Adaptive Channel Attention-Based Deformable Generative Adversarial Network for Underwater Image Enhancement

Tingkai Chen[1], Ning Wang[2(✉)], Xiangjun Kong[1], and Yanzheng Chen[2]

[1] School of Marine Electrical Engineering, Dalian Maritime University, Dalian 116026, China
[2] School of Marine Engineering, Dalian Maritime University, Dalian 116026, China
n.wang@ieee.org

Abstract. In this paper, to effectively strengthen quality of underwater image enhancement from both channel and spatial viewpoints, an adaptive channel attention-based deformable generative adversarial networks (ACADGAN) framework is established. Main contributions are as follows. 1) By virtue of multi-branch convolution architecture with dilated convolution mechanism, the adaptive channel attention (ACA) is devised, such that channel weight can be adaptively recalibrated, and thereby significantly contributing to preserving content features from channel viewpoint. 2) By augmenting offset position of sampling point with respect to convolution kernel, the deformable convolution network (DCN) is created, such that detailed information of underwater image can be dramatically retained from spatial aspect. 3) The ACADGAN scheme is eventually proposed by integrating ACA and DCN modules with a deep generative adversarial network. Comprehensive experiments demonstrate the remarkable effectiveness and superiority of the developed ACADGAN scheme.

Keywords: Underwater image enhancement · Generative adversarial network · Adaptive channel attention · Deformable convolution network

1 Introduction

With the rapid development of intelligence [40,43], intelligent remotely operated vehicle possessing underwater image enhancement (UIE) capability is highly desired for visually pleasing quality in terms of marine archaeology [32], ocean exploration [48], benthonic organism detection [8], *etc.* It should be mentioned

This work is supported by the National Natural Science Foundation of China (Grant 52271306), Innovative Research Foundation of Ship General Performance (Grant 31422120), and the Cultivation Program for the Excellent Doctoral Dissertation of Dalian Maritime University (Grant 2022YBPY004).

H. R. Karimi and N. Wang (Eds.): S-Cube 2022, LNICST 487, pp. 17–32, 2023.
https://doi.org/10.1007/978-3-031-34899-0_2

that selective wavelength-dependent attenuation can inevitably lead to bluish and/or greenish tones owing to light absorption and scattering effects. To some extent, the water body would present different degradation characteristics, such as low contrast, blurring detail, color distortion, *etc.* In this context, the UIE possessing strong generalization in multi-domain underwater environment is still an active field, which can dramatically promote the development of autonomous intelligent systems [41,42].

To address aforementioned underwater image degradation challenges, numerous state-of-the-art approaches have been innovatively proposed, which can be primarily divided into three families covering model-free enhancement, physical model-based restoration and data driven-based image-to-image conversion. Actually, the model-free enhancement can be effectively achieved by directly adjusting pixel without explicit model, whereby the enhancement can be realized in both space and frequency domains [15,22,23,37]. To be specific, the contrast limited adaptive histogram equalization (CLAHE) [54] and white balance [13] techniques have been developed by solving color distortion and low contrast issues in single color space. Besides, a two-step method has been proposed to conduct color correction and enhancement tasks in [16]. Moreover, by virtue of employing both bilateral and trilateral filtering techniques in each channel pertaining to CIELAB space, the multi-scale Retinex [51] framework has been devised to deal with image degradation. Deploying completely different concepts, the Fusion [4] scheme has been elaborately established via white balance and local histogram equalization strategies, whereby four representative weight maps with respect to laplacian contrast, local contrast, saliency and exposedness are utilized to conduct pixel-level fusion, and thereby contributing to presenting corresponding pixel with higher weight [3,38].

Compared to foregoing model-free enhancement works, the underwater image models [1,21,34] are sufficiently considered by model-based restoration approaches. In essence, the restoration is inversion solution to obtain potential underwater image. Concretely, to furthest increase visual quality, the dark channel prior (DCP)-based scheme has been proposed to estimate transmission map (TM) and background light [19,20,25]. Furthermore, the TM can be efficiently smoothed and optimized with the aid of median filter technique in [49]. Besides, in [9,10], the artificial light can be accurately estimated by determining intensity difference between background and foreground while the water depth can be estimated by computing residual energy ratio, respectively, such that the visual visibility can be tremendously reformed. It should be noted that, in [12], by comprehensively considering transformation difference between in-air and underwater scenes, the underwater DCP (UDCP) framework has been designed. In sequence, the phenomenon that the pixel value in blue channel is occasionally lower than that in red channel within complicated underwater environment has been intensively investigated, a dual channels-based scheme [31] is employed to estimate TM, such that the halo effect can be dramatically alleviated. Nevertheless, the foregoing existing physical model-based methods usually follow that the attenuation coefficients are only related to water body and keep consistent

among different channels, which can undoubtedly cause visual untruthfulness and unstable restoration, simultaneously.

Recently, with the high-speed prosperity of computational resource, the convolutional neural network (CNN)-based image-to-image conversion becomes available and feasible on large-scale synthetic and/or real datasets [24,39,44, 45,47]. Specifically, by training on simulated underwater scene constructed via combining air image and depth pairings, a two-stage WaterGAN [28] scheme has been devised, such that the color cast removal can be readily achieved. Similarly, by virtue of both revised underwater image formation model and scene parameters, the UWCNN [26] approach has been developed on the basis of synthetic underwater images. To eliminate the limitation of using paired underwater images, the Water CycleGAN [27] method has been proposed with the aid of cycle-consistency loss in a weakly supervised learning manner [53]. It should be noted that the training process of above-mentioned generative adversarial network (GAN) is extremely unstable. In the context, to dramatically increase training stability, the Wasserstein GAN [5] has been created by adequately computing distance between data and model distribution. Besides, the EBGAN [52], LSGANs [33] and conditional GAN [35] deploy the energy function, least-square loss function and restriction mechanism to address unstable training, respectively. However, accommodating multi-scale and geometric transformation for UIE task does not receive sufficient attention according to input information in an adaptive manner from both channel and spatial aspects.

To be specific, the receptive field of same convolutional processing layer usually shares same size, which could cause that multi-scale object information can hardly be efficiently captured at the same time from channel aspect [2,17,29]. In addition, preserving integrated information of underwater object with irregular transformation and unknown geometric scale would become impractical or infeasible due to inherently limited modeling mechanism in encoder structure from spatial side [11,46].

To circumvent foregoing weakness, an adaptive channel attention-based deformable generative adversarial network (ACADGAN) scheme has been devised for UIE in this paper. Significant highlights are summarized as follows:

- To effectively preserve image contents from channel perspective, the adaptive channel attention (ACA) is exclusively developed to adaptively recalibrate channel weight by virtue of multi-branch architecture using 3×3 and 5×5 convolutional kernels, simultaneously.
- To capture integrated information for object with irregular transformation and/or unknown geometric scale, the deformable convolution network (DCN) module is innovatively created, whereby offset position and modulation weight can be adaptively accommodated from the spatial aspect.
- Furthermore, combining ACA and DCN modules can contribute to entire ACADGAN scheme from both channel and spatial viewpoints, and thereby contributing to improving UIE quality.

The rest of paper is organized as follows. The problem formulation is presented in Sect. 2. Section 3 provides the developed ACADGAN scheme.

Experiment studies and comparisons are given in Sect. 4. Eventually, the conclusion is summarized in Sect. 5.

2 Problem Formulation

The classical underwater image formulation of Jaffe-McGlamery model [21] can be given by

$$\boldsymbol{U}_\lambda = \boldsymbol{I}_\lambda(x)e^{-\beta_\lambda \boldsymbol{d}(x)} + \boldsymbol{B}_\lambda(1 - e^{-\beta_\lambda \boldsymbol{d}(x)}) \tag{1}$$

where $\boldsymbol{U}$ is output image, $\lambda \in \{r, g, b\}$ denotes red, green and blue channels, respectively, $\boldsymbol{I}$ represents original input image, $\boldsymbol{d}$ is distance between camera and object, β denotes underwater scattering coefficient associated with water body, vertical depth and turbidity [6].

Notably, instead of employing foregoing underwater image formation, complex nonlinear mapping between degraded and enhanced images can be adaptively learned without explicit modeling, which can be formulated by

$$\boldsymbol{I} = \mathcal{F}(\boldsymbol{U}) \tag{2}$$

where $\mathcal{F}$ is an end-to-end multi-layer CNN with adversarial mechanism in an encoder-decoder pattern.

3 ACADGAN Scheme

In this section, the devised ACADGAN scheme is systematically presented, and is composed of three modules, i.e., ACA, DCN and GAN.

3.1 Adaptive Channel Attention

The ACA module is devised by recalibrating channel weight. Note that enhancing feature via employing convolutional kernel with different sizes from the channel aspect can dramatically improve visual quality. Accordingly, as shown in Fig. 1, the convolutional operation of utilizing 3×3 and 5×5 kernel can be given by

$$\boldsymbol{u}_k^\star = \boldsymbol{X} \otimes \boldsymbol{F}_k^\star = \sum_{i=1}^{c} \boldsymbol{x}^i \otimes \boldsymbol{f}_k^{i,\star} \tag{3}$$

where $\star \in \{3, 5\}$ is 3×3 and 5×5 convolutional filters, respectively, $\boldsymbol{x}^i \in \mathbb{R}^{w \times h}$, $\boldsymbol{X} = [\boldsymbol{x}^1, \boldsymbol{x}^2, \cdots, \boldsymbol{x}^c] \in \mathbb{R}^{c \times w \times h}$ denotes input feature, $\boldsymbol{f}_k^{i,\star} \in \mathbb{R}^{\star \times \star}$, $\boldsymbol{F}_k^\star = [\boldsymbol{f}_k^1, \boldsymbol{f}_k^2, \cdots, \boldsymbol{f}_k^c] \in \mathbb{R}^{c \times \star \times \star}$ refers to kernel weight of the k-th filter.

Subsequently, the fusion operation is formulated by

$$\boldsymbol{U} = \widetilde{\boldsymbol{U}} + \widehat{\boldsymbol{U}} \tag{4}$$

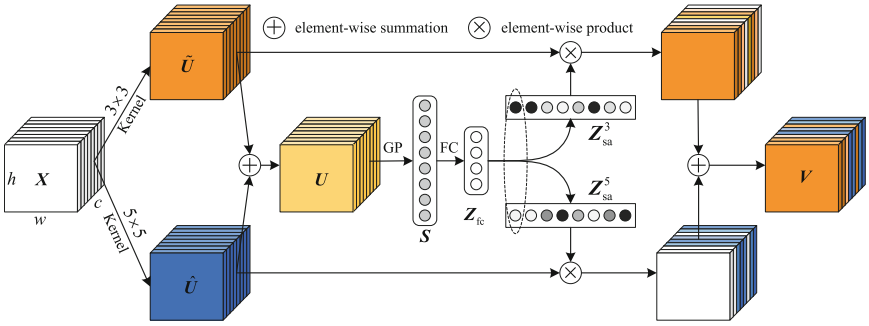


Fig. 1. Adaptive channel attention.

where $\widetilde{\boldsymbol{U}}$ and $\widehat{\boldsymbol{U}}$ represent convolutional output derived from previous layer by deploying 3×3 and 5×5 kernels, respectively, $\boldsymbol{U} \in \mathbb{R}^{c \times w \times h}$ is fusion output by element-wise summation operation.

Afterwards, the global average pooling is implemented to obtain representative response for each feature map, and can be derived by

$$s^i = \frac{1}{w \times h} \sum_{m=1}^{w} \sum_{n=1}^{h} \boldsymbol{u}^i(m,n), i = 1, 2, \cdots, c \tag{5}$$

where $\boldsymbol{S} = [s^1, s^2, \cdots, s^c] \in \mathbb{R}^{c \times 1}$, $\boldsymbol{U} = [\boldsymbol{u}^1, \boldsymbol{u}^2, \cdots, \boldsymbol{u}^c] \in \mathbb{R}^{c \times w \times h}$, w and h are width and height of feature map, respectively.

Besides, the single-hidden layer feedforward network is exploited to increase nonlinear fitting capability, and can be described by

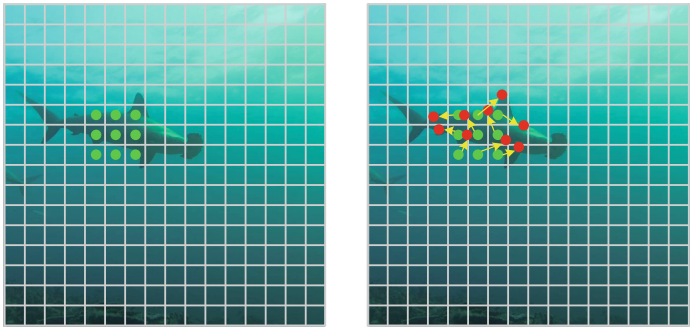
$$\boldsymbol{Z}_{\mathrm{fc}} = \mathrm{R}(\boldsymbol{w}_{\mathrm{fc}} \boldsymbol{S}) \tag{6}$$

where $\boldsymbol{Z}_{\mathrm{fc}} \in \mathbb{R}^{d \times 1}$ is output of fully connected layer, $\boldsymbol{w}_{\mathrm{fc}} \in \mathbb{R}^{d \times c}$ refers to corresponding weight, $d = \max(c/r, L)$, r denotes reduction ratio of channel dimension, L presents minimal value of d, R is Leaky ReLU activation function.

Furthermore, the soft attention associated with channel is adopted to adaptively select proper spatial scale according to feedback loss, and can be determined by

$$\begin{aligned} \boldsymbol{Z}_{\mathrm{sa}}^3 &= \frac{e^{\boldsymbol{w}_{\mathrm{sa}}^3 \boldsymbol{Z}_{\mathrm{fc}}}}{e^{\boldsymbol{w}_{\mathrm{sa}}^3 \boldsymbol{Z}_{\mathrm{fc}}} + e^{\boldsymbol{w}_{\mathrm{sa}}^5 \boldsymbol{Z}_{\mathrm{fc}}}} \\ \boldsymbol{Z}_{\mathrm{sa}}^5 &= \frac{e^{\boldsymbol{w}_{\mathrm{sa}}^5 \boldsymbol{Z}_{\mathrm{fc}}}}{e^{\boldsymbol{w}_{\mathrm{sa}}^3 \boldsymbol{Z}_{\mathrm{fc}}} + e^{\boldsymbol{w}_{\mathrm{sa}}^5 \boldsymbol{Z}_{\mathrm{fc}}}} \end{aligned} \tag{7}$$

where $\boldsymbol{w}_{\mathrm{sa}}^3 \in \mathbb{R}^{c \times d}$ and $\boldsymbol{w}_{\mathrm{sa}}^5 \in \mathbb{R}^{c \times d}$ are weights of soft attention, $\boldsymbol{Z}_{\mathrm{sa}}^3 \in \mathbb{R}^{c \times 1}$ and $\boldsymbol{Z}_{\mathrm{sa}}^5 \in \mathbb{R}^{c \times 1}$ are outputs of soft attention for 3×3 and 5×5 convolutional branches, respectively.



(a) (b)

Fig. 2. Deformable convolution network. (a) Standard sampling positions (green points). (b) Deformed sampling positions (red points). (Color figure online)

Eventually, the recalibrated feature map $\boldsymbol{V}$ is defined by

$$\boldsymbol{V} = \boldsymbol{Z}_{\mathrm{sa}}^{3} \cdot \widetilde{\boldsymbol{U}} + \boldsymbol{Z}_{\mathrm{sa}}^{5} \cdot \widehat{\boldsymbol{U}} \tag{8}$$

where $\boldsymbol{V} \in \mathbb{R}^{c \times w \times h}$ represents final feature map by fusing two attention maps.

Remark 1. Fusing multiple convolutional outputs from different branches using 3×3 and 5×5 kernels, appropriate receptive field pertaining to involved object with different scales can be adaptively selected in encoder stage.

Remark 2. In order to furthest decrease the number of parameters and keep consistent receptive field characteristic, simultaneously, by using 3×3 convolutional kernel with dilation philosophy, the foregoing 5×5 convolution is effectively achieved.

3.2 Deformable Convolution Network

The DCN module is achieved by adaptively adjusting fixed sampling position of convolution kernel, which is conductive to establishing integrated feature descriptor for object with unknown transformation and/or large geometric dimension from the spatial viewpoint. As shown in Fig. 2, the sampling positions of standard convolutional kernel can be represented by

$$P = \{(u, v) \mid u, v \in \{-k, -(k-1), \cdots, k\}\} \tag{9}$$

where $k = |\frac{w}{2}|_{\mathrm{odd}}$, and w refers to width of current convolutional kernel.

Accordingly, the deformed sampling position can be given by

$$\widetilde{p}_m = p_0 + p_m + \Delta p_m \tag{10}$$

where p_0 denotes the center point of convolutional kernel, $p_m = (x_m, y_m)$ is the m-th standard sampling position, $m = 1, 2, \cdots, n$, $n = |P|$, Δp_m represents offset

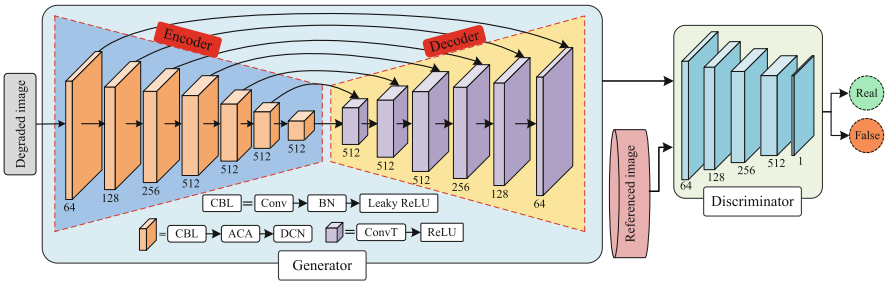


Fig. 3. The entire ACADGAN framework.

of the m-th element within convolutional kernel, $\widetilde{p}_m = (\widetilde{x}_m, \widetilde{y}_m)$ is deformed sampling position.

Note that the deformed sampling position should satisfy

$$\begin{aligned} 0 \leqslant \widetilde{x}_m \leqslant w \\ 0 \leqslant \widetilde{y}_m \leqslant h \end{aligned} \tag{11}$$

Accordingly, the bilinear interpolation strategy is used to compute feature value of non-integer position from X direction, and can be derived by

$$\boldsymbol{F}(\widetilde{x}_m, y_j) = \frac{\widetilde{x}_m - x_i}{x_{i+1} - x_i}\boldsymbol{F}(x_{i+1}, y_j) + \frac{x_{i+1} - \widetilde{x}_m}{x_{i+1} - x_i}\boldsymbol{F}(x_i, y_j) \tag{12}$$

where $p_{\mathrm{tl}} = (x_i, y_j)$, $p_{\mathrm{tr}} = (x_{i+1}, y_j)$, $p_{\mathrm{bl}} = (x_i, y_{j+1})$, $p_{\mathrm{br}} = (x_{i+1}, y_{j+1})$ are top-left, top-right, bottom-left and bottom-right integer coordinate positions which are closest to the deformed sampling position, respectively.

Similarly, the feature value $\boldsymbol{F}(\widetilde{x}_m, y_{j+1})$ can be described by

$$\boldsymbol{F}(\widetilde{x}_m, y_{j+1}) = \frac{\widetilde{x}_m - x_i}{x_{i+1} - x_i}\boldsymbol{F}(x_{i+1}, y_{j+1}) + \frac{x_{i+1} - \widetilde{x}_m}{x_{i+1} - x_i}\boldsymbol{F}(x_i, y_{j+1}) \tag{13}$$

Furthermore, the feature value $\boldsymbol{F}(\widetilde{x}_m, \widetilde{y}_m)$ is given by

$$\boldsymbol{F}(\widetilde{x}_m, \widetilde{y}_m) = \frac{\widetilde{y}_m - y_j}{y_{j+1} - y_j}\boldsymbol{F}(\widetilde{x}_m, y_{j+1}) + \frac{y_{j+1} - \widetilde{y}_m}{y_{j+1} - y_j}\boldsymbol{F}(\widetilde{x}_m, y_j) \tag{14}$$

where $\boldsymbol{F}(\widetilde{x}_m, \widetilde{y}_m)$ is the feature value in deformed position $(\widetilde{x}_m, \widetilde{y}_m)$.

Ultimately, the deformable output can be formulated as follows:

$$\boldsymbol{O}(x_0, y_0) = \sum_{m=1}^{n} \boldsymbol{w}_m^{\mathrm{F}} \boldsymbol{F}(\widetilde{x}_m, \widetilde{y}_m)\sigma(\boldsymbol{w}_m) \tag{15}$$

where $\boldsymbol{O}(x_0, y_0)$ is eventual output after performing deformation convolution operation corresponding to position (x_0, y_0), $\boldsymbol{w}_m^{\mathrm{F}}$ denotes the m-th element weight within convolutional kernel, $\sigma(\boldsymbol{w}_m)$ refers to modulation operation utilized to emphasize significance of the m-th offset position and $\boldsymbol{w}_m$ is constrained within the range (0,1).

Table 1. The objective image quality evaluation comparisons on UIEB dataset.

Scheme	SSIM	PSNR	UIQM	UICM	UISM	UIConM	UCIQE
IBLA	0.6505	15.2930	1.9546	6.6576	4.4284	0.1285	0.4902
UDCP	0.4433	10.7593	1.6257	5.6709	4.4114	0.0456	**0.4913**
CLAHE	0.7119	**17.6431**	2.7026	5.0877	5.4467	0.2261	0.4170
UCM	0.7344	16.3005	2.5844	**7.2730**	5.5500	0.2071	0.4861
FUnIE-GAN	0.6593	15.2940	**2.7248**	5.0241	**6.1299**	0.2162	0.4351
UWCNN	0.7112	15.5927	2.2504	3.8042	4.5116	0.2267	0.3847
ACADGAN	**0.8097**	16.3171	2.7191	4.4317	5.9843	**0.2711**	0.4007

Table 2. The objective image quality evaluation comparisons on URPC dataset.

Scheme	UIQM	UICM	UISM	UIConM	UCIQE
IBLA	2.3393	3.6788	4.2001	0.2783	0.375
UDCP	1.7634	**4.4328**	4.0194	0.1263	0.3995
CLAHE	2.5963	3.1805	4.8888	0.2808	0.3456
UCM	2.5847	3.5418	4.8931	0.2908	0.4054
FUnIE-GAN	2.6713	2.2178	**5.1713**	0.3025	0.3525
UWCNN	2.2726	1.5492	3.4277	**0.3403**	0.3022
ACADGAN	**2.7301**	2.8468	5.1205	0.3347	**0.4517**

3.3 Generative Adversarial Network

As shown in Fig. 3, the designed GAN is divided into generator and discriminator, whereby ACA and DCN modules are organically integrated into the encoder framework with respect to generator. Besides, the WGAN-GP loss, $L1$ loss and image gradient difference loss are combined to implement adversarial training [14].

Remark 3. By combining ACA and DCN modules within encoder structure of generator in the sequential manner, the proposed ACADGAN scheme with residual mechanism can be eventually developed by integrating discriminator, and thereby contributing to dramatically avoiding vanishing gradient and promoting to generate high-quality underwater image, simultaneously.

4 Experimental Studies and Comparisons

In order to sufficiently investigate the effectiveness and superiority of the proposed ACADGAN scheme, the physical model-based restoration approaches, model-free enhancement methods and image-to-image conversion techniques are

Table 3. The Feature Expression Comparisons on UIEB Dataset.

Scheme	SIFT	Harris	Canny
IBLA	1381.4444	652.9444	0.0763
UDCP	1030.6778	501.4889	0.0588
CLAHE	1908.6667	**822.6444**	0.1022
UCM	1804.6000	792.3000	0.1046
FUnIE-GAN	1615.9444	631.0444	0.0656
UWCNN	803.7778	498.5333	0.0456
ACADGAN	**1991.4333**	522.0778	**0.1060**

Table 4. The Feature Expression Comparisons on URPC Dataset.

Scheme	SIFT	Harris	Canny
IBLA	1106.5333	241.4368	0.0317
UDCP	738.9759	250.4428	0.0258
CLAHE	1198.1516	356.7569	0.0528
UCM	**1409.0626**	330.9964	**0.0608**
FUnIE-GAN	555.9627	210.1793	0.0273
UWCNN	260.8893	94.3755	0.0106
ACADGAN	1398.9302	**358.7016**	0.0575

employed to implement comparisons under typical underwater challenging situations covering artificial light (first column), bluish light (second column), greenish light (third column), low light (fourth column) and violet light (fifth column). As shown in Fig. 4, from which we can clearly observe that the physical model-based image restoration approaches can hardly achieve satisfactory visual performance. To be specific, the UDCP approach can bring in more serious aggravation of color deviation effect in aforementioned scenarios. Obviously, the model-free-based underwater image enhancement approach, i.e., UCM, can tend to introduce reddish color cast, and thereby resulting in enhancement failure by directly adjusting pixel value. It should be noted that the FUnIe-GAN and UWCNN schemes can only obtain pretty limited enhancement effect. Meanwhile, the FUnIE-GAN method can cause apparent color deviation under violet-light situation. Intuitively, the devised ACADGAN framework can ultimately achieve the optimal visual enhancement quality by employing ACA and DCN modules in a sequential manner. Furthermore, the consistent conclusion can also be summarized from Fig. 5.

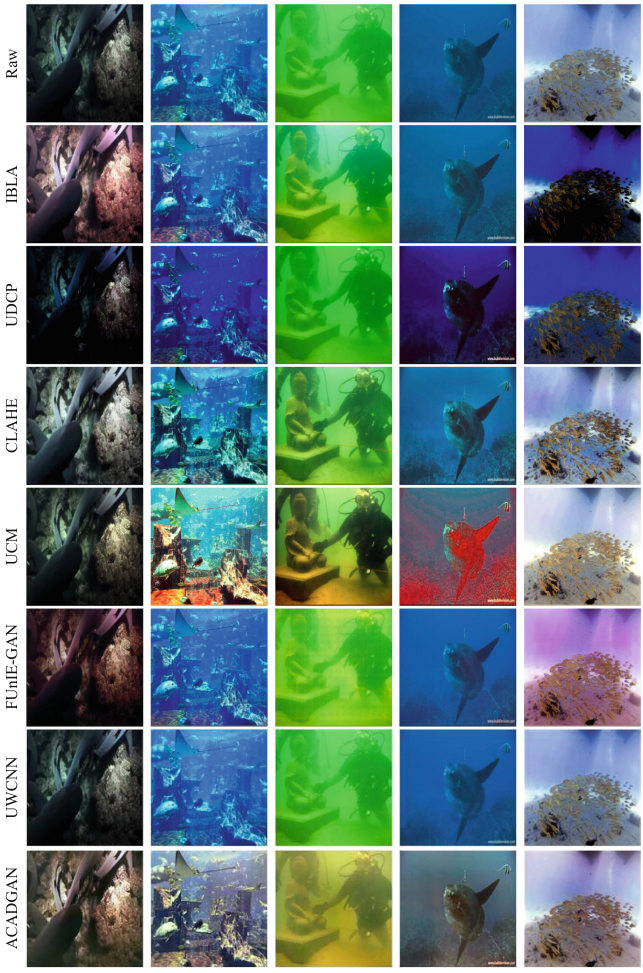


Fig. 4. The subjectively visual comparisons on UIEB dataset.

4.1 Quantitative Comparisons

To further sufficiently demonstrate superiority of the developed ACADGAN scheme, the peak signal to noise ratio (PSNR), structural similarity (SSIM), underwater color image quality evaluation (UCIQE) [50], underwater image quality measures (UIQM) related to underwater image colorfulness, sharpness, contrast (i.e., UICM, UISM and UIConM) [36] indices are comprehensively considered, and the comparison results are summarized in Table 1 and Table 2 derived from average value of 90 and 831 test images on UIEB and URPC datasets, respectively, from which we can readily obtain that the proposed ACADGAN scheme can achieve the optimal (highlighted in bold) or sub-optimal (highlighted in underscore) performances on most metrics. Concretely, on the one hand, the

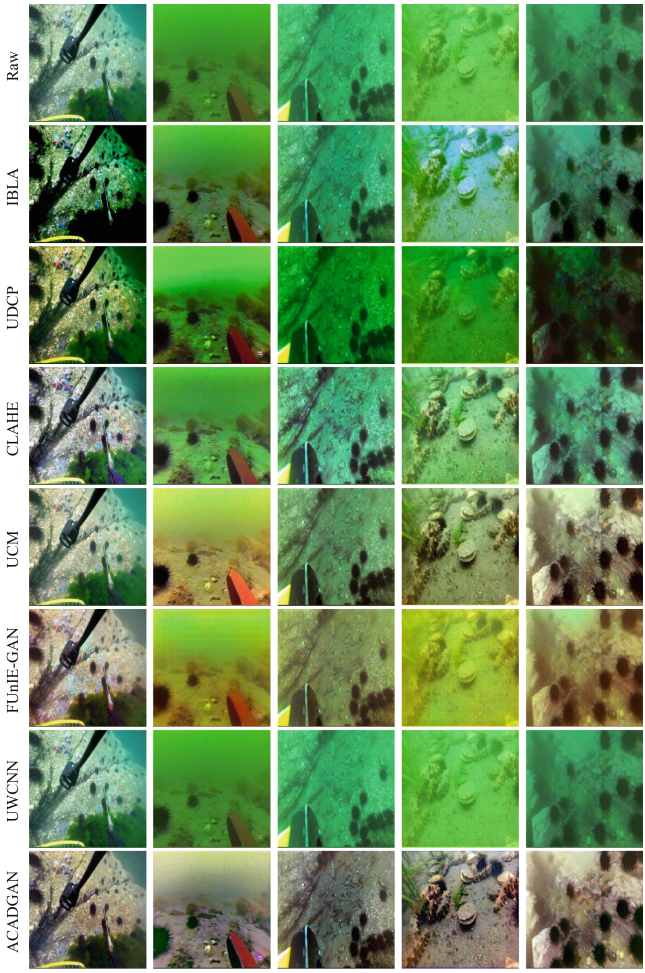


Fig. 5. The subjectively visual comparisons on URPC dataset.

image structure, texture and contrast are effectively preserved according to the optimal SSIM and UIConM metrics on UIEB dataset, simultaneously. More ambitiously, the proposed ACADGAN scheme can perfectly accomplish preservation in terms of the image content. Additionally, in terms of evaluation on URPC dataset, the developed ACADGAN scheme achieves the optimal UIQM score, which implies that the enhanced images are more consistent with human visual perception. More importantly, the proposed ACADGAN scheme can realize better balance in terms of chroma, saturation and contrast on the basis of the optimal UCIQE score.

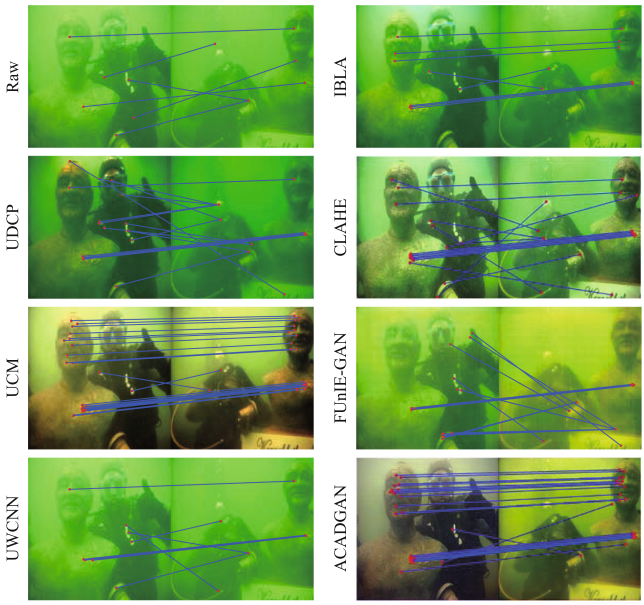


Fig. 6. The keypoint matching comparisons.

4.2 Feature Expression Comparisons

In order to verify effectiveness and superiority of the devised ACADGAN scheme from the perspective of fundamental feature expression, the SIFT [30], Harris [18] and Canny [7] approaches are utilized to extract key point, corner and pixel-level edge ratio. Correspondingly, the key point matching result is shown in Fig. 6, from which we can clearly see that few key points can be correctly extracted and matched on raw underwater image due to the poor quality. Meanwhile, the key point matching performance can be enhanced by employing restoration or enhancement approaches. Furthermore, the proposed ACADGAN scheme can achieve the optimal keypoint matching performance, which implies that the proposed ACADGAN scheme can make degraded underwater image restore more fundamental features. It should be emphasized that the average performance on UIEB and URPC datasets are provided in Table 3 and Table 4, from which we can obviously observe that the proposed ACADGAN scheme can achieve optimal or sub-optimal performance in terms of extraction of SIFT keypoint, Harris corner and Canny edge, which indicates the proposed ACADGAN scheme can contribute to extraction of fundamental features.

5 Conclusion

In this paper, an ACADGAN framework has been exclusively devised for UIE task. To be specific, the ACA module has been created to preserve content feature from channel perspective. In addition, the DCN module has been developed to augment sampling capability pertaining to convolutional kernel. Integrating ACA and DCN modules can eventually contribute to the proposed ACADGAN scheme. Comprehensive experiments and comparisons have thoroughly demonstrated that the proposed ACADGAN scheme can remarkably outperform plenty of state-of-the-art UIE approaches.

References

1. Akkaynak, D., Treibitz, T.: A revised underwater image formation model. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, pp. 6723–6732 (2018)
2. Alipour-Fard, T., Paoletti, M., Haut, J.M., Arefi, H., Plaza, J., Plaza, A.: Multibranch selective kernel networks for hyperspectral image classification. IEEE Geosci. Remote Sens. Lett. **18**(6), 1089–1093 (2020)
3. Ancuti, C.O., Ancuti, C., Bekaert, P.: Effective single image dehazing by fusion. In: International Conference on Image Processing, Hong Kong, China, pp. 3541–3544 (2010)
4. Ancuti, C., Ancuti, C.O., Haber, T., Bekaert, P.: Enhancing underwater images and videos by fusion. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, pp. 81–88 (2012)
5. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein generative adversarial networks. In: Proceedings of the International Conference on Machine Learning, Sydney, Australia, pp. 214–223 (2017)
6. Berman, D., Treibitz, T., Avidan, S.: Diving into haze-lines: color restoration of underwater images. In: Proceedings of the British Machine Vision Conference, London, UK, vol. 1, pp. 1–12 (2017)
7. Canny, J.: A computational approach to edge detection. IEEE Trans. Pattern Anal. Mach. Intell. **8**(6), 679–698 (1986)
8. Chen, T., Wang, N., Wang, R., Zhao, H., Zhang, G.: One-stage CNN detector-based benthonic organisms detection with limited training dataset. Neural Netw. **144**, 247–259 (2021)
9. Chiang, J.Y., Chen, Y.C.: Underwater image enhancement by wavelength compensation and dehazing. IEEE Trans. Image Process. **21**(4), 1756–1769 (2011)
10. Chiang, J.Y., Chen, Y.C., Chen, Y.F.: Underwater image enhancement: using wavelength compensation and image dehazing (WCID). In: International Conference on Advanced Concepts for Intelligent Vision Systems, Ghent, Belgium, pp. 372–383 (2011)
11. Dai, J., et al.: Deformable convolutional networks. In: Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, pp. 764–773 (2017)
12. Drews, P., Nascimento, E., Moraes, F., Botelho, S., Campos, M.: Transmission estimation in underwater single images. In: Proceedings of the IEEE International Conference on Computer Vision Workshops, Sydney, Australia, pp. 825–830 (2013)
13. Ebner, D.H.: Color constancy. Vis. Res. **51**(7), 674–700 (2011)

14. Fabbri, C., Islam, M.J., Sattar, J.: Enhancing underwater imagery using generative adversarial networks. In: International Conference on Robotics and Automation, Brisbane, QLD, Australia, pp. 7159–7165 (2018)
15. Feifei, S., Xuemeng, Z., Guoyu, W.: An approach for underwater image denoising via wavelet decomposition and high-pass filter. In: International Conference on Intelligent Computation Technology and Automation, Shenzhen, China, vol. 2, pp. 417–420 (2011)
16. Fu, X., Fan, Z., Ling, M., Huang, Y., Ding, X.: Two-step approach for single underwater image enhancement. In: International Symposium on Intelligent Signal Processing and Communication Systems, Xiamen, China, pp. 789–794 (2017)
17. Gao, W., Zhang, L., Huang, W., Min, F., He, J., Song, A.: Deep neural networks for sensor-based human activity recognition using selective kernel convolution. IEEE Trans. Instrum. Meas. **70**, 1–13 (2021)
18. Harris, C., Stephens, M., et al.: A combined corner and edge detector. In: Proceedings of the Alvey Vision Conference, vol. 15, pp. 10–5244. Citeseer (1988)
19. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. In: Proceedings of the IEEE Conference on Computer Vision Pattern Recognition, Miami Beach, FL, USA, pp. 1956–1963 (2009)
20. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. IEEE Trans. Pattern Anal. Mach. Intell. **33**(12), 2341–2353 (2010)
21. Jaffe, J.S.: Computer modeling and the design of optimal underwater imaging systems. IEEE J. Oceanic Eng. **15**(2), 101–111 (1990)
22. Jia, D., Ge, Y.: Underwater image de-noising algorithm based on nonsubsampled contourlet transform and total variation. In: International Conference on Computer Science and Information Processing, Xi'an, China, pp. 76–80 (2012)
23. Jian, S., Wen, W.: Study on underwater image denoising algorithm based on wavelet transform. In: International Conference on Control Engineering and Artificial Intelligence, Kuala Lumpur, Malaysia, vol. 806, p. 012006 (2017)
24. Jing, Y., Yang, Y., Feng, Z., Ye, J., Yu, Y., Song, M.: Neural style transfer: a review. IEEE Trans. Vis. Comput. Graph. **26**(11), 3365–3385 (2019)
25. Lee, S., Yun, S., Nam, J.H., Won, C.S., Jung, S.W.: A review on dark channel prior based image dehazing algorithms. EURASIP J. Image Video Process. **2016**(1), 4–26 (2016)
26. Li, C., Anwar, S., Porikli, F.: Underwater scene prior inspired deep underwater image and video enhancement. Pattern Recogn. **98**, 107038 (2020)
27. Li, C., Guo, J., Guo, C.: Emerging from water: underwater image color correction based on weakly supervised color transfer. IEEE Signal. Proc. Lett. **25**(3), 323–327 (2018)
28. Li, J., Skinner, K.A., Eustice, R.M., Johnson-Roberson, M.: WaterGAN: unsupervised generative network to enable real-time color correction of monocular underwater images. IEEE Robot. Autom. Lett. **3**(1), 387–394 (2017)
29. Li, X., Wang, W., Hu, X., Yang, J.: Selective kernel networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, pp. 510–519 (2019)
30. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
31. Lu, H., Li, Y., Zhang, L., Serikawa, S.: Contrast enhancement for images in turbid water. J. Opt. Soc. Am. A-Opt. Image Sci. **32**(5), 886–893 (2015)
32. Ludvigsen, M., Sortland, B., Johnsen, G., Singh, H.: Applications of geo-referenced underwater photo mosaics in marine biology and archaeology. Oceanography **20**(4), 140–149 (2007)

33. Mao, X., Li, Q., Xie, H., Lau, R.Y., Wang, Z., Paul Smolley, S.: Least squares generative adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, pp. 2794–2802 (2017)
34. McGlamery, B.: A computer model for underwater camera systems. In: Ocean Optics, Monterey, CA, USA, vol. 208, pp. 221–231 (1980)
35. Mirza, M., Osindero, S.: Conditional generative adversarial nets (2014). https://arxiv.org/abs/1411.1784
36. Panetta, K., Gao, C., Agaian, S.: Human-visual-system-inspired underwater image quality measures. IEEE J. Ocean. Eng. **41**(3), 541–551 (2015)
37. Priyadharsini, R., Sharmila, T.S., Rajendran, V.: A wavelet transform based contrast enhancement method for underwater acoustic images. Multidimension. Syst. Signal Process. **29**(4), 1845–1859 (2018)
38. Singh, R., Biswas, M.: Adaptive histogram equalization based fusion technique for hazy underwater image enhancement. In: IEEE International Conference on Computational Intelligence and Computing Research, Chennai, India, pp. 1–5 (2016)
39. Wang, N., Er, M.J.: Self-constructing adaptive robust fuzzy neural tracking control of surface vehicles with uncertainties and unknown disturbances. IEEE Trans. Control Syst. Technol. **23**(3), 991–1002 (2014)
40. Wang, N., Er, M.J.: Direct adaptive fuzzy tracking control of marine vehicles with fully unknown parametric dynamics and uncertainties. IEEE Trans. Control Syst. Technol. **24**(5), 1845–1852 (2016)
41. Wang, N., Er, M.J., Sun, J.C., Liu, Y.C.: Adaptive robust online constructive fuzzy control of a complex surface vehicle system. IEEE Trans. Cybern. **46**(7), 1511–1523 (2015)
42. Wang, N., Karimi, H.R., Li, H., Su, S.F.: Accurate trajectory tracking of disturbed surface vehicles: a finite-time control approach. IEEE/ASME Trans. Mechatron. **24**(3), 1064–1074 (2019)
43. Wang, N., Qian, C., Sun, J.C., Liu, Y.C.: Adaptive robust finite-time trajectory tracking control of fully actuated marine surface vehicles. IEEE Trans. Control Syst. Technol. **24**(4), 1454–1462 (2015)
44. Wang, N., Wang, Y., Er, M.J.: Review on deep learning techniques for marine object recognition: architectures and algorithms. Control. Eng. Pract. **118**, 104458 (2022)
45. Wang, W., Wu, X., Yuan, X., Gao, Z.: An experiment-based review of low-light image enhancement methods. IEEE Access **8**, 87884–87917 (2020)
46. Wang, X., Chan, K.C., Yu, K., Dong, C., Change Loy, C.: EDVR: video restoration with enhanced deformable convolutional networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, pp. 1954–1963 (2019)
47. Wang, Z., Chen, J., Hoi, S.C.: Deep learning for image super-resolution: a survey. IEEE Trans. Pattern Anal. Mach. Intell. **43**(10), 3365–3387 (2020)
48. Whitcomb, L., Yoerger, D.R., Singh, H., Howland, J.: Advances in underwater robot vehicles for deep ocean exploration: navigation, control, and survey operations. In: Hollerbach, J.M., Koditschek, D.E. (eds.) Robotics Research, pp. 439–448. Springer, London (2000). https://doi.org/10.1007/978-1-4471-0765-1_53
49. Yang, H., Chen, P., Huang, C., Zhuang, Y., Shiau, Y.: Low complexity underwater image enhancement based on dark channel prior. In: International Conference on Innovations in Bio-inspired Computing and Applications, Shenzhen, China, pp. 17–20 (2011)
50. Yang, M., Sowmya, A.: An underwater color image quality evaluation metric. IEEE Trans. Image Process. **24**(12), 6062–6071 (2015)

51. Zhang, S., Wang, T., Dong, J., Yu, H.: Underwater image enhancement via extended multi-scale Retinex. Neurocomputing **245**, 1–9 (2017)
52. Zhao, J., Mathieu, M., LeCun, Y.: Energy-based generative adversarial network (2016). https://arxiv.org/abs/1609.03126
53. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, pp. 2223–2232 (2017)
54. Zuiderveld, K.: Contrast limited adaptive histogram equalization. In: Graphics Gems, pp. 474–485 (1994). https://doi.org/10.1016/B978-0-12-336156-1.50061-6