

# Underwater image enhancement via efficient generative adversarial network

XIN QIAN<sup>1</sup>, PENG GE<sup>1,2\*</sup>

<sup>1</sup>School of Electronics and Information Engineering, South China University of Technology, 510640 Guangzhou, China

<sup>2</sup>School of Physics and Opto-electronic, South China University of Technology, 510640 Guangzhou, China

\*Corresponding author: scpge@scut.edu.cn

Underwater image enhancement has been receiving much attention due to its significance in facilitating various marine explorations. Inspired by the generative adversarial network (GAN) and residual network (ResNet) in many vision tasks, we propose a simplified designed ResNet model based on GAN called efficient GAN (EGAN) for underwater image enhancement. In particular, for the generator of EGAN we design a new pair of convolutional kernel size for the residual block in the ResNet. Secondly, we abandon batch normalization (BN) after every convolution layer for faster training and less artifacts. Finally, a smooth loss function is introduced for halo-effect alleviation. Extensive qualitative and quantitative experiments show that our methods accomplish considerable improvements compared to the state-of-the-art methods.

Keywords: underwater, image dehazing, generative adversarial network (GAN).

## 1. Introduction

Recently, growing research attention has been focused on underwater image enhancement for its fundamental role in multiple sea world applications [1, 2]. As underwater environments are affected by many factors, it is a challenging task for the real-world underwater image enhancement. Conventionally, underwater image is distorted by light scattering and wavelength-dependent absorption [3–5]. These negative effects result in low contrast, dim visibility, and color shift in the obtained images which pose considerable obstacles to the existing algorithms in achieving demanding results. For underwater scene, severe light scattering come from the suspended multisize particles, resulting in backward and forward light propagation direction changing. Light absorption appears due to the fact that the longer the light wavelength is, the faster it is absorbed. Thus the energy of red light reduces faster than the blue and green light. Based

on that, the underwater scenes usually appear in bluish or greenish tone. The reasons behind those negative phenomena are various: operation depth, water temperature, and weather conditions, to name a few. Because of that, it is difficult to explicitly model the distortion conditions, and restoring natural colors and fine details for distorted underwater image still remain an open problem [6]. In the past few years, many inspiring methods tackle this issue based on the image formation model (IFM). These include the methods based on: dark channel prior (DCP) [7], red channel prior [8], and haze-line [9]. In [10], LU *et al.* used color lines to estimate background light and DCP to estimate transmission map. HAN *et al.* [11], started from estimating the point-spread-function using slant-edge-method. And PUROHIT *et al.* [12], proposed a weighted combination of multilevel image details to restore image texture. Specially the fusion strategy [13], proposed by ANCUTI *et al.* utilized the color corrected version and contrast enhanced version of the raw underwater image to produce the enhanced result. In [14], PENG *et al.* proposed to estimate scene depth based on image blurriness and light absorption. SONG *et al.* [15] proposed another way for scene depth estimation using underwater light attenuation prior. Although those methods could produce satisfactory results in some cases, their generalization ability remains uncertain.

In the last several years, the convolution neural network (CNN) has proved its great effectiveness in many computer vision tasks. Hence there are increasing interests on CNN-based underwater image enhancement. In [16], ANWAR *et al.* utilized a synthetic underwater image database to train a convolutional neural network (UWCNN) and proved its generality in a variety of underwater scenes. In [17], LI *et al.* used the fusion strategy [13], to pre-process raw underwater images first, then put them into a trained CNN model to get the enhanced result. Apart from that, as generative adversarial network (GAN) was making remarkable progress in single image super resolution [18], the atmospheric image dehazing [19], and image translation [20], its practicability was also brought into underwater scenarios. In [21], LIU *et al.* proposed a conditional GAN model in which a multiscale feature extraction manner was introduced to form the generator. In [22], LI *et al.* proposed a WaterGAN to generate synthetic sea-world images, then use those data to train a two-stage network for underwater image color correction. FABBRI *et al.* [23], first used CycleGAN to produce the restored underwater image as training data, then an Underwater-GAN was constructed to learn a model that translates a haze image into a high resolution image.

Generally speaking, both IFM-based solutions and CNN-based solutions have feature respective merits as well as several limitations. For methods in [13], and [14], prior knowledge may fail in some cases, resulting in incorrect color restoration and image detail losses. The process models built in [17, 21, 22] are relatively complicated, which is inconvenient in practical use. Based on experimental comparison, we augment that a relatively slim-build network could accomplish both fast processing and satisfying results at the same time.

In this paper, we design a deep-learning-based model call efficient GAN (EGAN) for fast and good performance in underwater image enhancement. In particular, we first

adopt the residual network (ResNet) as the backbone of our generator, then we improve the generator architecture by introducing two adjustments: removing batch normalization (BN) and replacing convolution (Conv) kernel sizes. Those adjustments are proved in training resource saving and image quality improving, which are demonstrated by graphic and statistic measurements in the following contents. Besides, as a halo-effect is found to be one of annoying collateral issues coming with deep-learning-based methods, we introduce a smooth loss function to make the artifacts less visible in output images. Extensive qualitative and quantitative comparison experiments are conducted with the state-of-the-art methods, showing the competitiveness of our method.

## 2. Proposed method

A common feature shared by the CNN and the GAN is the ability to learn complex nonlinear functions. Different from CNN, GAN consists of two separated CNNs to form an adversarial architecture, as this manner pushes each of the CNN network to improve until the whole model reaches the point of Nash equilibrium. Specifically, there are generator network (G) and discriminator network (D) in GAN, G aims to transform an input (raw image) into a target (enhanced image) that is close to the ground truth (reference image), while D aims to judge whether an input is the ground truth or the enhanced image. As shown in Fig. 1, the total structure of our EGAN consists of two parts: a generator and a discriminator. The generator takes raw image as an input and outputs the according enhanced result, which serves as the major part in the EGAN image process procedure. Meanwhile the discriminator takes the enhanced image and the reference image as an input, and it outputs their possible distribution. The possible distribution tells whether an enhanced image is likely to be judged as a reference image or raw image.

### 2.1. Generator structure

Usually for many deep-learning-based vision tasks, more Conv layers lead to more parameters to be predicted during training, and, as a result, the trained model could simulate complicated nonlinear functions, for example the underwater image distortion model. However, stacking Conv layer is not always a proper solution, for vanishing emergence of gradients causing model degradation the training accuracy saturates and the error rises [24]. To tackle this issue, He *et al.* [25] proposed deep residual learning that introduced a residual connection between Conv layers. Benefiting from that, backward propagation gradients could spread a longer path through the model, and deeper CNN structure becomes practically feasible. Inspired by that, we designed our generator (G) as shown in Fig. 1.

Firstly, the model takes a patch ( $64 \times 64 \times 3$ ) of a raw underwater image as an input, followed by Conv layer with  $3 \times 3$  kernel size and 64 output feature channels. Then

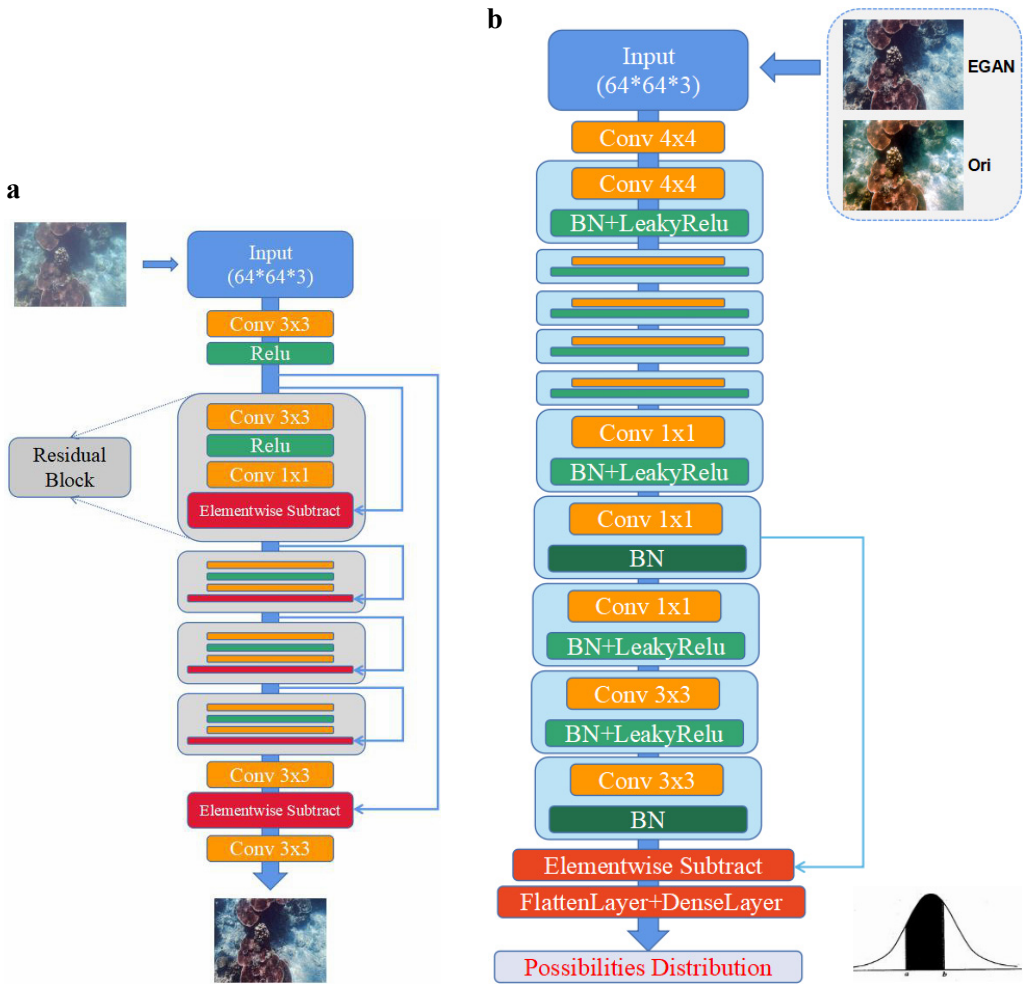


Fig. 1. Structures of our EGAN network. (a) Generator, and (b) discriminator.

those features are activated by rectified linear unit (ReLU). For the backbone design we borrow the layout from [19], which is first adopted for in-air dehazing. Then three adaptations are made for underwater enhancement. Firstly, we implement 4 residual blocks instead of 16 as originally designed, because we found that less layers could achieve equivalent or even exceeded performance (more details shown in Section 4). Every residual block consists of two Conv layers and one element-wise subtract layer, with one ReLU after the first Conv. All the Conv layer outputs feature a map with 64 channels. Secondly, we replace the second Conv kernel size of  $3 \times 3$  with  $1 \times 1$ , based on the knowledge that small kernel size carries less parameters, hence resulting in faster training. Besides, as feature maps are recurrently computed between residual

blocks, smaller kernel size could also favour a small image texture extraction and lead to less halo-effect production (more details shown in Section 4). Our last adaptation is removing all the BN layers after every Conv layer in the whole generator structure, because, as LIM *et al.* [26] proved, for the relatively shallow model architecture removing BN could save training time as well as promote model performance (for example increase contrast as shown in Section 4).

Different from many U-Net [27] based methods, our generator could translate image of different sizes. This merit makes the whole network more suitable for practical application. In our EGAN generator, all Conv layers have fixed output channel number of 64, except the last one which has 3 to form a normal picture.

## 2.2. Discriminator structure

The adversarial learning procedure of GAN requires another CNN model called discriminator (D). Similar as the literally expression, the discriminator serves to encode both generated and ground truth images into high dimensional vectors. Then those vectors were judged by corresponding loss function to determine whether an input is ground truth image or generated one. The existence of D pushes G to generate images that are closer to ground truth, while the development of G in turn pushes D to extract more distinct features in the inputs to facilitate more precise judgement. For the EGAN discriminator we have followed the discriminated architecture guideline proposed in [28] which contains eleven Conv layers. The first six Conv layers have an increasing number of feature channels by the factor of 2 from 64 to 2048. The next two Conv layers decrease output feature channels to 1024 and 512, then connected with the output after three Conv layers through residual connection. The resulting features are followed by a dense layer and a sigmoid activation function to get the possibility distribution. The discriminator structure is shown in Fig. 1.

## 2.3. Loss functions

Two different loss functions are needed for G and D training because of their different design purposes, namely the generator loss and the discriminator loss.

### 2.3.1. Generator loss

As the generator aims to produce ground-truth alike images, the generator loss is designed to contribute to that purpose by comparing generated images with reference images and computing the differences. In our EGAN model, we firstly adopt two loss function items proposed in [18], depicted as the content loss and the VGG (visual geometry group network) loss. The content loss includes MSE (mean square error) loss that computes the mean squared pixel-wise error, and the Adv (adversarial) loss that computes the impossibility that indicates how likely an enhanced image be judged as a reference image. VGG loss computes the mean squared error of the deep extracted

features, with the features extracted by the VGG19 network [29]. The formulations are depicted as follow:

$$L_{\text{MSE}} = \frac{1}{WH} \sum_{x=1}^W \sum_{y=1}^H [I_{x,y} - G(I_{x,y})]^2 \quad (1)$$

$$L_{\text{VGG}} = \frac{1}{WH} \sum_{x=1}^W \sum_{y=1}^H [\varphi(G(I_{x,y}))_{5,4} - \varphi(R_{x,y})_{5,4}]^2 \quad (2)$$

$$L_{\text{adv}} = -\log[D(G(I_{x,y}))] \quad (3)$$

where  $x, y$  mean axes along the image width and height,  $I_{x,y}$  represent the raw input image,  $R_{x,y}$  represent the reference ground truth,  $W$  and  $H$  mean the width and the height of the inputs.  $G(\cdot)$  and  $D(\cdot)$  mean the transformation function of the generator and the discriminator operation. The  $\varphi(\cdot)_{i,j}$  mean the features extracted in the  $j$ -th Conv before  $i$ -th MaxPooling layer in the VGG19 network.

Apart from that, we introduce a novel smooth loss for the halo-effect reduction:

$$L_{\text{smooth}} = \frac{1}{WH} \sum_{x=1}^W \sum_{y=1}^H \left\{ [\nabla_x(G(I_{x,y})) - \nabla_x(I_{x,y})]^2 + [\nabla_y(G(I_{x,y})) - \nabla_y(I_{x,y})]^2 \right\} \quad (4)$$

where  $\nabla_x(\cdot)$  and  $\nabla_y(\cdot)$  mean computing the gradients along  $x$ - and  $y$ -axis. By using the gradients of raw images as the guideline, the smooth loss helps to generate images with less artificial noises, hence being more similar to natural ones. In Section 4 we conducted more experiments to prove the improvement.

The whole generator loss function consists of four loss items:

$$L_{\text{gen}} = \lambda_1 L_{\text{MSE}} + \lambda_2 L_{\text{VGG}} + \lambda_3 L_{\text{adv}} + \lambda_4 L_{\text{smooth}} \quad (5)$$

where  $\lambda_{1-4}$  stands for different weight values for each loss item. We choose their optimal values by the following steps: firstly, we set them based on referred papers, secondly, we set other three parameters constant and adjust one parameter to find the optimal value. Finally, we repeat the step two to find the optimal values for all lambda parameters.

### 2.3.2. Discriminator loss

The discriminator loss computes the logarithmic probabilities of different discriminator outputs. As the values of  $D(G(I_{x,y}))$  and  $D(R_{x,y})$  are between 0 and 1, the generated probabilities were compared to 0 while the reference ones were compared to 1. Similar as the  $L_{\text{adv}}$  in generator loss, the  $L_{\text{dis}}$  is as follows:

$$L_{\text{dis}} = -\log[D(R_{x,y})] + \log[D(G(I_{x,y}))] \quad (6)$$

### 3. Experiment and analysis

#### 3.1. Preparation

##### 3.1.1. Data

Although there are many synthetic underwater image datasets generated by CNN- or GAN-based methods, their practicability still remains uncertain for the reason of relatively similar color casting patterns and lacking of objects diversity, as shown in Fig. 2. Through comparison, we choose the UIEBD propose in [17] for training, for it covers a range of extreme underwater situations and contains abundant colorful objects. The training set contains 890 raws and according reference images and 60 challenging images. The training set contains 890 raw images with reference and other 60 challenging images. In practice, we first separate the original 890 images into 800 and 90 subsets for training and validation. Then we resize the training images to  $112 \times 112$ , and use flip and rotate operation to expand the training set to 6400 in total. The rest 60 challenge images are used for testing.

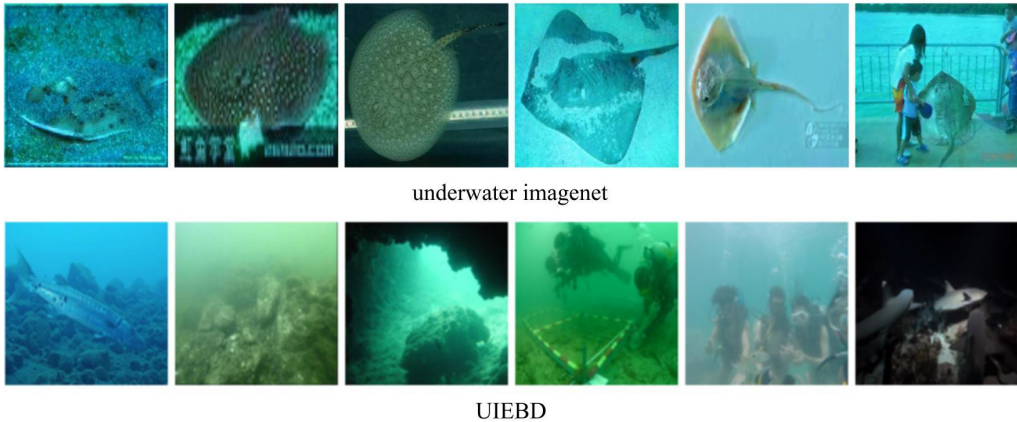


Fig. 2. Comparison of different underwater image datasets. The underwater imagenet synthesized by GAN [23] and the UIEBD collected from real scenes [17].

##### 3.1.2. Training settings

Based on experimental adjustment we set the generator loss weights  $\lambda_{1-4}$  as 5, 0.001, 0.0002, and 8. For EGAN training we adopt generator pre-training for 10 epochs with a learning rate of 0.0002, then we use adversarial training for the whole EGAN for 40 epochs with a learning rate of 0.0002 decayed by 0.5 every 10 epochs. The batch size is 16 and the patch size is  $64 \times 64$ . For optimization we use the Adam optimizer with Beta1 set as 0.9. The tensorflow is used for the network implementation. The whole model is trained on a server with an NVIDIA Tesla M40 GPU of 10G memory, and a 8G memory Intel CPU with 2.6 GHz frequency.

### 3.1.3. Comparing methods

For validation and test result performance we select the following state-of-the-art methods for comparison, including 1) IFM-based methods: IBLA [14], ULAP [15], FusionBased [13]; their results are obtained by the public available codes. 2) Deep-learning-based methods: UWCNN [30] and Duwie [17]; their results are obtained by the pre-trained model provided by the authors.

## 3.2. Experiment results

### 3.2.1. Evaluation on the validation set

As shown in Fig. 3, for different underwater distortions, EGAN produces enhanced images with more natural color distribution and finer object texture restorations. Specifically, for the images in the first row, our results achieve consistent greenish tone removal over corals regardless the distance. Similar performance can also be found in the fifth row. Besides, in the second row our results appear in clearer textures compared to the rest methods even including the reference image. The outstanding performance of EGAN in color correction accuracy and texture restoration can also be found in the third and fourth row images. Besides, PSNR and SSIM evaluation in Table 1 also confirm our competitiveness. To further demonstrate our competitiveness in fast image processing, we compared referenced methods with EGAN regarding the computing time. The experiment was conducted on single image and the result is shown in Table 2. Although the Duwie is based on deep learning too, they use most of time for image pre-processing. We conclude the results to the end-to-end manner of EGAN and the

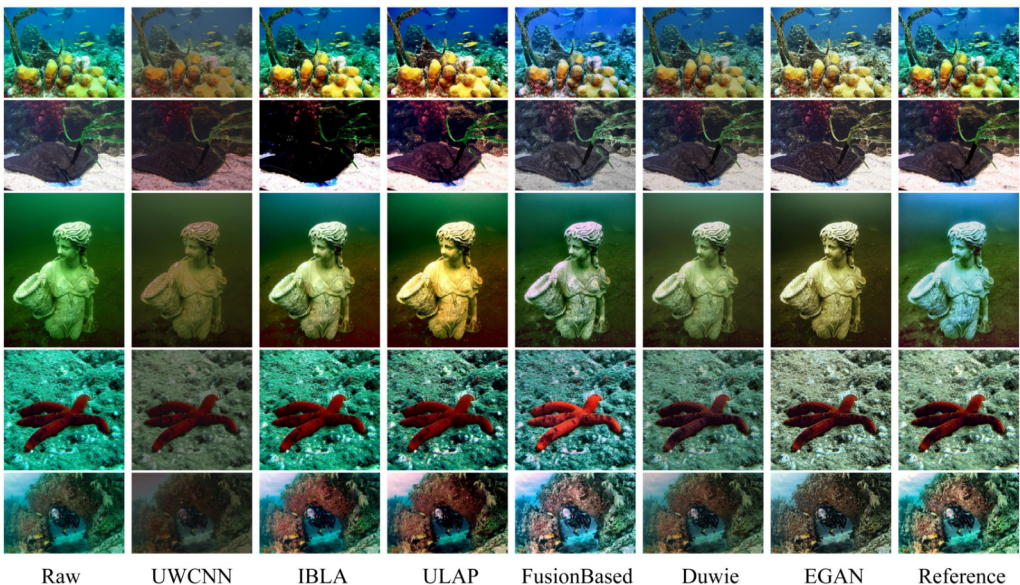


Fig. 3. Comparison of validation enhancement results. From left to right: raw images, UWCNN [30], IBLA [14], ULAP [15], FusionBased [13], Duwie [14], and our proposed EGAN, and reference images.



Table 1. PSNR and SSIM evaluation on validation set.

	Raw	UWCNN	IBLA	ULAP	FusionBased	Duwie	EGAN
PSNR	16.9136	13.5984	15.0323	15.8905	16.2535	19.1679	19.4463
SSIM	0.762	0.6579	0.6365	0.7359	0.7796	0.8326	0.8637

Table 2. Comparison of computing time.

UWCNN	IBLA	ULAP	FusionBased	Duwie	EGAN
30 s	526.21 s	30 s	5 s	5 s	0.01 s

generator structure adaptations we adopted: BN removal and Conv kernel replacement. By removing BN layers, the EGAN is able to reduce the training time, which will be a benefit for fast model training and image processing. By replacing Conv kernel size from  $3 \times 3$  to  $1 \times 1$ , more training and processing time were saved and less artificial noises were produced. More details about the methods effectiveness are illustrated in Section 4.

### 3.2.2. Evaluation on the test set

In the test set comparison shown in Fig. 4, our method also outperforms other methods in detail enhancement, shown in row 4 and 5 on the fish body texture. Moreover, the color correction is relatively more balanced as shown in row 1–4. Specifically, the outputs of UWCNN [30] appear dim in light intensity. Outputs of FusionBased [13], turn out to be over compensated in red channel.

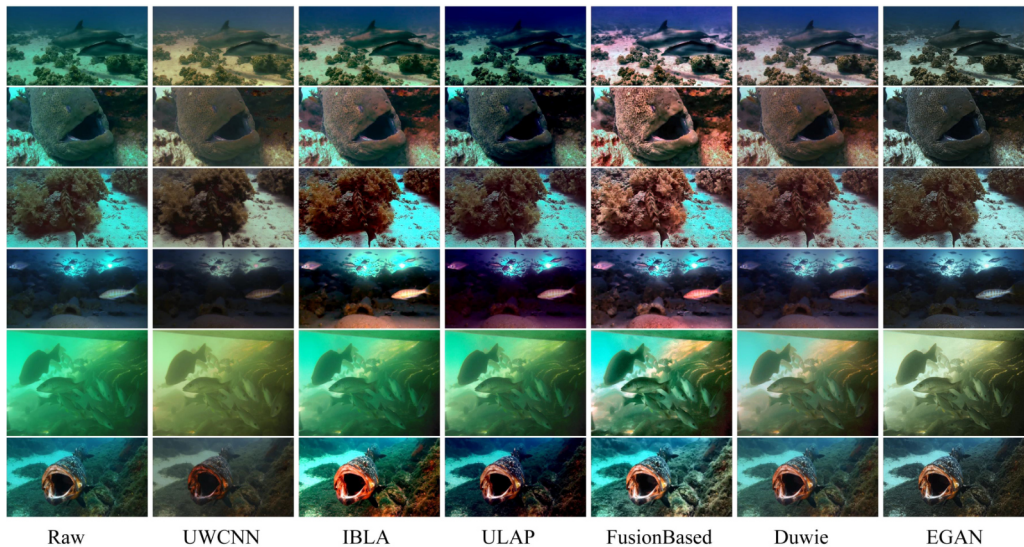


Fig. 4. Comparison of test sets. From left to right: raw images, UWCNN [30], IBLA [14], ULAP [15], FusionBased [13], Duwie [17], and our proposed EGAN.

## 4. Methods effectiveness study

### 4.1. BN removing and Conv kernel size replacing

We started our research to make the underwater image enhancement process fast and effective, and we adopted BN removing and Conv kernel size replacing to achieve this goal. To further study the merits of removing BN and changing Conv kernel size in generator structure, we conduct four experiments controlling: with or without BN in generator;  $3 \times 3$  or  $1 \times 1$  kernel for second Conv in the generator residual block. We label them as: BN +  $3 \times 3$ , BN +  $1 \times 1$ , noBN +  $3 \times 3$ , and noBN +  $1 \times 1$ . The comparing results are shown in Fig. 5.

As shown in the original and zoomed view, there are two distinguishable differences in halo-effect and contrast. Specifically, in the second row, between column 1–2 and 3–4, the output of  $1 \times 1$  kernel has less halo around the jellyfish, demonstrates the effect of  $1 \times 1$  kernel in halo-effect reducing. Comparing column 1 to 3 or 2 to 4 in

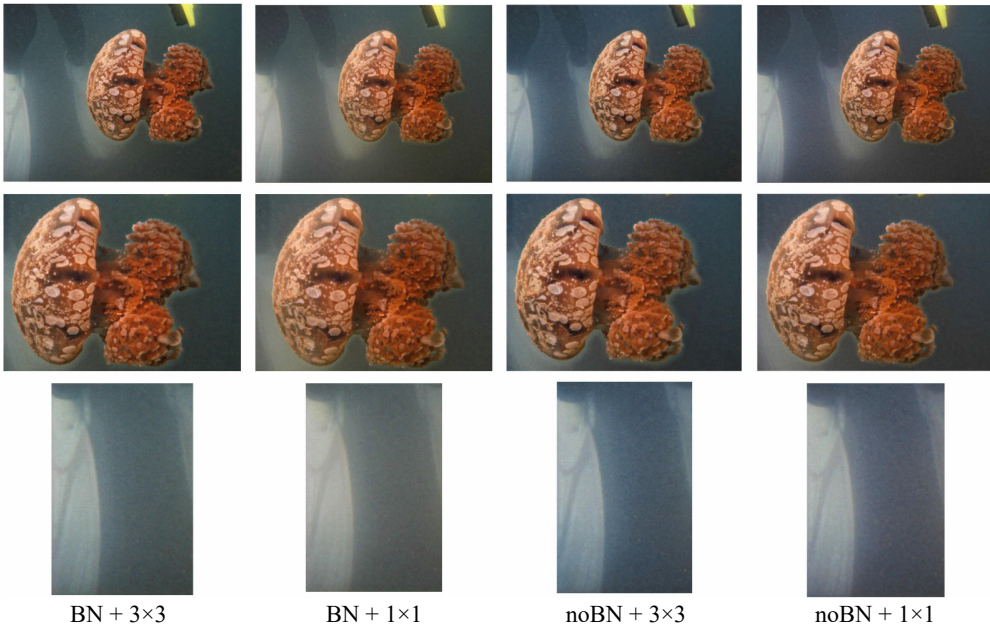


Fig. 5. Methods comparison on BN and  $3 \times 3$  kernel size. Row 1: original view, row 2: zoomed view of coral, and row3: zoomed view of human.

T a b l e 3. Methods comparison.

	BN + $3 \times 3$	BN + $1 \times 1$	noBN + $3 \times 3$	noBN + $1 \times 1$
Training time (per epoch)	166 s	162 s	150 s	146 s
PSNR	19.2356	19.2876	19.3688	19.4463
SSIM	0.8608	0.8569	0.8629	0.8636

row 3, we could see that noBN pictures have relatively higher contrast, showing that removing BN could boost contrast.

As shown in Table 3, those two adaptations also individually reduce the per epoch training time, proving the ability in efficiency boost of these two methods.

## 4.2. Smooth loss

The artificial noises coming along with deep-learning methods have been a common problem. In this paper, we designed a novel smooth loss for EGAN generator. To further demonstrate the effect of the proposed smooth loss in halo-effect alleviation, we conduct two experiments: training with smooth loss and training without smooth loss. The detailed comparison is shown in Fig. 6. As the result shows, the output with the smooth loss has relatively invisible white halo around the jelly fish, proving the effectiveness of our innovation.

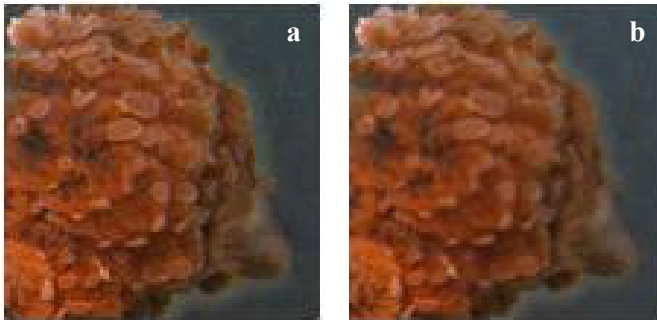


Fig. 6. Methods comparison on the smooth loss performance, (a) without smooth loss, and (b) with smooth loss.

## 4.3. Generator structure

We have known that deep-learning-based methods require a considerable amount of training time due to the complicated network structure. The proposed EGAN is able to achieve better enhance results in less training time, which is more suitable for practical application. For the whole generator design of EGAN, we adopt three architectural improvements comparing to the original generator structure proposed in POGAN [19]. Namely, we first remove all BN after Conv layers, then  $1 \times 1$  kernel is introduced to the second Conv in the residual block. Finally we reduce the total residual block number to 4 for faster training. For generator structure comparison, we choose four similar ResNet-based generator architectures: the original generator in POGAN [19] with 16 residual blocks as POGAN(res16), and its 4 residual block edited version as POGAN(res4), and generator from the RoR [31] and the ESRGAN [32]. Those models are trained in the same discriminator structure as EGAN, with same hyper-parameters settings and loss functions and other settings. As shown in Table 4, the generator structure of EGAN

Table 4. Generator structure comparison.

	POGAN(res16)	POGAN(res4)	RoR	ESRGAN	EGAN
Training time (per epoch)	253 s	166 s	167 s	186 s	147 s
PSNR	18.9979	19.2356	19.2571	19.2524	19.4463
SSIM	0.8570	0.8608	0.8562	0.8620	0.8637

achieves the best training time and PSNR/SSIM value, approving the effectiveness of the EGAN generator architecture.

### 5. Conclusion and future work

In this paper we propose an efficient generative adversarial network called EGAN for underwater image enhancement, additionally a smooth loss function is introduced for halo-effect alleviation. Extensive experiments are conducted to validate the effectiveness of our entire model and its specific features. However, during our research we also found the imperfection of our model in extreme color correction, specifically under the distortion of green color drifting. As shown in Fig. 7, the raw image (Raw),

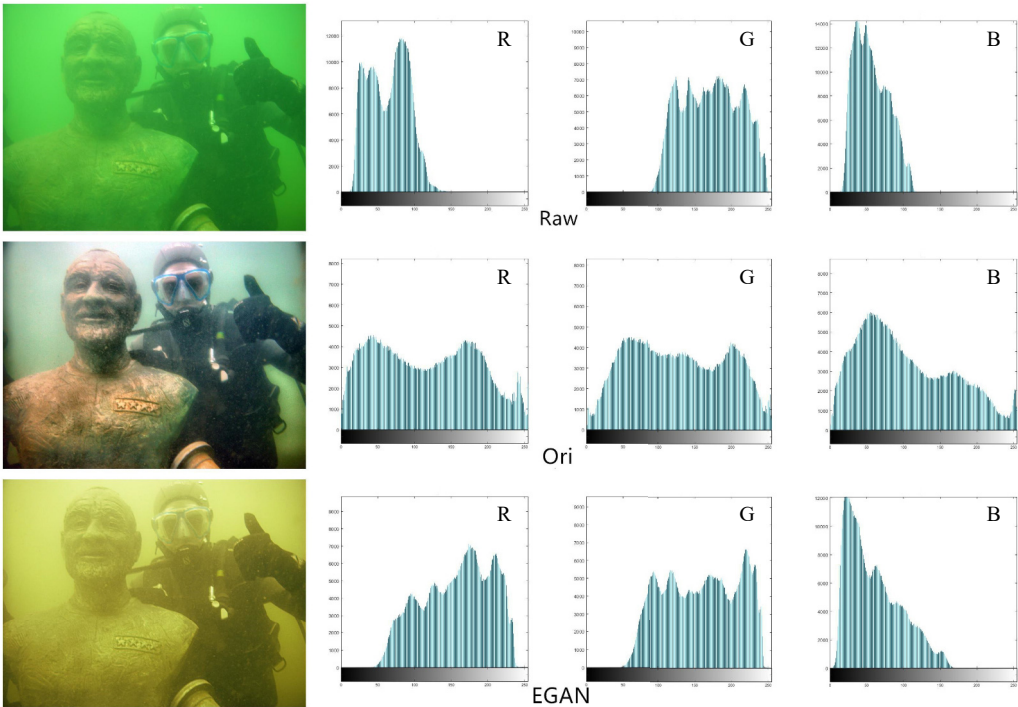


Fig. 7. RGB channels comparison.

reference image (Ori), and our result (EGAN) are listed vertically with their RGB pixels distribution diagrams (with R, G, B from left to right), plotted by the pixel value on the horizontal axis and the pixel number on the vertical. From Fig. 7 we could observe that the raw image has most of its green pixels shifted to high pixel value region, which leads to unsatisfied result of our model because most of our training images appeared with red channel shifting. Inadequate greenish and blueish training data are the main reason of imperfection and we will dedicate more work on that issue in our future work.

*Funding*—This work was supported by the Natural Science Foundation (61805082), the Key Technologies R&D Program of Guangzhou City (201704020038, 201704020182, 201803030008), and the Fundamental Research Funds for the Central Universities (2019NS125).

## References

- [1] JAFFE J.S., *Underwater optical imaging: the past, the present, and the prospects*, IEEE Journal of Oceanic Engineering **40**(3), 2015, pp. 683–700, DOI: [10.1109/JOE.2014.2350751](https://doi.org/10.1109/JOE.2014.2350751).
- [2] WANG Y., SONG W., FORTINO G., QI L.Z., ZHANG W., LIOTTA A., *An experimental-based review of image enhancement and image restoration methods for underwater imaging*, IEEE Access **7**, 2019, pp. 140233–140251, DOI: [10.1109/ACCESS.2019.2932130](https://doi.org/10.1109/ACCESS.2019.2932130).
- [3] JAFFE J.S., *Computer modeling and the design of optimal underwater imaging systems*, IEEE Journal of Oceanic Engineering **15**(2), 1990, pp. 101–111, DOI: [10.1109/48.50695](https://doi.org/10.1109/48.50695).
- [4] AKKAYNAK D., TREIBITZ T., *A revised underwater image formation model*, [In] *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 6723–6732, DOI: [10.1109/CVPR.2018.00703](https://doi.org/10.1109/CVPR.2018.00703).
- [5] HOU W., *A simple underwater imaging model*, Optics Letters **34**(17), 2009, pp. 2688–2690, DOI: [10.1364/OL.34.002688](https://doi.org/10.1364/OL.34.002688).
- [6] GIBSON R., ATKINSON R., GORDON J. [Eds], *A Review of Underwater Stereo-Image Measurement for Marine Biology and Ecology Applications*, Oceanography and Marine Biology: An Annual Review, Vol. 47, 2016, pp. 257–292.
- [7] DREWS P. JR, DO NASCIMENTO E., MORAES F., BOTELHO S., CAMPOS M., *Transmission estimation in underwater single images*, [In] *2013 IEEE International Conference on Computer Vision Workshops*, 2013, pp. 825–830, DOI: [10.1109/ICCVW.2013.113](https://doi.org/10.1109/ICCVW.2013.113).
- [8] GALDRAN A., PARDO D., PICÓN A., ALVAREZ-GILA A., *Automatic Red-Channel underwater image restoration*, Journal of Visual Communication and Image Representation **26**, 2015, pp. 132–145, DOI: [10.1016/j.jvcir.2014.11.006](https://doi.org/10.1016/j.jvcir.2014.11.006).
- [9] BERMAN D., LEVY D., AVIDAN S., TREIBITZ T., *Underwater single image color restoration using haze-lines and a new quantitative dataset*, IEEE Transactions on Pattern Analysis and Machine Intelligence **43**(8), 2021, pp. 2822–2837, DOI: [10.1109/TPAMI.2020.2977624](https://doi.org/10.1109/TPAMI.2020.2977624).
- [10] LU H., LI Y., ZHANG L., SERIKAWA S., *Contrast enhancement for images in turbid water*, Journal of the Optical Society of America A **32**(5), 2015, pp. 886–893, DOI: [10.1364/JOSAA.32.000886](https://doi.org/10.1364/JOSAA.32.000886).
- [11] HAN P., LIU F., YANG K., MA J., LI J., SHAO X., *Active underwater descattering and image recovery*, Applied Optics **56**(23), 2017, pp. 6631–6638, DOI: [10.1364/AO.56.006631](https://doi.org/10.1364/AO.56.006631).
- [12] PUROHIT K., MANDAL S., RAJAGOPALAN A.N., *Multilevel weighted enhancement for underwater image dehazing*, Journal of the Optical Society of America A **36**(6), 2019, pp. 1098–1108, DOI: [10.1364/JOSAA.36.001098](https://doi.org/10.1364/JOSAA.36.001098).
- [13] ANCUTI C., ANCUTI C.O., HABER T., BEKAERT P., *Enhancing underwater images and videos by fusion*, [In] *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 81–88, DOI: [10.1109/CVPR.2012.6247661](https://doi.org/10.1109/CVPR.2012.6247661).

- [14] PENG Y.T., COSMAN P.C., *Underwater image restoration based on image blurriness and light absorption*, IEEE Transactions on Image Processing **26**(4), 2017, pp. 1579–1594, DOI: [10.1109/TIP.2017.2663846](https://doi.org/10.1109/TIP.2017.2663846).
- [15] SONG W., WANG Y., HUANG D., TJONDRONEGORO D., *A rapid scene depth estimation model based on underwater light attenuation prior for underwater image restoration*, [In] Hong R., Cheng W.H., Yamasaki T., Wang M., Ngo C.W. [Eds], *Advances in Multimedia Information Processing – PCM 2018. PCM 2018. Lecture Notes in Computer Science*, Vol. 11164, Springer, Cham, pp. 678–688, DOI: [10.1007/978-3-030-00776-8\\_62](https://doi.org/10.1007/978-3-030-00776-8_62).
- [16] ANWAR S., LI C., PORIKLI F., *Deep underwater image enhancement*, [arXiv:1807.03528](https://arxiv.org/abs/1807.03528) [cs.CV], 2018.
- [17] LI C., GUO C., REN W., CONG R., HOU J., KWONG S., TAO D., *An underwater image enhancement benchmark dataset and beyond*, IEEE Transactions on Image Processing **29**, 2020, pp. 4376–4389, DOI: [10.1109/TIP.2019.2955241](https://doi.org/10.1109/TIP.2019.2955241).
- [18] LEDIG C., THEIS L., HUSZÁR F., CABALLERO J., CUNNINGHAM A., ACOSTA A., AITKEN A., TEJANI A., TOTZ J., WANG Z., SHI W., *Photo-realistic single image super-resolution using a generative adversarial network*, [In] *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 105–114, DOI: [10.1109/CVPR.2017.19](https://doi.org/10.1109/CVPR.2017.19).
- [19] DU Y., LI X., *Recursive image dehazing via perceptually optimized generative adversarial network (POGAN)*, [In] *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019, pp. 1824–1832, DOI: [10.1109/CVPRW.2019.00233](https://doi.org/10.1109/CVPRW.2019.00233).
- [20] ISOLA P., ZHU J.Y., ZHOU T., EFROS A.A., *Image-to-image translation with conditional adversarial networks*, [arXiv:1611.07004](https://arxiv.org/abs/1611.07004) [cs.CV], 2017.
- [21] LIU X., GAO Z., CHEN B.M., *MLFcGAN: multilevel feature fusion-based conditional GAN for underwater image color correction*, IEEE Geoscience and Remote Sensing Letters **17**(9), 2020, pp. 1488–1492, DOI: [10.1109/LGRS.2019.2950056](https://doi.org/10.1109/LGRS.2019.2950056).
- [22] 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 Robotics and Automation Letters **3**(1), 2018, pp. 387–394, DOI: [10.1109/LRA.2017.2730363](https://doi.org/10.1109/LRA.2017.2730363).
- [23] FABBRI C., ISLAM M.J., SATTAR J., *Enhancing underwater imagery using generative adversarial networks*, [In] *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 7159–7165, DOI: [10.1109/ICRA.2018.8460552](https://doi.org/10.1109/ICRA.2018.8460552).
- [24] HE K., SUN J., *Convolutional neural networks at constrained time cost*, [In] *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 5353–5360, DOI: [10.1109/CVPR.2015.7299173](https://doi.org/10.1109/CVPR.2015.7299173).
- [25] HE K., ZHANG X., REN S., SUN J., *Deep residual learning for image recognition*, [In] *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778, DOI: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [26] LIM B., SON S., KIM H., NAH S., LEE K.M., *Enhanced deep residual networks for single image super-resolution*, [In] *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017, pp. 1132–1140, DOI: [10.1109/CVPRW.2017.151](https://doi.org/10.1109/CVPRW.2017.151).
- [27] RONNEBERGER O., FISCHER P., BROX T., *U-Net: convolutional networks for biomedical image segmentation*, [In] Navab N., Hornegger J., Wells W., Frangi A. [Eds], *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science*, Vol. 9351, Springer, Cham, 2015, DOI: [10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28).
- [28] RADFORD A., METZ L., CHINTALA S., *Unsupervised representation learning with deep convolutional generative adversarial networks*, [arXiv:1511.06434](https://arxiv.org/abs/1511.06434) [cs.LG], 2015.
- [29] SIMONYAN K., ZISSERMAN A., *Very deep convolutional networks for large-scale image recognition*, [arXiv:1409.1556](https://arxiv.org/abs/1409.1556) [cs.CV], 2014.
- [30] LI C., ANWAR S., PORIKLI F., *Underwater scene prior inspired deep underwater image and video enhancement*, Pattern Recognition **98**, 2020, article 107038, DOI: [10.1016/j.patcog.2019.107038](https://doi.org/10.1016/j.patcog.2019.107038).

- [31] ZHANG K., SUN M., HAN T.X., YUAN X., GUO L., LIU T., *Residual networks of residual networks: multilevel residual networks*, IEEE Transactions on Circuits and Systems for Video Technology **28**(6), 2018, pp. 1303–1314, DOI: [10.1109/TCSVT.2017.2654543](https://doi.org/10.1109/TCSVT.2017.2654543).
- [32] WANG X., YU K., WU S., GU J., LIU Y., DONG C., QIAO Y., LOY C.C., *ESRGAN: enhanced super-resolution generative adversarial networks*, [In] Leal-Taixé L., Roth S. [Eds], *Computer Vision – ECCV 2018 Workshops. ECCV 2018. Lecture Notes in Computer Science*, Vol. 11133, Springer, Cham, 2018, DOI: [10.1007/978-3-030-11021-5\\_5](https://doi.org/10.1007/978-3-030-11021-5_5).

*Received July 24, 2020  
in revised form October 7, 2020*