian Provincial Big Data Research Institute of Intelligent Manufacturing, in part by the Quanzhou scientific and technological planning projects under Grant 2017G024, Grant 2018C113R, Grant 2019C028R, Grant 2019C029R, Grant 2019C076R, and Grant 2019C099R, and in part by the Subsidized Project for Postgraduates’ Innovative Fund in Scientific Research of Huaqiao University under Grant 1611422001.

ABSTRACT

Underwater images suffer from serious color distortion and detail loss because of the wavelength-dependent light absorption and scattering, which seriously influences the subsequent underwater object detection and recognition. The latest methods for underwater image enhancement are based on deep models, which focus on finding a mapping function from the underwater image subspace to a ground-truth image subspace. They neglect the diversity of underwater conditions which leads to different background colors of underwater images. In this paper, we propose a Class-condition Attention Generative Adversarial Network (CA-GAN) to enhance an underwater image. We build an underwater image dataset which contains ten categories generated by the simulator with different water attenuation coefficient and depth. Relying on the underwater image classes, CA-GAN creates a many-to-one mapping function for an underwater image. Moreover, in order to generate the realistic image, attention mechanism is utilized. In the channel attention block, the feature maps in the front-end layers and the back-end layers are fused along channels, and in the spatial attention block, feature maps are pixel-wise fused. Extensive experiments are conducted on synthetic and real underwater images. The experimental results demonstrate that CA-GAN can effectively recover color and detail of various scenes of underwater images and is superior to the state-of-the-art methods.

INDEX TERMS

Underwater image, generative adversarial network.

I. INTRODUCTION

Underwater image processing is very critical to deep ocean exploration [1]. The quality of underwater image is extremely important to vision applications such as object recognition, detection and tracking. However, underwater images always suffer from serious color distortion and detail loss because of the wavelength-dependent light absorption and scattering. With the increase of water depth, the image becomes darker and darker as more and more light are absorbed [2]. Besides, the degradation is affected by a large number of complex factors including water component, quality and temperature. Thus, underwater image enhancement is still a challenging task. Moreover, the research of underwater image enhancement is still at the stage of the begining, and there are still many problems to be discussed.

Recently, attention has been attracted on underwater image enhancement. The existing literatures are categorized into two classes: the traditional methods and the deep learning based methods. The former employs the low-level image processing tools to enhance underwater image, such as color correction, contrast stretch, and dehaze [4], [8]. They have difficulties in handling the varying background of underwater images, thus, they are deficiency in generalization.

The associate editor coordinating the review of this manuscript and approving it for publication was Gulistan Raja.
Inspired by the success of deep learning in object classification and speech recognition, deep models are increasingly implemented on image restoration, such as dehazing [9], derain, deblur, and image inpainting. As we know, a good deep model requires large labeled training dataset. However, there is only a few underwater image datasets until now. There are big space to improve the performance of underwater image enhancement. Recently, Generative Adversarial Network (GAN) [3] was rised up, and it is recognized that GAN favor the visual effect over other deep learning based models. Gradually, GAN has been employed for underwater image enhancement. Most existing GAN based methods for underwater image enhancement treat the image enhancement as an image-to-image translation [10], [18] and have achieved promising results. However, the existing GAN models for underwater image enhancement are limited in three folds: 1) the underwater image dataset is deficient in diversity. 2) the existing methods usually learn only a mapping function for any underwater images, neglect that the underwater image varies with the increase of water depth [11]. 3) The existing GAN models have space to be improved with the development of GAN.

In this paper, we build an underwater image dataset in which underwater images are synthesized by a simulator based on physical optical model with randomly generated depth parameters. Our dataset contains 70000 pairs images which are classified into 10 classes according to the water depth. Moreover, we propose a Class-conditioned Attention GAN (CA-GAN) for underwater image enhancement in which an underwater image is classified first and then the class label guides the generating of enhanced images. Actually, we design a water class embedding block (WCEB) which normalizes different classes of underwater images to an in-air image. Furthermore, we implement attention mechanism on the generator in order to amplify the relevant feature and to suppress the irrelevant feature for underwater image enhancement. A concurrent channel and spatial attention feature fusion block (CS-AFFB) is employed to recalibrate the feature maps in shallow layers and deeper layers.

The contributions are summarized as follows:

- We propose a class-conditioned attention GAN for underwater image enhancement. The class embedding block normalizes an underwater image in color. With the help of class prediction, our GAN model focuses on mapping different classes of underwater images to an in-air image.
- We build a new dataset that contains 70,000 pairs images with eight classes classified according to water depth.

II. RELATED WORKS

Underwater image enhancement is a challenging while important task, as underwater images suffer from dramatically degrades of visibility and color distortions. There mainly exist two classes of methods for underwater image enhancement: hand-crafted priors based methods and deep models for underwater image enhancement.

A. HAND-CRAFTED PRIORS BASED METHODS

The traditional methods mainly depend on hand-crafted prior together with the physical scattering model. Yang et al. [21] use dark channel prior to estimate the depth map of an underwater image and adopt a color contrast method to enhance contrast. Carlevaris et al. [22] compare the maximum intensity among different color channels to estimate the scene depth, and use MAP estimator to recover the scene. Different to [21] and [22], Peng et al. [23] utilize image blurriness and light absorption to estimate the scene depth which can produce better enhancement results in different light conditions and color tones. Li et al. [24] restore an underwater image by minimizing the information loss, then base on the histogram distribution to enhance the contrast and brightness. Yang et al. [25] firstly implement the histogram equalization on an underwater image and then decompose the results into the structure and texture layers based on total variation and L1 norm minimization, followed by enhancing the contrast of the structure layer and denoising the texture layer. Ancuti et al. [25], [30] try to unify the intensity correction with the edge sharpening in a multi-scale fusion way. The hand-crafted prior based methods are deficiency in generalization. When the assumption is not satisfied, they cannot achieve good performance. Thus, in wild world, they often perform poorly, e.g. undesired artifacts, color distortion, and so on.

B. DEEP MODELS FOR UNDERWATER IMAGE ENHANCEMENT

Recently, many GAN based methods have been employed for underwater image enhancement. In [11], a UIE-Net was proposed to enhance an underwater image, which contains a Color Correction subnetwork(CC-Net) that estimates attenuation coefficient and a Haze Removal subnetwork (HR-Net) that estimates the transmission map, after which we obtain the enhanced image based on the formation of underwater image. Different to [11], many works enhance underwater images based on CycleGAN which enable unpaired image-to-image transformations. Li et al. [19] proposed a waterGAN method to generate underwater image sets relying on a physical model of underwater image, and train a two-stage image restoration network for color correction using synthetic images. Although waterGAN can get the state-of-the-art results, it needs to train various models for different water scence and they cannot cope with the diverse water environment. Fabbri et al. [18] directly adopt CycleGAN to generate paired images by using the subsets of ImageNet [26] to train the Underwater GAN (UGAN) restoration network with L1 loss and Gradient Difference Loss as the objective function. Li et al. [10] designed a GAN model to map real underwater images to in-air images by minimizing the adversarial loss, cycle consistency loss and SSIM loss to make the content and structure of recovered image look like without water. Different from these works, we use in-air images and their corresponding depth map to synthesis underwater
images to construct a dataset which contains eight types of images in various water depths. We also propose a GAN model that embedded with water class information to map various classes of underwater images to one in-air scene with just one trained model.

### III. DATASET CONSTRUCTION

We employ the underwater imaging model similar to [1] to synthesize underwater images which is formulated as:

$$I_c = t_c(p) J_c(p) + (1 - t_c(p)) A_c$$

where $c \in \{R, G, B\}$, $I_c(p)$ is the captured underwater image in the color channel $c$, $J_c(p)$ is the scene radiance, $A_c$ is the homogeneous global veiling-light component, and $p$ is a pixel location in underwater scene. $t_c(p)$ is the transmission of the color channel $c$, which depends on the distance $d(p)$ of the scene and the water attenuation coefficient $\beta_c$:

$$t_c(p) = \exp(-\beta_c d(p))$$

Different from image dehaze model [9], $\beta_c$ is wavelength-dependent and affected by the variations of season, geography and climate, causing objects to appear blue, green or yellow.

To synthesize the underwater images, we use the attenuation coefficients which are measured by Jerlov and Colleagues [6] during the Swedish Deep Sea Expedition of 1947-48. It includes five coastal waters (1, 3, 5, 7, 9) and five ocean waters (I, IA, IB, II, III). The water 1 is the clearest and water 9 is the most turbid coastal. Likewise, the water I is the clearest and water III is the most turbid.

We build the Realistic Single Underwater Image Enhancement (RSUIE) dataset based on RESIDE dataset [7] which is a large-scale benchmark dataset for image dehazing algorithms, it contains 8970 nature scene images and their corresponding depth map generated by [5]. In order to adapt to the problem of underwater image enhancement, we screen the original RESIDE dataset and delete the image that looks too dark or too foggy and finally we obtain a dataset which containing 70,000 images.

We implement Eq.(1), Eq.(2) on RESIDE to synthetic the 10 types of coastal water and ocean water images. As the coastal water classes face huge attenuation in deep water, such as the water type 9 above 5m and water 3 above 10m, which makes the objects almost invisible. However, some classes of water have small attenuation in shallow water, such as water type I, IA, IB about 1m to 5m, which have little effect on objects. Therefore we set different depth-ranges for different water classes: set the depth-range of water 9, 7 and 5 to [1, 5], set the depth-range of water 3 and 1 to [1, 15] and set the depth range of water I, IA, IB, II, III to [5, 20]. At the same time, we select a random global veiling-light $A_c \in [0, 1]$. Finally, we use the first 6,600 clean color images and depth images to synthesize training set and the other 400 images to synthesize validation (test) set. Therefore the RSUIE dataset contains a training set of 66,000 samples and a validation (test) set of 4,000 samples. Figure 4 shows 10 different samples for one RGB image and its corresponding depth map.

### IV. PROPOSED METHOD

The goal in this paper is to get the clear image $J$ when given an underwater image $I$ and its corresponding water type $c$ as inputs. The enhance process is done by a Class-condition Attention Generative Adversarial Network (CA-GAN), the details of proposed network architecture are shown in Fig.1, and for each image I the generator estimates its corresponding clear image $J$. The generator of CA-GAN contains two strided convolution blocks with stride 2, nine residual blocks and two upsample blocks with 3*3 convolution layer. Every convolution layers except the last one in CA-GAN are followed by spectral normalization [13], water class embedding block (WCEB) and ReLU activion layer. The spectral normalization layer used here can not only avoid unusual gradients, but also reduce the update times of discriminator [13]. And we supply the generator with class-conditional gians and bias in WCEB to map different types of underwater images to one clear natural scene image. Particularly, to recalibrate front-end feature map produced in the encoder layers and back-end feature map produced in the decoder layer, we introduce a concurrent channel and spatial attention feature fusion block (CS-AFFB). In addition, during the training phase, we apply SNGAN discriminator [13] and train both networks in an adversarial manner.

In the following subsections, we will introduce the details of WCEB and CS-AFFB in Section 4.1 and Section 4.2, respectively. Then, we describe our loss functions that allow simultaneously recover general content of underwater images and enhance their details.

#### A. WATER CLASS EMBEDDING BLOCK

The class information can be fed into the network in various method. In [14] Odena concats a one-hot class vector with the input vector, and maximize the the log-likelihood of the real image and the log-likelihood of the correct class. In [15] the conditional instance normalization was proposed to generate images in completely different styles by feeding the generator with class-conditional gians and bias in instance normalization layer.

In this paper, we propose a water class embedding block (WCEB) to constrain the feature space of every class of underwater images. As shown in Figure 2, we first encode the water class into a one-hot vector. Next we use a fully connected layer to map the encoded vector to a 48-dimensional vector. Then use two fully connected layers to map the 48-dimensional vector to gians $\gamma$ and bias $\beta$, respectively. Finally, we get class-conditional instance normalization results by $WCEB(x, c) = \gamma \left( \frac{x - \mu(x)}{\sigma(x)} \right) + \beta$, where $x$ is the feature map, $c$ is the class of input image, $\mu(x)$ and $\sigma(x)$ are mean and standard deviation of the feature map that computed across spatial dimensions independently for each channel and each sample.
1) CHANNEL ATTENTION FEATURE FUSE BRANCH
We introduce channel attention feature fuse branch (C-AFFB) to emphasize the important channels and ignore the less important ones. We consider the given front-end input feature map (represented by dotted line) $F_l = [f_{l1}, \ldots, f_{li}, \ldots, f_{lC}]$ and back-end input feature map (represented by solid line) $F_h = [f_{h1}, \ldots, f_{hi}, \ldots, f_{hC}]$ as a combination of channels $f_{li} \in R^{H \times W}$ and $f_{hi} \in R^{H \times W}$, respectively, where C is the number of channels, W is the width of feature map, H is the height of feature map. As shown in Fig. 3, the channel feature fuse branch reweight different channels of the front-end feature map and fuse the back-end feature map. First, the two feature maps $F_l$ and $F_h$ are concatenated. Next, transform the concatenated feature map to a $1 \times 1 \times C$ tensor by a global average pooling layer followed by two fully-connected layers and a ReLU layer. Then, the weight range $W_c = [w_{l1}, w_{l2}, \ldots, w_{lc}]$ of $F_l$ is constrained to [0, 1] by a Sigmoid layer. Finally, the weighted $F_l$ is added to $F_h$ and get the channel fuse feature map $F_{cf} = ReLU (F_h + W_c * F_l) = ReLU (\{f_{h1} + w_{cf}f_{l1}, \ldots, f_{hi} + w_{cf}f_{li}, \ldots, f_{hC} + w_{cf}f_{lC}\})$.

2) SPATIAL ATTENTION FEATURE FUSE BRANCH
We introduce the spatial attention feature fuse branch (S-AFFB) to squeeze the feature map for fine-grained underwater image enhancement. Here we consider the front-end input feature map (represented by dotted line) $F_{ls} = \left[ f_{ls}^{1,1}, \ldots, f_{ls}^{i,j}, \ldots, f_{ls}^{H,W} \right]$ and back-end input feature map (represented by solid line) $F_{hs} = \left[ f_{hs}^{1,1}, \ldots, f_{hs}^{i,j}, \ldots, f_{hs}^{H,W} \right]$ as a slicing of input channel, where $f_{is}^{i,j} \in R^{1 \times 1 \times C}$ and $f_{hs}^{i,j} \in R^{1 \times 1 \times C}$. The spatial feature weight $W_s = \left[ w_s^{1,1}, \ldots, w_s^{i,j}, \ldots, w_s^{H,W} \right]$ is achieved through a 1*1 convolution layer followed by a Sigmoid layer. And the spatial fusion feature map $F_{sf}$ is formulated as:

$$F_{sf} = ReLU (F_h + W_s * F_l)$$
$$= ReLU \left( \left[ f_{hs}^{1,1} + w_{sf}^{1,1} f_{ls}^{1,1}, \ldots, f_{hs}^{W,H} + w_{sf}^{W,H} f_{ls}^{W,H} \right] \right).$$  (3)

3) CONCURRENT CHANNEL AND SPATIAL ATTENTION FEATURE FUSION
Finally, we obtain concurrent channel and spatial attention fuse feature map $F_{cf}$ by element-wise addition of $F_{cf}$ and $F_{sf}$. $F_{cf} = F_{cf} + F_{sf}$. This fuse method encourages the network to pay more attention to meaningful features, which have higher activation both on channel-wise and saptial-wise.

C. LOSS FUNCTION
We formulate the loss function as a combination of Mean Squared Error (MSE), feature-based loss $L_F$ and adversarial loss $L_G$. The MSE loss and feature-based loss focus on...
FIGURE 4. Ten types of synthesized underwater images using a RGB image and its corresponding depth image.

enhancing general contents while adversarial loss focus on enhancing texture details.

1) MSE LOSS
MSE loss of generated image $I_{en}$ and target image $I_t$ is:

$$L_{MSE} = \frac{1}{CWH} \| I_t - I_{en} \|_2^2.$$  \hspace{1cm} (4)

2) FEATURE-BASED LOSS
Feature-based loss aims to measure the difference of the generated and target image CNN feature maps, $L_F$ is defined as

$$L_F = \frac{1}{CWH} \| F(I_t) - F(I_{en}) \|_2^2,$$  \hspace{1cm} (5)

where $F$ represents a non-linear CNN transformation and the size of feature map is $W \times H$ with $C$ channels. In this paper the feature map is obtained by the 3-th convolution layer before the 3-th pooling layer within the VGG19 model.

3) ADVERSARIAL LOSS
In CA-GAN, the discriminator and generator are trained in an alternating fashion by minimizing the hinge version of the adversarial loss:

$$L_D = -E[\min(0, -1 + D(I_t, y))] - E[\min(0, -1 - D(I_{en}, y))].$$  \hspace{1cm} (6)

$$L_G = -E[D(I_{en}, y)].$$  \hspace{1cm} (7)

Combine (4), (5) and (6) we get our final objective function:

$$L = L_{MSE} + \lambda_F L_F + \lambda_D L_G$$  \hspace{1cm} (8)

where $\lambda_F$ and $\lambda_D$ are weighting parameter to control the trade-off among the adversarial loss, the MSE loss and the feature-based loss.

V. EXPERIMENTAL RESULTS
In this section we present the implementation details and evaluation results on synthetic and real underwater images.

The proposed CA-GAN algorithm is compared with several state-of-the-art underwater image enhancement methods on both synthetic and real-world underwater images. These methods can be separated into prior information based methods and deep CNN based methods. The prior information based methods include Histogram Distributions Prior (HDP) [24], Retinex based enhancing approach (Retinex) [29], Fusion enhance (Fusion) [30], White Balance Correction and Image Decomposition (WBCID) [31]. The deep CNN based methods are DPATN [27], UGAN-P [18], WaterNet [35] and Un-GAN [33]. We run all the comparison methods using their recommended parameter settings, specially.

| Methods | Types | PSNR(dB) | SSIM | MSE |
|---------|-------|----------|------|-----|
| CA-GAN  | Type-9| 22.924   | 0.837| 57.844 |
|         | Type-3| 23.4     | 0.879| 69.635 |
|         | Type-II| 26.356  | 0.913| 64.374 |
| -woCS-AFFB | Type-9| 22.257   | 0.832| 75.692 |
|         | Type-3| 22.744   | 0.841| 82.711 |
|         | Type-II| 24.442  | 0.89 | 97.91  |
| -woWCEB | Type-9| 18.904   | 0.814| 121.03 |
|         | Type-3| 18.005   | 0.773| 121.621|
|         | Type-II| 22.273  | 0.876| 104.712|

FIGURE 5. Visual comparison of underwater image enhancement results on Type-9, Type-3 and Type-II test set with and without WCEB and CS-AFFB in our proposed CA-GAN.
A. IMPLEMENTATION DETAILS
In our experiment, the CA-GAN was implemented using the Pytorch deep learning framework and executed on a computer with Intel(R) Core(TM) i7-7700 CPU @ 4.2GHz CPU, Nvidia GTX 1080Ti GPU. During the CA-GAN training, we use random rotation and horizontal flipping for data augmentation and images in the training set of RSUIE datasets are resized to 256×256. We train the proposed CA-GAN in an alternating manner that performs one gradient descent step on discriminator first, and one step on generator using Adam optimizer with initial learning rate \( lr = 0.0002, \) \( \beta_1 = 0.5, \) \( \beta_2 = 0.999. \) The batch size is set to 1, the weighting parameter \( \lambda_F \) is set to 0.5 and \( \lambda_G \) is set to 0.1. For testing, it takes about 0.14s for resolution 512×512 on the above-mentioned machine, which is about 7 FPS. Particularly, the underwater images belong to 10 different classes corresponding to the synthesized data, and we train a VGG16 classifier using the RSUIE training set to predict the water class of input image while testing. The test accuracy of the VGG16 classifier is 95.68% on RSUIE test set.

B. ABLATION STUDY
In this subsection, we would like to demonstrate the effectiveness of the proposed WCEB and CS-AFFB in our CA-GAN by conducting ablation study. Both quantitative and qualitative evaluations on RSUIE test set are reported by removing WCEB (\(-\text{woWCEB}\)) and CS-AFFB (\(-\text{woCS-AFFB}\)) components, respectively.

Due to the limitation of space, we only tabulate the average PSNR, SSIM and MSE results on Type-9, Type-3 and Type-II synthetic images in RSUIE test set in Table 1 as they apparently show different colors. From Table 1, we can see that fed the water type information into the net by WCEB can significantly improve the performance of underwater image enhancement, especially for the turbid Type-9 subset. We also observe that CS-AFFB could achieve better performance compared with other feature fusion strategies. A visual comparison is presented in Fig. 5. As we can see, the model without WCEB can not enhance all types of images, and the model with CS-AFFB can get much more realistic results and have less artifacts.

C. RESULTS ON SYNTHETIC DATASET
The performance of CA-GAN on RSUIE synthetic dataset is evaluated in terms of PSNR, SSIM and MSE, we report the average scores of 10 different classes of underwater images in Tabel 2 and bold the best results. As we can see, Un-GAN and Fusion are effective only on ocean water classes, i.e., Type-1 and Type-3. While HDP, HLP and WBCID can hardly achieve good performance on any classes of underwater images. On the contrary, the proposed CA-GAN achieves the best PSNR, SSIM and MSE values on various classes of underwater images, which indicates that our CA-GAN can better restore the content, structure and texture information of an underwater image.

Furthermore, Fig. 6 presents a visual comparisions for 10 different classes of underwater images in the test set of RSUIE synthetic dataset. It is visible that most methods cannot recover appropriate color of the scenes and produce color deviations. Among them, DPATN introduces halo artifacts and color distortion in most images, and Fusion method can obviously improve the contrast but over-enhance the red channel on some images. In comparison, the results of proposed CA-GAN have more realistic colors that closer to the ground truth images, and it’s effective for deep underwater images that severely degraded.

D. RESULTS ON REAL-WORLD UNDERWATER IMAGES
We employ the underwater image quality set (UIQS) [28] and underwater color cast set (UCCS) [28] which are recently
benchmark dataset to evaluate the performance of CA-GAN on real-world underwater images. The images in RUIE are taken in the water near Zhangzi island in the Yellow Sea, and the water depth varies from 5 to 9 meters, the scene depths range from 0.5 to 8 meters. Specially UIQS contains 3630 images in various underwater conditions which is used to test the algorithms’ improvement of image visibility; and UCCS contains 300 images with bluish, greenish and blue-green tones that is used to test the algorithms’ ability of correcting color cast.

The qualitative comparison results are presented in Fig. 7, we can observe that our proposed CA-GAN presents the best visual experience on all images, that it can not only correct the color cast, but also can enhance the visibility of details. But WBCID have almost no effect on various underwater images. HDP, HLP, Fusion and DPATN can partially enhance the underwater images, but hard to recover the scene color. And the results of Un-GAN tends yellowish color.

As ground truth images of real underwater images are unavialable, we employ the PI [12] to evaluate the visual effect with computable index which has been verified effective in single image super-resolution task. The smaller the PIs the better the results. From Fig. 7, we can also observe that CA-GAN achieves the best PI among the state-of-the-art methods.

We also give a visual results on serval widely used real-world test images in Fig. 8. For fair comparison, we directly download the results of Fusion17 [25] and HLP18 [34] from their project page. Fig. 8 also demonstrates the effective of CA-GAN on underwater images of various color tones. As we...
FIGURE 7. Visual comparison of underwater image enhancement results on RUIE dataset. PIs are shown on the top of each image which are the smaller the better.

FIGURE 8. Visual comparison of Fusion18, HLP and CA-GAN on real-world images.

can see, different from Fusion17, which over-enhances the image detail and noise, the proposed CA-GAN can properly restore the serious underwater image details loss. At the same time, the results of proposed CA-GAN can display more realistic scene colors and fewer artifacts comparing to HLP18.

VI. CONCLUSION

In this paper, we have introduced a class-condition attention generative adversarial network (CA-GAN) for underwater image enhancement. It employs a water class embedding block to map different types of underwater images to one clear natural scene image and introduce a concurrent channel and spatial attention feature fusion block to recalibrate front-end feature map produced in the encoder layers and back-end feature map produced in the decoder layer. Experimental results on both synthetic and real underwater images demonstrate that CA-GAN can effectively recover the color and detail of various scenes of underwater images and is superior to the state-of-the-art methods.

REFERENCES

[1] B. L. McGlamery, “A computer model for underwater camera systems,” Proc. SPIE, vol. 0208, Mar. 1980, Art. no. 958279.
[2] J. Ronald and V. Zaneveld, Light and Water: Radiative Transfer in Natural Waters, 1995.
[3] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in Proc. Adv. Neural Inf. Process. Syst., 2014, pp. 2672–2680.
[4] D. Berman, T. Treibitz, and S. Avidan, “Diving into haze-lines: Color restoration of underwater images,” in Proc. Brit. Mach. Vis. Conf. (BMVC), 2017, vol. 1, no. 2, pp. 1–12.
[5] F. Liu, C. Shen, G. Lin, and I. Reid, “Learning depth from single monocular images using deep convolutional neural fields,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, no. 10, pp. 2024–2039, Oct. 2016.
[6] Y. Y. Schechner and N. Karpe, “Recovery of underwater visibility and structure by polarization analysis,” IEEE J. Ocean. Eng., vol. 30, no. 3, pp. 570–587, Jul. 2005.
[7] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, “Benchmarking single image dehazing and beyond,” 2017, arXiv:1712.04143. [Online]. Available: http://arxiv.org/abs/1712.04143
[8] K. Nomura, D. Sugimura, and T. Hamamoto, “Underwater image color correction using exposure-bracketing imaging,” IEEE Signal Process. Lett., vol. 25, no. 6, pp. 893–897, Jun. 2018.

[9] H. Zhang and V. M. Patel, “Densely connected pyramid dehazing network,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 3194–3203.

[10] Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor, “The 2018 pirm challenge on perceptual image super-resolution,” in Proc. ECCV Workshops, vol. 5, 2018, pp. 334–355.

[11] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, “Spectral normalization for generative adversarial networks,” 2018, arXiv:1802.05957. [Online]. Available: http://arxiv.org/abs/1802.05957

[12] A. Odena, C. Olah, and J. Shlens, “Conditional image synthesis with auxiliary classifier GANs,” in Proc. 34th Int. Conf. Mach. Learn., vol. 70, 2017, pp. 2642–2651.

[13] N. Van Noord and E. Postma, “A learned representation of artist-specific colourisation,” in Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW), Oct. 2017, pp. 2907–2915.

[14] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, “DeblurGAN: Blind motion deblurring using conditional adversarial networks,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 8183–8192.

[15] H. Zhang, V. Sindagi, and V. M. Patel, “Image de-reaining using a conditional generative adversarial network,” IEEE Trans. Circuits Syst. Video Technol., early access, Jun. 3, 2019, doi: 10.1109/TCSVT.2019.2924007.

[16] C. Fabbri, M. J. Islam, and J. Sattar, “Enhancing underwater imagery using generative adversarial networks,” in Proc. IEEE Int. Conf. Robot. Autom. (ICRA), May 2018, pp. 7159–7165.

[17] 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.

[18] A. Brock, J. Donahue, and K. Simonyan, “Large scale GAN training for high fidelity natural image synthesis,” 2018, arXiv:1809.11096. [Online]. Available: http://arxiv.org/abs/1809.11096

[19] H.-Y. Yang, P.-Y. Chen, C.-C. Huang, Y.-Z. Zhuang, and Y.-H. Shiau, “Low complexity underwater image enhancement based on dark channel prior,” in Proc. 2nd Int. Conf. Innov. Bio-Inspired Comput. Appl., Dec. 2011, pp. 17–20.

[20] P. Carletaris-Bianco, A. Mohan, and R. M. Eustice, “Initial results in underwater single image dehazing,” in Proc. OCEANS MTS/IEEE SEATTLE, Sep. 2010, pp. 1–8.

[21] Y.-T. Peng and P. C. Cosman, “Underwater image restoration based on image blurriness and light absorption,” IEEE Trans. Image Process., vol. 26, no. 4, pp. 1579–1594, Apr. 2017.

[22] 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.

[23] J. Yang, X. Wang, H. Yue, X. Fu, and C. Hou, “Underwater image enhancement based on structure-texture decomposition,” in Proc. IEEE Int. Conf. Image Process. (ICIP), Sep. 2017, pp. 1207–1211.

[24] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2009, pp. 248–255.

[25] R. Liu, X. Fan, M. Hou, Z. Jiang, Z. Luo, and L. Zhang, “Learning aggregated transmission propagation networks for haze removal and beyond,” IEEE Trans. Neural Netw. Learn. Syst., vol. 30, no. 10, pp. 2973–2986, Oct. 2019.

[26] R. Liu, X. Fan, M. Zhu, M. Hou, and Z. Luo, “Real-world underwater enhancement: Challenges, benchmarks, and solutions,” 2019, arXiv:1901.05320. [Online]. Available: http://arxiv.org/abs/1901.05320

[27] X. Fu, P. Zhuang, Y. Huang, Y. Liao, X.-P. Zhang, and X. Ding, “A retinex-based enhancing approach for single underwater image,” in Proc. IEEE Int. Conf. Image Process. (ICIP), Oct. 2014, pp. 4572–4576.

[28] C. Ancuti, C. O. Ancuti, T. Haber, and P. Beckaert, “Enhancing underwater images and videos by fusion,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2012, pp. 81–88.

[29] R. He, Z. Wang, H. Xiong, and D. D. Feng, “Single image dehazing with white balance correction and image decomposition,” in Proc. Int. Conf. Digit. Image Comput. Techn. Appl. (DICTA), Dec. 2012, pp. 1–7.

[30] P. L. J. Drews, E. R. Nascimento, S. S. C. Botelho, and M. F. Montenegro Campos, “Underwater depth estimation and image restoration based on single images,” IEEE Comput. Graph. Appl., vol. 36, no. 2, pp. 24–35, Mar. 2016.

[31] 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.

[32] D. Berman, D. Levy, S. Avidan, and T. Treibitz, “Underwater single image color restoration using haze-lines and a new quantitative dataset,” 2018, arXiv:1811.01343. [Online]. Available: http://arxiv.org/abs/1811.01343

[33] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao, “An underwater image enhancement benchmark dataset and beyond,” IEEE Trans. Image Process., vol. 29, pp. 4376–4389, 2020.
JIANHUA DENG received the M.Sc. and Ph.D. degrees in radio physics from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2003 and 2007, respectively. From 2001 to 2006, he was a Research Assistant with the Millimeter-wave Laboratory, Southwest Institute of Electronic Technology of China, Chengdu. In 2009, he was a Research Fellow with the Department of Electronic Engineering, City University of Hong Kong. His current research interests are in the areas of optimization, modeling, and computer-aided design for microwave and millimeter-wave integrated circuits and antennas.

JIAFU ZHUANG was born in Quanzhou, China, in 1987. He received the B.Sc. degree in communication engineering from Huaqiao University and the M.Sc. degree in pattern recognition and intelligent systems from the Huazhong University of Science and Technology. He currently works with the Department Artificial Intelligent and Robotics Technology, Haixi Institutes, Chinese Academy of Sciences. His research interests include 3D shape analysis, computer vision, and deep learning.

PEIDONG LIANG received the B.E. degree from the Henan University of Science and Technology, Luoyang, China, in 2009, and the M.S. and Ph.D. degrees from the Harbin Institute of Technology Harbin, China, in 2011 and 2017, respectively. He has been a CTO with the Fujian (Quanzhou) HIT Engineering and Technology Institute, since 2016. His research interests include robot engineering and factory automation.

PEIZHONG LIU received the B.E. degree from Huaqiao University, Fujian, China, in 1998, and the M.S. and Ph.D. degrees from Xiamen University, Xiamen, Fujian, China, in 2007 and 2010, respectively. Since 2011, he has been a Visiting Scholar with Duke University, USA. He is currently a Professor with the College of Engineering, and act as the Secretary General of the Neural Network and Computational Intelligence Committee of the Artificial Intelligence Society of China. His research interests include multi-dimensional space biomimetic informatics, visual media retrieval, network model, and information security.