Underwater Image Enhancer based on Conditional GANs

Xiaodong Liu[1] , Zhijie Wang[2], Yu Herng Tan[1], Yingcai Bi[1] and Ben M. Chen[1]

***Abstract*— Vision-based tasks conducted by autonomous underwater vehicles (AUVs) are highly dependent on the quality of the underwater images captured. However, unlike optical imaging in air, the characteristics of underwater environment poses unique challenges, resulting in images that are hazy and possess a level of degraded quality, which will consequently cause AUVs to exhibit poor performance on vision tasks. In this paper, we propose a novel architecture that can jointly extract local features and global features and elegantly fuse them together to perform underwater image enhancement based on the framework of conditional generative adversarial networks (cGANs) with residual blocks. The network is trained in an end-to-end fashion. Our approach is validated with subjective and objective tests and comparison against the state of the art, where we show great improvements. Furthermore, we demonstrate the proposed method can effectively boost the performance of underwater vision tasks such as object detection and tracking for underwater robots.**

I. Introduction

The development of vision-based autonomous underwater vehicles (AUVs) is a promising field as they are potential substitutes for human divers and hence useful for tasks such as ship hull inspection [1], coral reefs monitoring [2], pipeline corrosion inspection [3] and seabed mapping [4]. For these tasks, the acquisition of clear underwater images is highly important. However, unlike optical imaging in the air, underwater environment has its unique characteristics which manifests in severe light absorption and light scattering [5], resulting in images to appear highly degraded. The degraded images will invariably hamper the performance of underwater vision tasks [6], therefore bringing about the need for developing underwater image enhancing algorithms.

Underwater image enhancement can be viewed as translating the degraded image into its corresponding enhanced version where the key of the translation is the mapping from the degraded images to undegraded ones. Intuitively, image enhancement can be considered as the inverse process of image degradation. Thus, the enhanced image can be inferred based on estimating the parameters in the image degradation model. The majority of work in this area is based on the image degradation model and its variants [7]–[12].

Although much investigation has been made into modeling the underwater image distortion process, the intrinsic nonlinear degradation is affected by a large number of factors such as the ambient light condition, level of turbidity, or the salinity condition of the water, which are still difficult to model. In addition, these factors may vary a lot in different waters or even at different times of the day. Therefore, model-based methods do not generalize well since they utilize various assumptions or specified parameters to achieve reasonable results. It is desirable to design the enhancing algorithm with good generalization ability to reduce this dependence and improve the results under unknown conditions.

(a) Raw underwater image.

(b) Enhanced underwater image.

Fig. 1. Results of underwater image enhancer with the raw underwater images collected from the Internet. It shows that the enhanced images are able to remove the color cast while preserving fine details.

Recently, deep neural networks have shown powerful ability in the approximation of complex nonlinear functions with good generalization performance [13]. Such the deep learning-based methods have also been introduced to handle underwater image enhancement. In [14], the authors implement the convolutional neural networks (CNNs) which is trained to minimize the Euclidean distance between the generated image and the ground truth. However, minimize the Euclidean distance may produce blurry results as it is achieved by averaging all possible outputs [15]. On the other hand, the recently proposed generative adversarial networks (GANs) [16] are trained to minimize an adversarial loss where the generator is regulated by the discriminator to generate less blurry images. GANs have recently been applied and shown great potential on underwater image enhancement [13], [17].

Authors in [13] propose WaterGAN to generate realistic underwater images from in-air RGB-D images. However, the additional depth information can be difficult to obtain for underwater scenes. The network proposed in [17] is trained on the paired images and it can achieve good color restoration performance but it is still likely to generate noisy patches for the textureless areas. As stated in [15], generating images with fine details by GANs is still an open problem [15].

The human vision system has the ability to make adjustments according to the overall light conditions. Similarly,

[1] Xiaodong Liu, Yu Herng Tan, Yingcai Bi and Ben M. Chen are with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore xiaodongliu@u.nus.edu

[2]Zhijie Wang is with the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China paul.wangzhijie@bupt.edu.cn

Preprint submitted to 2019 International Conference on Robotics and Automation (ICRA). Received September 15, 2018.

cameras also support various adjustable capture settings (ISO, shutter time) for different scenes in order to improve the quality images. Therefore global condition of the image, which reveals high level information, is useful to adjust local pixels, helping to generate fine details. Global features have been explored and demonstrated as useful information by vision tasks such as image colorization [18]. As far as we know, we are the first to introduce global features into conditional GANs framework with residual blocks for underwater image enhancement. As such, our method is developed based on cGANs and a novel architecture is proposed for better enhancing performance. Our main contributions are:

1) We propose a novel architecture which uses a combination of global image features and local image features with residual blocks to automatically enhance underwater images.
2) Our method shows better results than state-of-the-art methods and significantly boost the underwater object detection and tracking performance.

Extensive experiments are conducted and one example is shown in figure 1. The source code and more experiment results are made publicly available.[1].

II. Related Works

Many of the prior work on compensating the effects of underwater image formation are based on the image degradation model as in [7]–[12]. However, the model based methods are often not generalizable across different scenes where their assumptions in the model may not hold.

Alternatively, deep neural networks has already become the common workhorse for a variety of vision tasks due to the great ability of modeling nonlinear function. For image-to-image translation problem, GANs have been successfully used for image inpainting [19], style transfer [20] and tasks like edges to photos translation [15]. Underwater image enhancement can also be formulated into this type of problem similarly. The goal is to translate the degraded underwater images, suffering from color cast and hazy effect, into images that appear to be above water with fine details. There has already been implementations of GANs for underwater image enhancement as described in [13] and [17].

In [13], the authors propose WaterGAN to learn the water column properties and generate the realistic underwater images from in-air RGBD images and real underwater images from a particular site. However, unlike in-air images, additional depth information is not easy to obtain for underwater scenes.

Authors in [17], propose to use CycleGAN [21] to learn the intrinsic mapping from undegraded underwater images to degraded ones. Then the visually undegraded images are degraded by CycleGAN and thus the paired training images are generated. Afterwards, the U-net based generator under the GAN framework is trained with the paired images. However, we find the direct use U-net as generator for our task is prone to generating noisy patches in the textureless areas .

Different from [17], we design a novel generator that uses a combination of global image features, which are extracted efficiently from the entire image with residual blocks, and local image features, which are computed from small image patches. Global features and local features are then effectively fused together into our proposed network.

III. Network Design

For the image enhancement task, instead of only feeding the noise vector z into the GANs, we also feed the raw degraded underwater image x into the network, which makes GANs in the conditional setting. Therefore, the generator is trained to learn the mapping as $G : \{x, z\} \mapsto y$ where y denotes the undegraded images. Besides, unlike unconditional GAN, both the discriminator and generator will observe the raw underwater images, which shows better performance for image-to-image translation task [15].

A. Objective Function

As the original loss function of GANs is found to be difficult to optimize, suffering from gradient vanish and model collapse, here, we adopt the improved loss of Wasserstein GAN with gradient penalty (WGAN-GP) as proposed in [22]. The adversarial loss can be denoted as

$$\begin{aligned}\mathcal{L}_{cWGAN} = E_{x,y}[D(x,y)] - E_{x,z}(D(x,G(x,z))) \\ +\lambda_{GP} E_{\hat{x}}[(|| \bigtriangledown_{\hat{x}} D(\hat{x})||_2 - 1)^2]\end{aligned} \tag{1}$$

where $\hat{x}$ is the sample alone the straight lines between the generated image $G(x, z)$ and the real undistorted image y. λ_{GP} is the weight factor.

In addition, as it is reported to be beneficial to mix the adversarial loss with other traditional loss [23].Hence, $L1$ loss is adopted to regularize the generator network

$$\mathcal{L}_{L1}(G) = E_{x,y,z}[||y - G(x,z)||_1] \tag{2}$$

Therefore the overall objective function $\mathcal{L}^*$ is

$$\mathcal{L}^* = \min_G \max_D \mathcal{L}_{cWGAN}(G,D) + \lambda_1 \mathcal{L}_{L1}(G) \tag{3}$$

where λ_1 is the weighting factor for $L1$ loss.

B. The Generator

As the generator is the core of our network and functions as our underwater image enhancer, we specially design the structure of the generator here. Our generator is based on the U-net [24] as used in [17]. Unlike generator in [17], we propose a novel structure to extract the local features and global features efficiently in the encoder part at the same time and fuse them together effectively.

[1]https://github.com/Xiaodong-Bran/underwater-image-enhancer

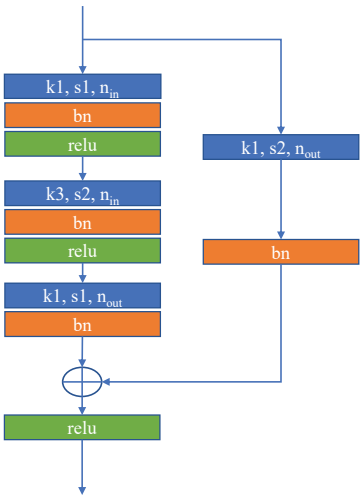


Fig. 2. Basic residual block. k is the size of the convolution kernel. k1 means the kernel size is 1×1. s stands for stride. s1 means the stride for convolution is 1. bn stands for batch normalization, relu stands for rectified linear unit layer. n_{in} is the channel of input feature maps. n_{out} is the channel of output feature maps.

1) Residual blocks: Normally, the local features are extracted through convolutional layer by layer and the size of the reception field grows slowly due to the small size of convolution filter. Therefore, to extract the global features which cover the entire image, a deeper network is required. To efficiently make the network deeper as well as easy to optimize, we adopt the residual blocks proposed in [25] in the encoder part. Unlike the direct connection from the input to the output in traditional residual block, we add an additional convolutional layer for dimensional matching as shown in figure 2.

2) Shared local features: In order to improve the model efficiency, the global feature extraction and the local feature extraction share the same parameters for the first 4 blocks. The local features are extracted from small image patches with small receptive field. The size of the input image is fixed at 256×256 due to memory limitations. After 4 blocks for down-sampling (with stride equals 2), it generates the feature map of size 16×16 (with 512 channels).

3) Global image features: The global features need a coverage of the entire image. To extract the global features, the feature map is further reduced along the global feature extraction path with another 4 residual blocks and is finally reduced to 1×1 (with 512 channels).

4) Fusing the global and local features: The extracted 1×1×512 global features are then duplicated into 16×16 copies generating 16×16×512 feature maps. Then they are fused with the low-level feature maps generated by the 4^{th} block of the encoder.

Remark: While, where to fuse the global features and local features is a compromise between the weight of the global feature and local feature on the final fused feature map. If the global features are fused to the 3^{rd} block with feature map size 32×32, and the encoder is truncated there, possibly the low level features are not efficiently extracted. If the global features are fused to the 5^{th} block with feature map size 8×8, then the local features may outweigh the global features. For maximizing the fusion effects, we test to fuse the global features to different blocks in the encoder i.e. the 3^{rd}, 4^{th} and 5^{th}. We find that the best performance is achieved when the global feature is concatenated to the 4^{th} blocks which is consistent to our previous analysis.

The decoder part of the enhanced U-net is performed on the fused feature map.

Recent work in [26] proposed to introduce a self-attention layer into the GAN framework, aiming to make use of global information for better image generation. However, we attempt to add the self-attention layer but no further added benefit is observed. The reason may be that self-attention layer could not add more information into our network since we have already maximized the fusion effects of the global and local features. The overall structure of the generator is shown in figure 3.

C. The Discriminator

The discriminator we used here is based on the 70×70 PatchGAN [15]. As we are using WGAN-GP loss, it is designed to penalize the norm of the gradient of the discriminator with respect to each input. Therefore, it should not involve the correlations between each input and the batch normalization layer of PatchGAN is removed.

IV. Experiment

In this section, we firstly test the proposed generator individually to verify the effectiveness of the proposed structure. Then comparisons are made between the state-of-the-art methods based on validation images and test images separately. Lastly, the improvements brought by our image enhancer for underwater vision tasks are evaluated.

Dataset and training details: The dataset we use here for training are provided by [17]. It contains 6128 image pairs with ground truth (undistorted underwater image) and distorted underwater image. We randomly select 6000 image pairs as training set and the remaining 128 images as the validation dataset. The network is trained using ADAM [27] optimizer with learning rate equals 0.0002 and $\beta_1 = 0.5$. The batch size is 1. $\lambda_{GP} = 10, \lambda_1 = 100$. The model is trained with a NVIDIA GTX 1080 GPU which takes around 40 hours for 100 epochs.

A. Evaluation of the Proposed Generator Structure

To evaluate the performance of our proposed generator, we use the test method which is similar to [28]. Since we only test the generator here, we do not consider discriminator loss. The testing is to minimize mean square error (MSE), i.e. maximize the peak signal-to-noise ratio (PSNR) between the generated image and the ground truth.

The comparison is made among the following three types of generator: A: the generator in [17]. B: the generator in [17] augmented with global features (without residual blocks). C: the proposed generator (with global features and residual

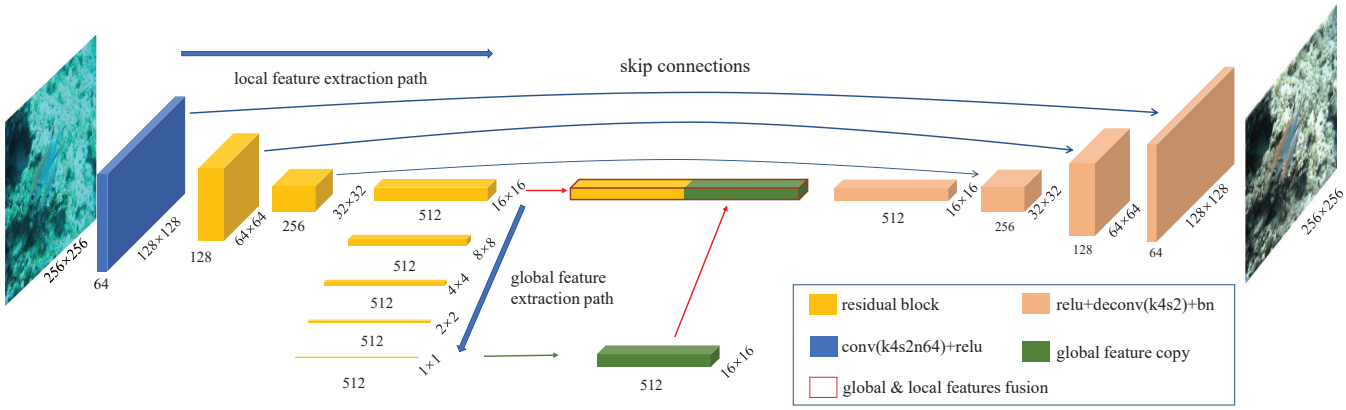


Fig. 3. Generator structure.The encoder part consists of residual block with number of output channels n_{out}= 128,256,512,512,512,512,512 respectively. The decoder part structure is based on U-net.

blocks). The evaluation metrics are: PSNR and Structure Similarity Index (SSIM) where PSNR indicates the average pixel similarity and SSIM measures the structure similarity between the enhanced images with ground truth. For both metrics, higher values suggest better performance. The recently proposed Inception Score [29] and Frchet Inception distance (FID) [30] for evaluation the quality of images generated by GANs are not suitable to our case as image enhancement does not generate a specified class of image.

Table I shows the average PSNR and SSIM values for the compared generators. The training is based on the dataset [17] with batch size of 16 and equivalent epochs. The tests are conducted three times for each generator and best performances are reported. By adding the global features into the generator in [17], the PNSR improved 0.52dB. Implemented with residual blocks, the generator can further boost its modeling ability so as to improve the PSNR to 27.22dB which outperforms the other configurations.

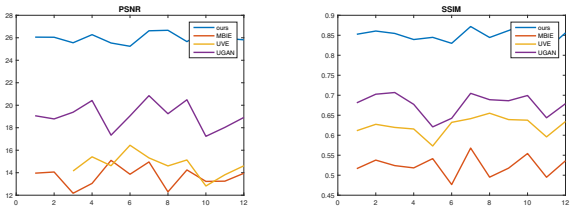
TABLE I
EVALUATION OF THE PROPOSED GENERATOR STRUCTURE.

	A	B	Ours
PSNR	26.49	27.01	**27.22**
SSIM	0.8006	0.8055	**0.8055**

B. Comparison with the State of the Art

We compare the our method with the state-of-the-art methods that are proposed in the past two years for underwater image enhancement: i.e. multi-band image enhancement (MBIE) [12], underwater visibility enhancement (UVE) [11] and enhancing underwater image with GAN (UGAN) [17]. The comparison are made based on the validation and test images separately.

Validation image set: We calculate the average PSNR and SSIM based on every 10 images instead of using the average of the whole validation dataset to prevent the case where a method performs well on some images but performing poorly on others. The results are shown in figure 4. It shows that our method achieves the highest PSNR and SSIM with a large margin and outperforms other methods.



(a) PSNR (b) SSIM

Fig. 4. Average PSNR and SSIM for validation dataset.

Test image set: To evaluate the generalization ability, 86 underwater images are collected from Flickr website tagged with underwater. Some of the results are shown in figure 5. It can be shown that UVE [11] can effectively enhance the image contrast as its last step utilize the contrast limited adaptive histogram equalization [31] for contrast enhancement. However, it shows limited effect to remove color cast. MBIE [12] can generate images with good saturation. But, it also performs poorly in handling color cast and introduces a superfluous red hue. This may because that the method estimates color distortion based on the estimated ambient light, however, the assumed parameters may not hold for some underwater scenes. UGAN [17] can achieve reasonable color restoration but generates noisy patches around the textureless area. On the contrary, by fusing the global features and local features, our method can enhance the details and suppress the unwanted artifacts. Further comparisons between UGAN and our method are shown in figure 6.

Mean opinion score test: As there is no ground truth and no effective metric to evaluate the performance of color restoration and the detail preserving for underwater image, similar to [32], we perform the mean opinion score (MOS) test. We randomly select 25 images from the test images for user study. 8 raters are invited to assign an integer score from 1(bad quality) to 4 (good quality). The MOS scores are presented in table II. In table II, our method receives

Preprint submitted to 2019 International Conference on Robotics and Automation (ICRA). Received September 15, 2018.

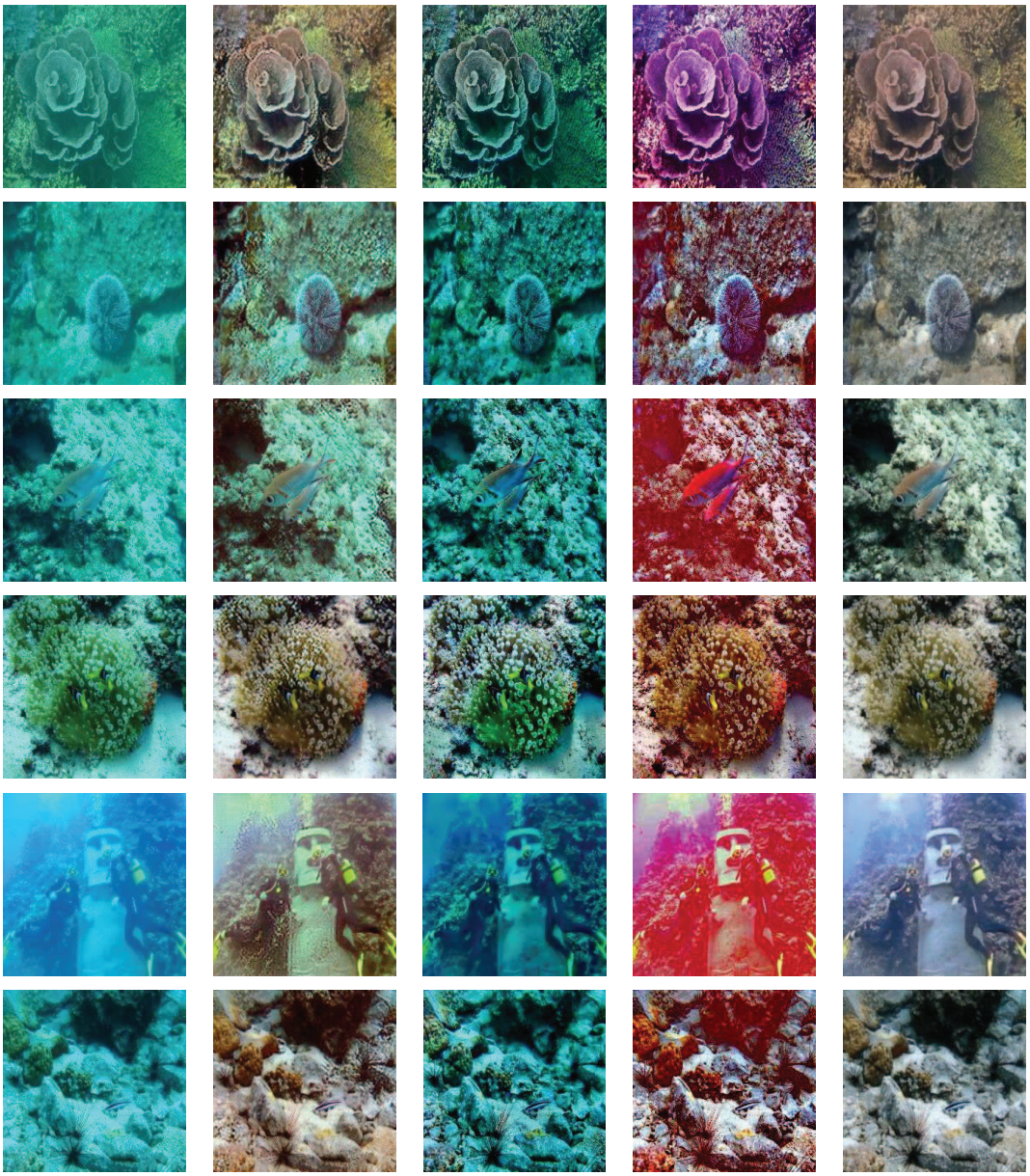
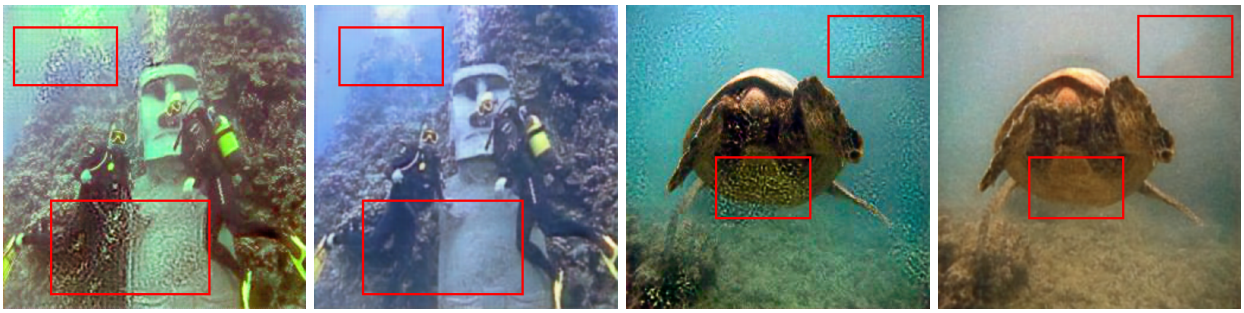


Fig. 5. From the left to right are the original raw images, the results of UGAN, UVE , MBIE and our method respectively.



(a) UGAN (b) Ours (c) UGAN (d) Ours

Fig. 6. Further comparison between UGAN and our method with details.

Preprint submitted to 2019 International Conference on Robotics and Automation (ICRA). Received September 15, 2018.

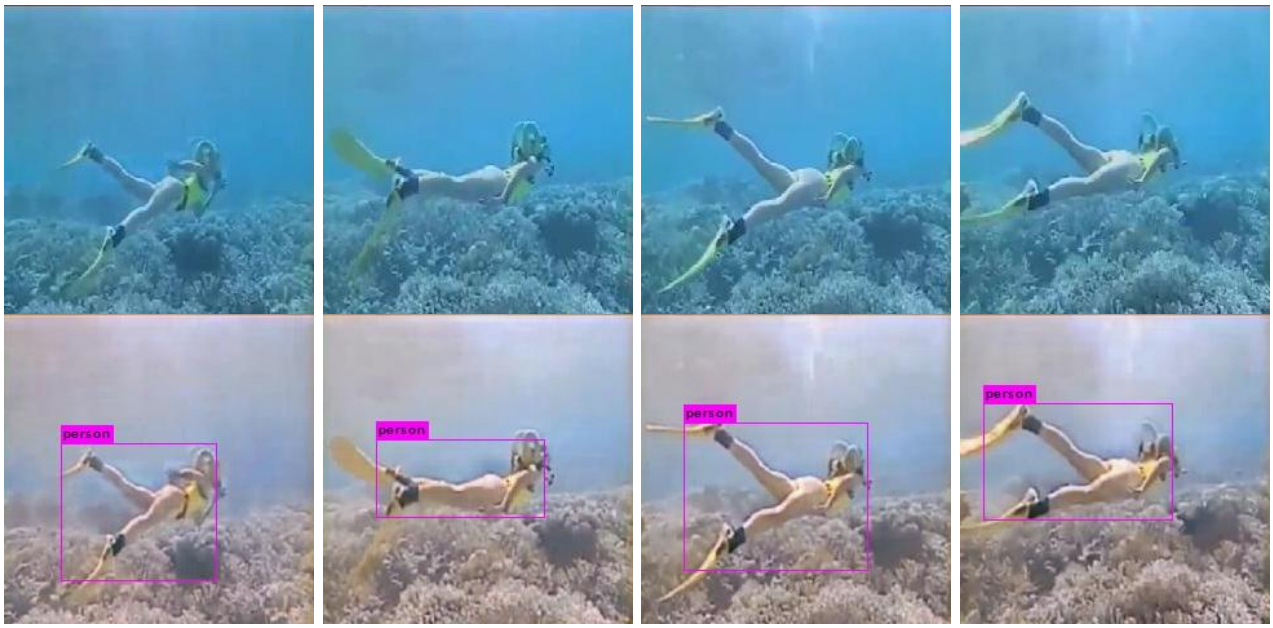


Fig. 7. Performance of the YOLOv3 on diver detection. Top row is the original video sequence, bottom row is the enhanced images by our method. YOLOv3 easily fails to detect the diver based on the original video sequence due to image degradation. With enhanced images, the detection rate is greatly improved.

TABLE II
THE MOS SCORE FOR SAMPLED TEST IMAGES.

	MBIE	UVE	UGAN	Ours
MOS	2.4028	2.5443	1.556	**3.5972**

the highest score, which indicates that from human visual perspective, our method produce more visually pleasing results than others.

C. Underwater Vision Task Evaluation

For underwater task such as archaeological site exploration that still needs human diver intervention. AUVs are required to track the divers as well as to wait for their commands [33]. Here, we take the diver detection and tracking for example to demonstrate the performance improvement brought by the proposed method. The person detection method used here is the state-of-the-art object detection framework YOLOv3 [34]. To evaluate the detection and tracking performance, three video clips[2] that contain human divers are randomly selected from YouTube. Examples of person detection based on YOLOv3 is shown in figure 7. The comparison between different methods are made and the number of correctly detected frames are summarized in table III. Total number of frames for three clips are 768, 570, 130 respectively. It is shown that our method can efficiently improve the correct detection rate and outperform the state-of-the-art methods. Furthermore, we test our enhancer on NVIDIA Jetson TX2 embedded platform, which can be used as onboard processor for underwater robots. It reaches 5Hz for image enhancement, which is sufficient for basic underwater vision tasks.

TABLE III
PERFORMANCE ON DIVER DETECTION AND TRACKING.

# of frames	Video1 (768)	Video2 (570)	Video3 (130)
Original	382	105	3
MBIE	281	147	52
UVE	339	115	52
UGAN	257	200	52
Ours	**442**	**326**	**58**

V. CONCLUSION

We have presented an end-to-end solution for underwater image enhancement based on the framework of cGANs. Under this framework, we propose a novel architecture with residual blocks to jointly extract both the global and local features. Experiment results demonstrate that our method can effectively enhance the underwater images with fine details and outperform the state of the art. In addition, we also demonstrate that the performance of underwater vision tasks can be greatly improved with our image enhancer. Future work will be considered to collect and train the network on a larger diverse set of underwater images with different level of turbidity to make our underwater image enhancer more generalizable.

ACKNOWLEDGMENT

The authors would like to thank Yonggun Cho, Ayoung Kim and Cameron Fabbri for sharing the dataset and the source code for [11], [12], [17].

[2]Links for the videos are
video1: https://www.youtube.com/watch?v=XBmYNoKMOuA&t=15s
video2: https://www.youtube.com/watch?v=P-zSeK5ve2M&t=165s
video3: https://www.youtube.com/watch?v=L5vvYLOlowo

REFERENCES

[1] P. Ozog, N. Carlevaris-Bianco, A. Kim, and R. M. Eustice, "Long-term mapping techniques for ship hull inspection and surveillance using an autonomous underwater vehicle," *Journal of Field Robotics*, vol. 33, no. 3, pp. 265–289, 2016.

[2] F. Shkurti, A. Xu, M. Meghjani, J. C. G. Higuera, Y. Girdhar, P. Giguere, B. B. Dey, J. Li, A. Kalmbach, C. Prahacs *et al.*, "Multi-domain monitoring of marine environments using a heterogeneous robot team," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*. IEEE, 2012, pp. 1747–1753.

[3] A. Khan, S. S. A. Ali, F. Meriaudeau, A. S. Malik, L. S. Soon, and T. N. Seng, "Visual feedback–based heading control of autonomous underwater vehicle for pipeline corrosion inspection," *International Journal of Advanced Robotic Systems*, vol. 14, no. 3, p. 1729881416658171, 2017.

[4] L. Paull, M. Seto, J. J. Leonard, and H. Li, "Probabilistic cooperative mobile robot area coverage and its application to autonomous seabed mapping," *The International Journal of Robotics Research*, vol. 37, no. 1, pp. 21–45, 2018.

[5] C. D. Mobley, *Light and water: radiative transfer in natural waters*. Academic press, 1994.

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

[7] N. Carlevaris-Bianco, A. Mohan, and R. M. Eustice, "Initial results in underwater single image dehazing," 2010.

[8] J. Y. Chiang and A. Ying-Ching Chen, "Underwater image enhancement by wavelength compensation and dehazing (wcid)," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1756–1769, 2012.

[9] A. Galdran, D. Pardo, A. Picón, and A. Alvarez-Gila, "Automatic red-channel underwater image restoration," *Journal of Visual Communication and Image Representation*, vol. 26, pp. 132–145, 2015.

[10] 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 Transactions on Image Processing*, vol. 25, no. 12, pp. 5664–5677, 2016.

[11] Y. Cho and A. Kim, "Visibility enhancement for underwater visual slam based on underwater light scattering model," in *Robotics and Automation (ICRA), 2017 IEEE International Conference on*. IEEE, 2017, pp. 710–717.

[12] Y. Cho, J. Jeong, and A. Kim, "Model assisted multi-band fusion for single image enhancement and applications to robot vision," *IEEE Robotics and Automation Letters*, 2018.

[13] 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 Robotics and Automation Letters*, vol. 3, no. 1, pp. 387–394, 2018.

[14] Y. Wang, J. Zhang, Y. Cao, and Z. Wang, "A deep cnn method for underwater image enhancement," in *Image Processing (ICIP), 2017 IEEE International Conference on*. IEEE, 2017, pp. 1382–1386.

[15] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *arXiv preprint*, 2017.

[16] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[17] C. Fabbri, M. J. Islam, and J. Sattar, "Enhancing underwater imagery using generative adversarial networks," *arXiv preprint arXiv:1801.04011*, 2018.

[18] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Let there be color!: joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 4, p. 110, 2016.

[19] R. A. Yeh, C. Chen, T.-Y. Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do, "Semantic image inpainting with deep generative models." in *CVPR*, vol. 2, no. 3, 2017, p. 4.

[20] C. Li and M. Wand, "Precomputed real-time texture synthesis with markovian generative adversarial networks," in *European Conference on Computer Vision*. Springer, 2016, pp. 702–716.

[21] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *arXiv preprint*, 2017.

[22] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," in *Advances in Neural Information Processing Systems*, 2017, pp. 5767–5777.

[23] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2536–2544.

[24] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[26] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," *arXiv preprint arXiv:1805.08318*, 2018.

[27] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[28] Y.-S. Chen, Y.-C. Wang, M.-H. Kao, and Y.-Y. Chuang, "Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 6306–6314.

[29] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," in *Advances in Neural Information Processing Systems*, 2016, pp. 2234–2242.

[30] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," in *Advances in Neural Information Processing Systems*, 2017, pp. 6626–6637.

[31] K. Zuiderveld, "Graphics gems iv," P. S. Heckbert, Ed. San Diego, CA, USA: Academic Press Professional, Inc., 1994, ch. Contrast Limited Adaptive Histogram Equalization, pp. 474–485. [Online]. Available: http://dl.acm.org/citation.cfm?id=180895.180940

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

[33] N. Mišković, A. Pascoal, M. Bibuli, M. Caccia, J. A. Neasham, A. Birk, M. Egi, K. Grammer, A. Marroni, A. Vasilijević *et al.*, "Caddy project, year 3: The final validation trials," in *OCEANS 2017-Aberdeen*. IEEE, 2017, pp. 1–5.

[34] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.