



WSUIE: Weakly Supervised Underwater Image Enhancement for Improved Visual Perception

Lin Hong, Xin Wang , Zhenlong Xiao, Gan Zhang, and Jun Liu 

Abstract—Underwater images inevitably suffer from degradation and blur due to the scattering and absorption of light as it propagates through the water, which hinders the development of underwater visual perception. Existing deep underwater image enhancement methods mainly rely on the strong supervision of a large-scale dataset composed of aligned raw/enhanced underwater image pairs for model training. However, aligned image pairs are not available in most underwater scenes. This work aims to address this problem by proposing a novel weakly supervised underwater image enhancement (named WSUIE) method. Firstly, a novel generative adversarial network (GAN)-based architecture is designed to enhance underwater images by unpaired image-to-image transformation from domain X (raw underwater images) to domain Y (arbitrary high-quality images), which alleviates the need for aligned underwater image pairs. Then, a new objective function is formulated by exploring intrinsic depth information of underwater images to increase the depth sensitivity of our method. In addition, a dataset with unaligned image pairs (named UIIE) is provided for the model training. Many qualitative and quantitative evaluations of the WSUIE method are performed on this dataset, and the results show that this method can provide improved visual perception performance while enhancing visual quality of underwater images.

Index Terms—Underwater image enhancement, weakly supervised learning, generative adversarial networks (GAN), underwater visual perception.

I. INTRODUCTION

WITH the increasing exploration of the underwater world, many scholars and institutions have participated in the research of underwater robots in order to find new ways to perform underwater tasks. Underwater vision, as an attractive sensing modality to perceive the underwater environment, has become a hot research topic due to its non-intrusiveness, passive nature, and high information content. High-quality underwater

images can provide a lot of valuable information for underwater robots to implement some tasks, such as underwater inspection [1], marine archaeology [2], and underwater grasping [3], etc. Over the last decades, in-air visual perception has gained rapid progress. However, the performance of advanced algorithms designed for in-air images usually degrades when applied to the underwater world. Thus, a series of methods have been proposed to bridge the gap between the in-air and underwater vision communities by enhancing the visual quality of underwater images, ranging from physical model-based methods to data-driven methods. The physical model-based methods inspired by the Jaffe-McGlamery model [4], [5] and dark channel prior (DCP) [6] have shown the remarkable effectiveness for single underwater image enhancement/restoration.

Compared with traditional methods based on physical model priors, the exploration of deep underwater image enhancement models is increasing recently. As a result, many data-driven underwater image enhancement methods have been proposed, mainly based on convolutional neural networks (CNNs) [7] and GANs [8]. Nowadays, there is still much room for improvement in existing methods as the learning perceptual enhancement for underwater imagery is a challenging ill-posed problem. One of the bottlenecks of these solutions is the strong need for aligned raw/enhanced image pairs for supervised networks training. Due to high labor costs and difficulty in collecting large-scale underwater datasets composed of annotated aligned image pairs, most data-driven models use small datasets or synthetic images as ground truth, and cannot capture natural variability of the underwater world [7], [9]. Recently, several unpaired image-to-image transformation works have been proposed to tackle this problem [8], [10]. They focused on relaxing the strong need for aligned image pairs in model training and designed some effective models inspired by DualGAN [11] and CycleGAN [12]. However, such a transformation can not guarantee a high visual perceptual quality for the underwater image. Few of these methods have seen designing robust and effective underwater image enhancement models and studying their applicability in improving underwater visual perception. This work attempts to solve these problems by introducing a novel underwater image enhancement method in a weakly supervised manner, and aims to improve the performance of underwater visual perception. Unlike previous data-driven underwater image enhancement works, this method can learn the semantic color of high-quality images while ensuring that the raw underwater image retains its key 2D and 3D attributes (*e.g.*, content, texture, depth information, etc), and our method does not require aligned image pairs for model training.

The contribution of this work includes three aspects:

Manuscript received March 22, 2021; accepted August 5, 2021. Date of publication August 18, 2021; date of current version August 31, 2021. This letter was recommended for publication by Associate Editor Nina Mahmoudian and Editor Pauline Pounds upon evaluation of the reviewers' comments. This work was supported in part by the Joint Funds of the Natural Science Foundation of China under Grant U1913206, in part by the Special Project for Research and Development in key areas of Guangdong Province under Grant 2019B090920001, in part by the Shenzhen Bureau of Science Technology and Information under Grant JCYJ20180306172134024. (*Corresponding author: Xin Wang.*)

Lin Hong, Xin Wang, Zhenlong Xiao, and Gan Zhang are with the School of Mechanical Engineering and Automation, Harbin Institute of Technology, Shenzhen 518055, China (e-mail: 20B953023@stu.hit.edu.cn; wangxinsz@hit.edu.cn; xzl_xiao@163.com; 20S153177@stu.hit.edu.cn).

Jun Liu is with the Department of Mechanical Engineering, City University of Hong Kong, Kowloon 518057, Hong Kong (e-mail: jliu287@cityu.edu.hk).

All the code and dataset are publicly available at <https://github.com/LinHong-HIT/WSUIE>.

Digital Object Identifier 10.1109/LRA.2021.3105144

- 1) A novel 3D visual descriptor is explored by estimating the camera-to-object distance of all image pixels, which helps generate the corresponding depth map of each monocular underwater image. Based on the depth maps, a new depth loss term is proposed and added to the full objective function of our method, thereby achieving the combination of 2D (local content and texture) and 3D (depth map) visual features.
- 2) A new large-scale data set (named UUIE) is provided for underwater image enhancement, which is composed of unaligned image pairs. The UUIE dataset is dedicated to weakly supervised model training and opens the door for further design of advanced underwater image enhancement methods more effectively and efficiently.
- 3) Many qualitative and quantitative evaluations of the WSUIE method have been conducted. The results show that the enhanced underwater images provide visual appeal to the public and improve the performance of some underwater visual perception tasks. This suggests that it is very promising to achieve underwater image enhancement in a weakly supervised way and to assist underwater visual perception.

II. RELATED WORKS

A. Underwater Image Enhancement

Data-driven underwater image enhancement methods have made rapid progress due to the emergence of many deep learning-based algorithms and the availability of large-scale datasets [13]. These methods can generally be divided into CNN-based and GAN-based. Most of them rely on large-scale datasets composed of aligned underwater image pairs for model training [14]–[17]. Li *et al.* [14] proposed a CNN-based underwater image enhancement model called UWCNN, which was an end-to-end model trained on a synthetic underwater image dataset. Sun *et al.* [15] proposed a deep model for underwater image enhancement by designing an encoder-decoder framework, which employed skip connection to avoid low-level features losing while accelerating the training process. Based on the dataset provided by [18], Liu *et al.* [16] proposed a deep multiscale feature fusion network for underwater image color correction inspired by the conditional GAN. Guo *et al.* [17] provided a new multiscale dense GAN to enhance underwater images, and the residual multiscale dense block was involved in the generator. In most cases, aligned underwater image pairs are difficult to obtain, several unsupervised methods [8]–[10] have been proposed to tackle this problem. Li *et al.* [9] proposed WaterGAN, which is used to generate realistic underwater images from the in-air images and their depth pairs in an unsupervised pipeline for color correction of underwater images. Islam *et al.* [10] proposed a fast underwater image enhancement method for real-time preprocessing in the autonomy pipeline by visually-guided underwater robots based on CycleGAN [12]. In [8], a weakly supervised method was proposed to correct color distortion of underwater images. In addition to the development of many effective supervised methods, underwater image enhancement research with unpaired image-to-image transformation is also worth exploring.

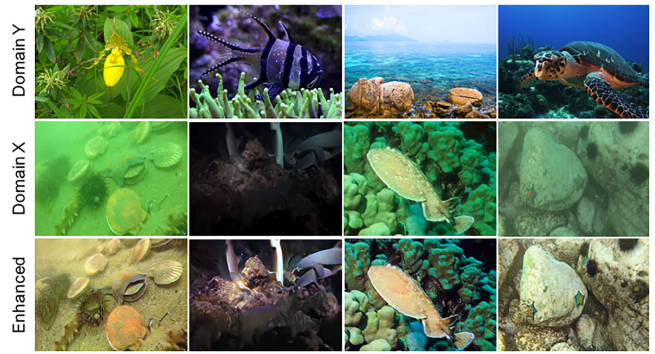


Fig. 1. Results of underwater image enhancement by the WSUIE method. The bottom row is enhanced underwater images generated by the WSUIE in an unpaired image-to-image transformation from the domain X (raw underwater images) to the domain Y (arbitrary high-quality images).

B. Underwater Visual Perception

Underwater visual perception has a wide range of applications. It has attracted increasing attention with the emergence of many underwater image enhancement/restoration algorithms. Boudhane *et al.* [19] proposed an underwater image enhancement method for underwater fish localization and detection. Chen *et al.* [20] proposed a GAN-based underwater image restoration scheme (GAN-RS) and applied the enhanced underwater images to marine products grasping. Li *et al.* [8] achieved underwater image color correction based on weakly supervised color transfer, and the enhanced image can provide better saliency detection performance than the raw underwater image. And in a recent work, Islam *et al.* [10] introduced a novel underwater image enhancement method (FUNIE-GAN). Its performance was verified based on some deep visual models for underwater objects detection, human body-pose estimation, and visual saliency prediction. As a promising sensing modality, many vision sensors will be widely used in many underwater tasks, and underwater image enhancement is undoubtedly vital for these tasks.

III. PROPOSED MODEL

The WSUIE method is proposed to enhance underwater images by unpaired image-to-image transformation inspired by the CycleGAN [12]. Fig. 2 shows the framework of our method, the two generators G_1 and G_2 are designed to learn cyclic mappings by evolving with adversarial discriminators D_X and D_Y through an iterative min-max game.

A. Network Architecture

1) *Generator*: as shown in Fig. 3, network architecture of the generator G_1 and G_2 follows [21]. The G_1 and G_2 have the same convolutional encoder-decoder framework but inverse direction: the G_1 learns the mapping from domain X to domain Y , and the G_2 learns the inverse mapping from domain Y to domain X . The input of the generator is RGB images of shape $3 \times 256 \times 256$, followed by two down-convolution blocks with stride 2, nine residual blocks [22], two transposed convolutional blocks with stride 1/2 for upsampling. Except for the first and last layers, which use 9×9 kernels, all convolutional layers use 3×3

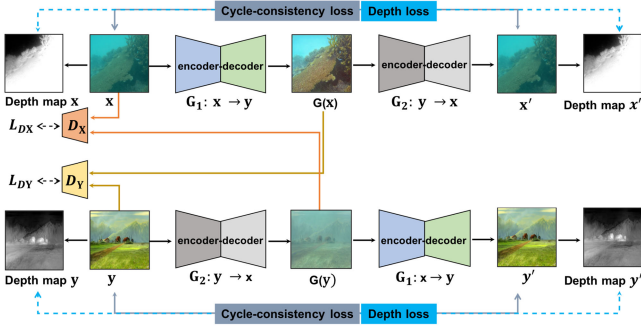


Fig. 2. Framework of WSUIE method. **The top stream** shows that a raw underwater image x in the domain X is transferred to that with style of high-quality images in the domain Y by the discriminator D_X . The generated image $G(x)$ is then transferred to its original state by cycle-consistency loss and depth loss. **The bottom stream** illustrates the inverse process.

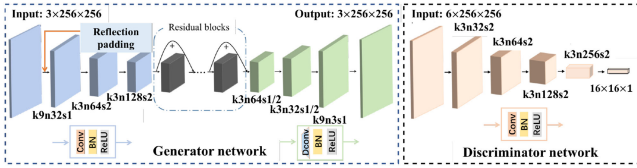


Fig. 3. Pipelines of the generator and discriminator. (k means the kernel size, n is the number of feature maps and s represents the stride, Conv is convolutional operation, DConv is transposed convolutional operation, BN is batch normalization.).

kernels. A nine-residual-block stack with each block containing two 3×3 convolutional layers, is designed to learn the identity function since, in most cases, the output image should share structure with the input image. Except for the output layer, all non-residual convolutional layers are followed by spatial batch normalization (BN) and ReLU nonlinearities.

2) *Discriminator*: for the discriminator, we use a Markov patch GAN [23], which assumes the independence of pixels beyond the size of the patch and only discriminates based on the patch-level information. Fig. 3 shows the architecture of the discriminator. The $6 \times 256 \times 256$ input represents a generated image and a target image. The following four convolutional layers are used to transform the $6 \times 256 \times 256$ input to a $1 \times 16 \times 16$ output representing the discriminator's averaged validity responses. In each layer, the 3×3 convolutional filter with a stride of 2 is used. The application of non-linearity and BN is the same as that of the generator.

B. Objective Function Formulation

The goal of the WSUIE method is to learn mappings between two domains X and Y given training samples $\{x_i\}_{i=1}^N \in X$ and $\{y_i\}_{i=1}^M \in Y$. As shown in Fig. 2, our objective function consists of three kinds of terms: two adversarial losses, a cycle consistency loss, and a depth loss.

1) *Standard GAN Loss*: there are two standard GAN losses to learn mappings $G_1: X \rightarrow Y$ and $G_2: Y \rightarrow X$, $X(Y)$ represents the source (target) domain. The standard adversarial loss



Fig. 4. Quad-tree subdivision to get the highest-intensity region of the underwater image for B_c estimation.

for the mapping function G_1 is:

$$\begin{aligned} \mathcal{L}_{GAN}(G_1, D_Y, X, Y) = & E_{y \sim P_{data}(y)} [\log D_Y(y)] \\ & + E_{x \sim P_{data}(x)} [\log(1 - D_Y(G_1(x)))] \end{aligned} \quad (1)$$

where G_1 tries to generate images from domain X that have the similar style with images in domain Y by $G_1(x)$, while D_Y aims to distinguish between generated images $G(x)$ and target samples y . For the mapping function G_2 , we also introduce a similar adversarial loss: $\mathcal{L}_{GAN}(G_2, D_X, Y, X)$.

2) *Cycle-Consistency Loss*: in the unpaired image-to-image transformation of our model, adversarial training can learn mappings G_1 and G_2 that generates outputs with similar style as target domains Y and X , respectively. However, the generated images may lose some 2D features (*e.g.*, semantic content, local texture) due to the random permutation, this problem can be solved by the cycle-consistency loss

$$\begin{aligned} \mathcal{L}_{cyc}(G_1, G_2) = & E_{x \sim P_{data}(x)} [\|G_2(G_1(x)) - x\|_1] \\ & + E_{y \sim P_{data}(y)} [\|G_2(G_1(y)) - y\|_1] \end{aligned} \quad (2)$$

3) *Depth Loss*: inspired by the image dehaze work [6], in addition to the blurring characteristics, monocular underwater images can also provide an intrinsic constructive clue: the camera-to-object distance of any image pixels corresponds to the degradation of image quality in these locations. (*i.e.*, the camera-to-object distance can be scaled by the degradation of underwater images). The degradation-derived distance provides a novel depth feature for underwater images. According to the light transmission model and followed by [6], in each color channel $c \in \{R, G, B\}$, the intensity of an underwater image at each pixel can be modeled as

$$I_c(\mathbf{x}, \lambda) = J_c(\mathbf{x}, \lambda) \cdot t_c(\mathbf{x}, \lambda) + B_c \cdot (1 - t_c(\mathbf{x}, \lambda)) \quad (3)$$

where \mathbf{x} is the pixel coordinate, λ is the wavelength, I_c is the pixel value of the image in color channel c , J_c is the unattenuated scene that is to be restored, B_c is the global veiling-light component, t_c is the medium transmission of color channel c . The red channel has the lowest transmission in the underwater environment [24], so we apply the DCP only on the blue and green channels of the underwater image. At least one color channel has some pixels whose intensity are very low and close to zero, we can get

$$J^{\text{dark}}(\mathbf{x}) = \min_{p \in \Omega(\mathbf{x})} \left(\min_{c \in \{g, b\}} J^c(p) \right) \rightarrow 0 \quad (4)$$

To estimate the parameter B_c , we employ quad-tree subdivision [25] to select the highest-intensity region of underwater images, as shown in Fig. 4, and then calculate the average pixel

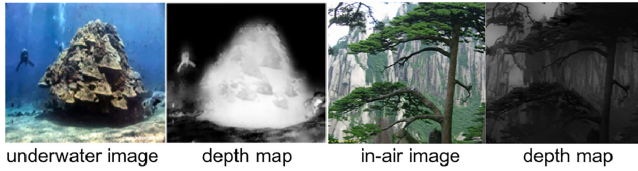


Fig. 5. Depth maps of the underwater images and the in-air images.

value of the selected region as the estimate result

$$B^c = \frac{1}{w_n \cdot h_n} \sum_{\mathbf{x} \in (I^{\text{dark}})_{n+1}} I^c(\mathbf{x}), c \in \{g, b\} \quad (5)$$

To estimate the t_c , the Eq. (3) can be expressed as

$$\frac{I^c(\mathbf{x})}{B^c} = \frac{J^c(\mathbf{x}) \cdot t^c(\mathbf{x})}{B^c} + 1 - t^c(\mathbf{x}) \quad (6)$$

Combining the Eq. (4) and Eq. (6), we can get

$$t^b(\mathbf{x}) = 1 - \min_{p \in \Omega(\mathbf{x})} \left(\min_{c \in \{g, b\}} \left(\frac{I^c(p)}{B^c} \right) \right) \quad (7)$$

According to the Lambert-Beer law [26], t_c can also be expressed as

$$t_c(\mathbf{x}, \lambda) = e^{-\beta(\lambda)d(\mathbf{x})} \quad (8)$$

For a given water medium, the attenuation coefficient $\beta(\lambda)$ is a constant, and then we can get the object-to-camera distance (depth value) of each pixel by

$$d(\mathbf{x}) = -\frac{1}{\beta(\lambda)} \ln t_c(\mathbf{x}) \quad (9)$$

Thus, we can get depth maps of underwater images according to the Eq. (9), as shown in Fig. 5.

In our method, a raw underwater image can be restored to its original state by the cyclic mapping $x \rightarrow y \rightarrow x'$. Theoretically, the depth values of any image pixel of x and x' should be the same after this mapping, thus we extract depth maps of underwater images at both ends of the cyclic mapping and take them as a novel 3D structural restriction to the cycle consistency by introducing a novel depth loss. It arouses the depth sensitivity of underwater images in 3D space. As shown in Fig. 2, for the depth map of image x , the G_1 and G_2 should satisfy forward and backward cycle consistency. We formulate this behavior by a depth loss

$$\begin{aligned} \mathcal{L}_{\text{depth}}(G_1, G_2) = & E_{x \sim Pdata(x)} [\|D(G_2(G_1(x))) - D(x)\|_1] \\ & + E_{y \sim Pdata(y)} [\|D(G_2(G_1(y))) - D(y)\|_1] \end{aligned} \quad (10)$$

where function $D(\cdot)$ is used to obtain depth maps of raw/enhanced images after the cycle consistency mapping.

C. Full Objective Function

The full objective function of our WSUIE method is

$$\begin{aligned} \mathcal{L}(G_1, G_2, D_X, D_Y) = & L_{GAN}(G_1, D_Y, X, Y) \\ & + L_{GAN}(G_2, D_X, Y, X) + \mathcal{L}_{\text{cycle}}(G_1, G_2) \\ & + \mathcal{L}_{\text{depth}}(G_1, G_2) \end{aligned} \quad (11)$$

In our method, we aim to solve

$$G_1^*, G_2^* = \arg \min_{G_1, G_2} \max_{D_X, D_Y} \mathcal{L}(G_1, G_2, D_X, D_Y) \quad (12)$$

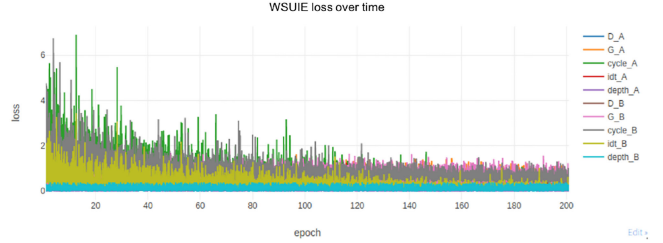


Fig. 6. Training loss of the WSUIE method on the UIIE dataset, D_A and D_B are the two adversarial losses, Cycle_A and Cycle_B constitute the cycle consistency loss, depth_A and depth_B constitute the depth loss.

IV. IMPLEMENTATION

A. UIIE Dataset

To train the proposed WSUIE method in a weakly supervised manner, we collected a large-scale dataset composed of unaligned image pairs from two different domains X and Y (X contains raw underwater images and Y contains arbitrary high-quality images). Previous works have proposed several datasets (e.g., EUVP [10], UIEB [7], etc.) for supervised network training of deep underwater image enhancement models. In the UIIE dataset, there are some images selected from the EUVP and UIEB dataset, and some high-quality in-air images are from the ImageNet [27] or downloaded from Google and Baidu Browsers. We also collected some raw underwater images during oceanic explorations in different locations and various illuminations. As a result, the UIIE dataset contains 4088 raw underwater images and 5629 high-quality images, and there are 50 raw underwater images and 50 high-quality images for the model test. As far as we know, the UIIE is the first public dataset dedicated to the research of weakly supervised underwater image enhancement.

B. Training Details

We implement the WSUIE model by using the PyTorch libraries, Python 3.8. It is trained on the UIIE dataset, one NVIDIA GeForce GTX 3090 graphics card is used, this model is trained for 200 epoch with a batch-size of 8. Other initial hyperparameter settings of model training refer to [12]. The training loss is shown in Fig. 6, and snapshots of the epoch 1, 10, 50, 100, 150, and 200 are shown in Fig. 7.

V. EXPERIMENTS AND EVALUATION

A. Qualitative Evaluation

We first perform a qualitative evaluation of our method on the test set of the UIIE dataset. The performance of our method is compared with the state-of-the-art (SOTA) models, including seven data-driven models: (i) underwater GAN with gradient penalty (UGAN-P) [18], (ii) Pix2Pix [23], (iii) least-squared GAN (LS-GAN [28]), (iv) GAN with residual blocks in the generator (Res-GAN) [29], (v) Wasserstein GAN with residual blocks in the generator (Res-WGAN) [30], (vi) CycleGAN, and (vii) FUnIE-GAN (UP) [10]); and two physic model-based methods: multi-band fusion-based enhancement (Mband-EN [31]) and haze-line-aware color restoration (Uw-HL [24]).

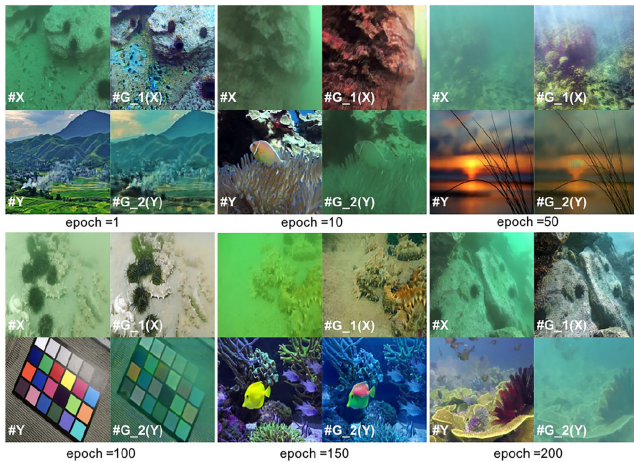


Fig. 7. Snapshots of some training epochs of the WSUIE based on the UIIE dataset: $\#X(Y)$ are input images from domain $X(Y)$, $\#G_1(X)$ and $\#G_2(Y)$ are images generated by the G_1 and G_2 , respectively.

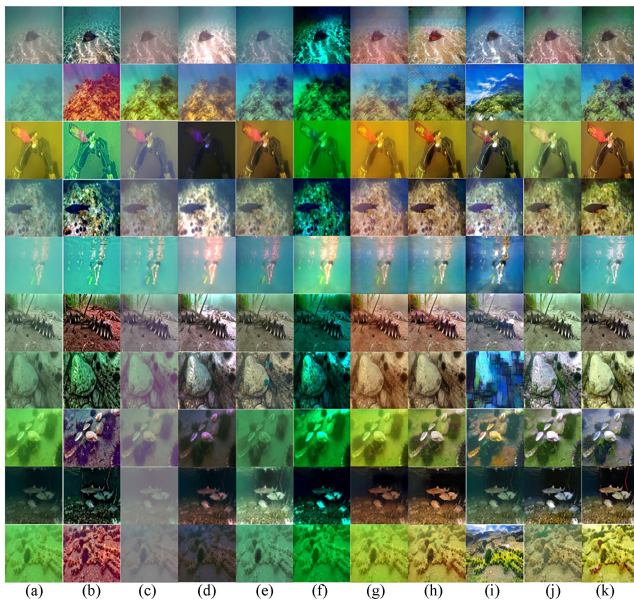


Fig. 8. Qualitative comparison between the WSUIE method and the SOTA methods in terms of underwater image enhancement performance. From left to right: a) raw underwater image; b) Mband-EN; c) Res-GAN; d) Res-WGAN; e) CycleGAN; f) LS-GAN; g) Uw-HL; h) UGAN-P; i) Pix2Pix; j) FUnIE-GAN; k) FUnIE-GAN-UP; l) WSUIE.

As shown in Fig. 8, all the enhanced underwater images generated by the SOTA method and our method can improve the primary characteristics of underwater images in terms of visual effects (qualitative evaluation metrics) to some extent. Compared with the SOTA methods, our enhanced images have clearer content and texture, richer colors, and more significant contrast. Additionally, the greenish hue and bluish hue in raw underwater images can be rectified by our method, which shows that our method has superiority in improving the primary visual features of underwater images. More specifically, three data-driven methods (CycleGAN, Pix2Pix, and FUnIE-GAN-UP) that have similar network architectures to our method are trained on the

TABLE I
QUANTITATIVE EVALUATION OF THE WSUIE ON PSNR AND SSIM

Methods	PSNR(\uparrow)	SSIM (\uparrow)
	Inputs: 17.27 ± 2.88	Inputs: 0.62 ± 0.075
Mband-EN	12.11 ± 2.55	0.4565 ± 0.097
Res-GAN	14.75 ± 2.22	0.4685 ± 0.122
Res-WGAN	16.46 ± 1.80	0.5762 ± 0.014
CycleGAN	17.14 ± 2.65	0.6400 ± 0.080
LS-GAN	17.83 ± 2.88	0.6725 ± 0.062
Uw-HL	18.85 ± 1.76	0.7722 ± 0.066
UGAN-P	19.59 ± 2.54	0.6685 ± 0.075
Pix2Pix	20.27 ± 2.66	0.7081 ± 0.069
FUnIE-GAN	21.36 ± 2.17	0.8164 ± 0.046
FUnIE-GAN-UP	21.92 ± 1.07	0.8876 ± 0.068
WSUIE	22.95 ± 1.74	0.8884 ± 0.026

TABLE II
QUANTITATIVE EVALUATION OF THE WSUIE ON UIQM AND UCIQM

Methods	UIQM (\uparrow)	UCIQE (\uparrow)
	Inputs: 1.8431	Inputs: 5.4725
Pix2pix	1.8054	6.1405
CycleGAN	1.9417	5.4800
FUnIE-GAN-UP	1.9290	6.3989
WSUIE	1.9435	6.4147

UIIE dataset, and their respective performances are shown in Fig. 8. Obviously, the enhanced underwater images generated by CycleGAN are more likely to be affected by the image style in the target domain. The Pix2Pix cannot handle greenish underwater images, while the FUnIE-GAN-UP can provide some high-quality enhanced images, but the performance is unstable. On the whole, our method obtains the best performance.

B. Quantitative Evaluation

In terms of quantitative evaluation, the performance of our method is compared with the SOTA methods from four quantitative evaluation metrics. The two full-reference metrics: peak signal-to-noise ratio (PSNR) and structural similarity (SSIM), and the two non-reference metrics: underwater image quality measure (UIQM) and underwater color image quality evaluation (UCIQE). To verify our method on the two full-reference metrics PSNR and SSIM, we apply our method to the test set (aligned image pairs) of the EUVP dataset to get the averaged PSNR and SSIM values. Then compare the performance of our method with that of the SOTA method. As shown in table I, our method can obtain the best performance of underwater image enhancement. To validate our method on the two non-reference metrics UIQM and UCIQE, the raw underwater images in the test set of the UIIE are enhanced by our method. Then we compare the performance of our method with three underwater image enhancement methods (CycleGAN, Pix2Pix, FUnIE-GAN-UP) designed in an unpaired image-to-image transformation manner. As shown in table II, our method achieves a slight improvement on both UIQM and UCIQE metrics.

C. User Study

This section conducts a user study to add human preference to the quantitative evaluations of our method. We apply four recently proposed data-driven methods and the WSUIE method to generate corresponding enhanced underwater images on the

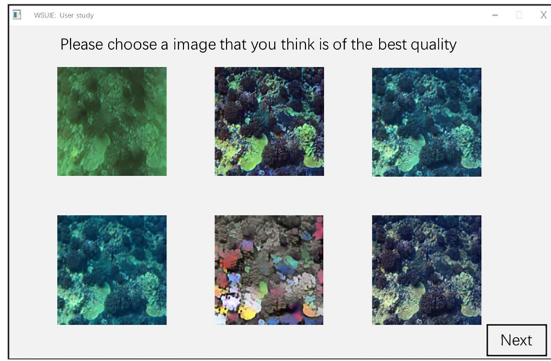


Fig. 9. Snapshots of the user study.

TABLE III
STATISTICS OF USERS' CHOICES FOR EACH METHOD

Pix2Pix	CycleGAN	FUnIE-GAN	FUnIE-GAN-UP	WSUIE
2%	6%	12%	14%	66%

test set. It provides us 50 groups of enhanced underwater images, one raw underwater image and its five enhanced images without semantic labels are taken as a group. Then a total of 17 subjects from different backgrounds are required to select the best-quality image from each group. As shown in Fig. 9, statistics of users' choices are displayed in table III. The public far more accepts the enhanced underwater images generated by our methods than other methods.

D. Improved Visual Perception

1) *Feature Extraction and Matching*: improved visual perception relies on rich basic visual features. In this evaluation, several feature descriptors (Canny [32]), Harris [33], and SIFT [34]) are extracted from raw/enhanced underwater images, and our WSUIE is compared with three SOTA methods from the perspective of basic feature extraction and matching. As shown in Fig. 10, all enhanced images have more edge contours and Harris features than raw underwater images. Compared with the three SOTA methods, the WSUIE can provide the enhanced underwater images with more basic vision features. As shown in Fig. 11, in raw underwater images, the correct feature matching based on SIFT feature rarely appears. With the assistance of the three methods and the WSUIE, more accurate SIFT feature matching appears, and the WSUIE performs the best.

2) *Object Detection*: To evaluate the performance of our WSUIE on the object detection task, the standard deep vision-based object detection model SSD [35] is trained for fish detection. We then apply it to the raw underwater images and the enhanced ones generated by the three SOTA methods and the WSUIE. As shown in Fig. 12, compared with the fish detection results on the raw underwater image and the enhanced images generated by the three SOTA methods, the SSD model has higher accuracy for the image enhanced by WSUIE, and for some challenging scenes (e.g., occlusion, overlap), the WSUIE can also achieve robust performance.

3) *Pose Estimation*: to verify the performance improvement the vision-based underwater human pose estimation task brought

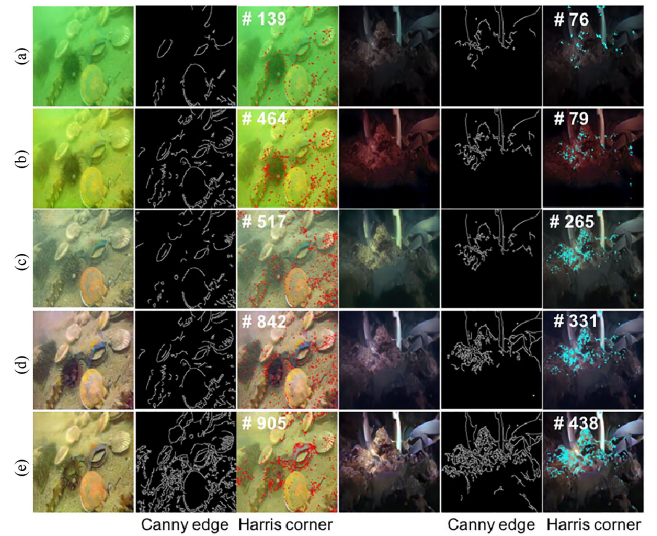


Fig. 10. Quantitative evaluation of the WSUIE based on feature extraction task: a) raw underwater images; b) Pix2Pix; c) CycleGAN; d) FUnIE-GAN-UP; e) WSUIE. #* is the number of the extracted Harris features.

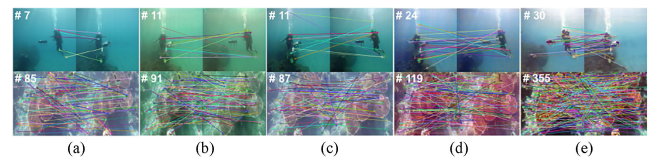


Fig. 11. Quantitative evaluation of the WSUIE based on feature matching task: a) raw underwater images; b) Pix2Pix; c) CycleGAN; d) FUnIE-GAN-UP; e) WSUIE. #* is the number of the matched SIFT features.

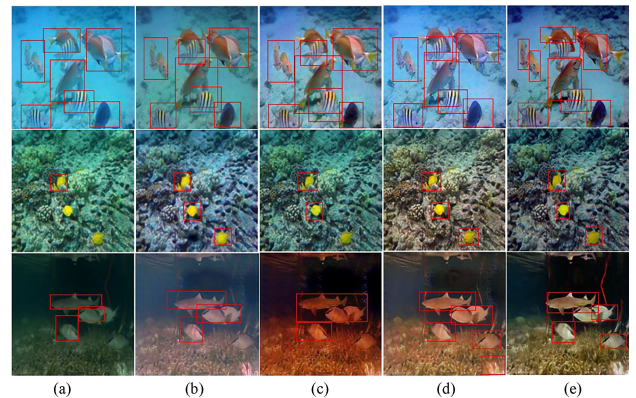


Fig. 12. Quantitative evaluation of the WSUIE based on the fish detection task: a) raw underwater images; b) Pix2Pix; c) CycleGAN; d) FUnIE-GAN-UP; e) WSUIE.

by our WSUIE, a comparison experiment is carried out based on the standard depth vision model [36], and the results are shown in Fig. 13. The enhanced underwater images can provide clearer image content and texture. Compared with the three SOTA methods, the vision-based human pose estimation results on the enhanced image generated by the WSUIE are more accurate.

4) *Saliency Prediction*: in our method, the camera-to-object distance (depth information) of each underwater image pixel is

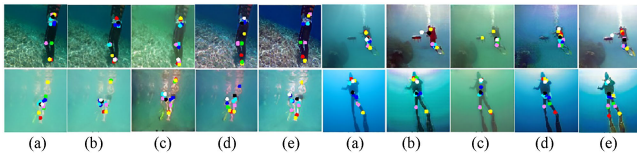


Fig. 13. Quantitative evaluation of the WSUIE based on the human pose estimation task: a) raw underwater images; b) Pix2Pix; c) CycleGAN; d) FUnIE-GAN-UP; e) WSUIE.

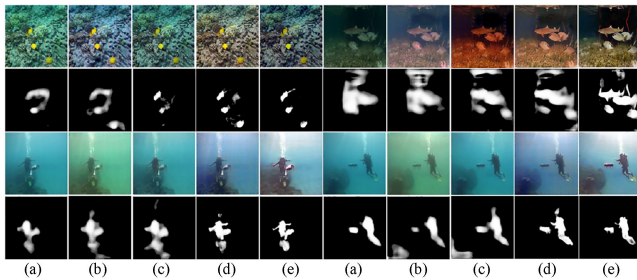


Fig. 14. Quantitative evaluation of WSUIE based on the visual saliency prediction task: a) raw underwater images; b) Pix2Pix; c) CycleGAN; d) FUnIE-GAN-UP; e) WSUIE.

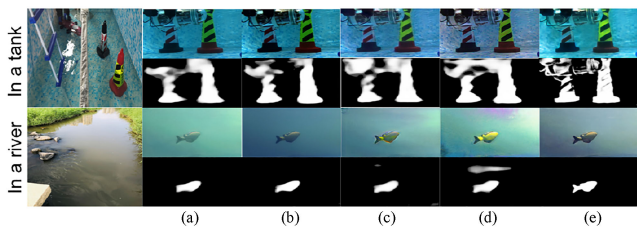


Fig. 15. Practical evaluation of the WSUIE based on the visual saliency prediction task: a) raw underwater images; b) Pix2Pix; c) CycleGAN; d) FUnIE-GAN-UP; e) WSUIE.

extracted and used to enhance underwater images with improved spatial structure. In order to evaluate the performance of our WSUIE on spatial structure-sensitive tasks such as underwater visual saliency prediction, we conducted some comparative experiments based on the raw underwater image and the enhanced images generated by the three SOTA methods. As shown in Fig. 14, the enhanced image generated by the WSUIE can provide the best performance in underwater visual saliency prediction tasks.

E. Practical Experiments

We equip an AUV with a waterproof camera, and conduct several practical experiments on underwater visual perception in rivers and tanks in the wild. As shown in the Fig. 15, To some extent, AUV can achieve improved performance of visual saliency prediction with the help of three SOTA methods and our WSUIE. Compared with the three SOTA methods, the WSUIE is more accurate, which shows that our method can effectively improve the underwater visual perception performance in practical underwater situations.

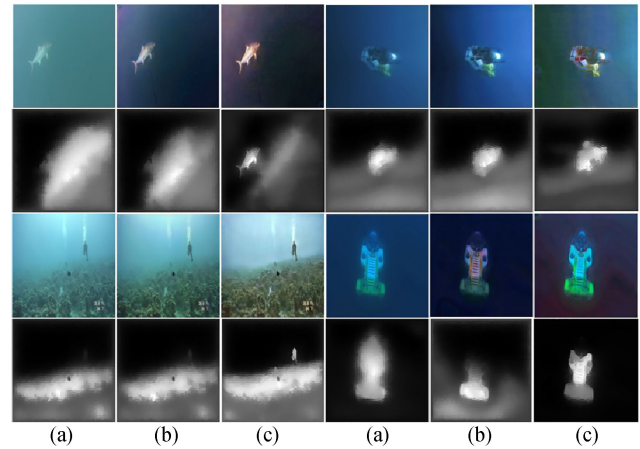


Fig. 16. Ablation study of the WSUIE method: a) raw underwater image; b) WSUIE without depth loss term; c) WSUIE with depth loss term.

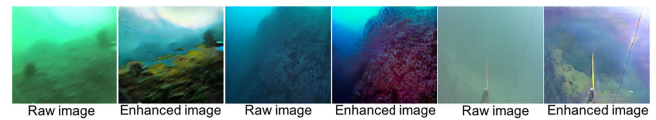


Fig. 17. Failure cases of the WSUIE method: not very effective for texture-less underwater images and images taken when the camera is moving fast.

F. Ablation Study

We perform an ablation study to further illustrate the advantage of the proposed depth loss term of the WSUIE method in underwater image enhancement. In our implementation, we train two kinds of WSUIE methods full objective function with/without the depth loss term based on the UUIE dataset, and compare the performance of the two methods on a depth sensitivity task: underwater visual saliency prediction. As shown in Fig. 16, removing depth loss substantially degrades the performance of saliency prediction.

G. Discussion

The proposed method has few failure cases, although it can achieve superior performance in most raw underwater images. Firstly, as shown in Fig. 17, the WSUIE method is not very effective for enhancing texture-less images and images taken when the camera is moving fast. The generated images in such cases are often oversaturated due to noise amplification and motion blur. Secondly, WSUIE is prone to training instability when high-quality in-air images have a strong background. In this case, the generated image lacks content consistency. It is a fairly common problem in unpaired training of GANs and requires meticulous hyper-parameter tuning. Finally, our method does not achieve a significant quantitative improvement of underwater images. Nevertheless, in most cases, the result of the unpaired image-to-image transformation is satisfactory. Furthermore, it suggests that the in-air images can be used in many data-driven underwater image enhancement works to bridge the gap between the in-air and underwater vision communities.

VI. CONCLUSION

This paper achieved underwater image enhancement in an unpaired image-to-image transformation manner, which gave birth to a novel weakly supervised underwater image enhancement method—WSUIE. This method formulates a new depth loss term, which is the first work that combines 2D and 3D visual features for underwater image enhancement. A large-scale dataset composed of unaligned image pairs is also provided for weakly supervised network training. Extensive qualitative and quantitative evaluations were conducted to verify the effectiveness of the WSUIE in improving the quality of underwater images and performances of some underwater visual perception tasks. In the future, we will further seek to improve the performance and stability of our method, and extend our method to unsupervised working methods, which will facilitate underwater exploration and protection tasks. Moreover, there is an interesting finding that our method can generate synthetic underwater images with the cyclic transformation. This inspires us to apply our method to some in-air image datasets and grasping datasets to generate many synthetic underwater images with annotated labels at low labor costs, thereby promoting the research of underwater visual perception and grasping.

ACKNOWLEDGMENT

The authors would like to thank the NSFC for offering us some underwater images.

REFERENCES

- [1] F. Bonin-Font, G. Oliver, and S. Wirth, "Visual sensing for autonomous underwater exploration and intervention tasks," *Ocean Eng.*, vol. 93, pp. 25–44, 2015.
- [2] M. Johnson-Roberson, M. Bryson, and A. Friedman, "High-resolution underwater robotic vision-based mapping and three-dimensional reconstruction for archaeology," *J. Field Robot.*, vol. 34, no. 4, pp. 625–643, 2017.
- [3] M. Cai, S. Wang, Y. Wang, R. Wang, and M. Tan, "Coordinated control of underwater biomimetic vehicle-manipulator system for free floating autonomous manipulation," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 51, no. 8, pp. 4793–4803, Aug. 2021.
- [4] B. McGlamery, "A computer model for underwater camera systems," in *Proc. Ocean Optics VI, Int. Soc. Opt. Photon.*, vol. 0208, 1980, pp. 221–231, doi: [10.1117/12.958279](https://doi.org/10.1117/12.958279).
- [5] J. S. Jaffe, "Computer modeling and the design of optimal underwater imaging systems," *IEEE J. Ocean. Eng.*, vol. 15, no. 2, pp. 101–111, Apr. 1990.
- [6] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.
- [7] C. Li, C. Guo, and W. Ren, "An underwater image enhancement benchmark dataset and beyond," *IEEE Trans. Image Process.*, vol. 29, pp. 4376–4389, 2020.
- [8] C. Li, J. Guo, and C. Guo, "Emerging from water: Underwater image color correction based on weakly supervised color transfer," *IEEE Signal Process. Lett.*, vol. 25, no. 3, pp. 323–327, Mar. 2018.
- [9] J. Li, K. A. Skinner, R. M. Eustice, and M. Johnson-Roberson, "WaterGAN: Unsupervised generative network to enable real-time color correction of monocular underwater images," *IEEE Robot. Automat. Lett.*, vol. 3, no. 1, pp. 387–394, Jan. 2018.
- [10] M. J. Islam, Y. Xia, and J. Sattar, "Fast underwater image enhancement for improved visual perception," *IEEE Robot. Automat. Lett.*, vol. 5, no. 2, pp. 3227–3234, Apr. 2020.
- [11] Z. Yi, H. Zhang, P. Tan, and M. Gong, "Dualgan: Unsupervised dual learning for image-to-image translation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2868–2876.
- [12] J. Zhu, T. Park, and P. Isola, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2242–2251.
- [13] M. Han, Z. Lyu, T. Qiu, and M. Xu, "A review on intelligence dehazing and color restoration for underwater images," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 50, no. 5, pp. 1820–1832, May 2020.
- [14] S. Anwar, C. Li, and F. Porikli, "Deep underwater image enhancement," vol. 98, 2020, Art. no. 107038.
- [15] X. Sun, L. Liu, Q. Li, J. Dong, E. Lima, and R. Yin, "Deep pixel-to-pixel network for underwater image enhancement and restoration," *IET Image Process.*, vol. 13, no. 3, pp. 469–474, 2019.
- [16] X. Liu, Z. Gao, and B. M. Chen, "MLFcGAN: Multilevel feature fusion-based conditional gan for underwater image color correction," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 9, pp. 1488–1492, Sep. 2020.
- [17] Y. Guo, H. Li, and P. Zhuang, "Underwater image enhancement using a multiscale dense generative adversarial network," *IEEE J. Ocean. Eng.*, vol. 45, no. 3, pp. 862–870, Jul. 2020.
- [18] C. Fabbri, M. J. Islam, and J. Sattar, "Enhancing underwater imagery using generative adversarial networks," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2018, pp. 7159–7165.
- [19] Mohcine, Boudhane, Nsiri, and Benayad, "Underwater image processing method for fish localization and detection in submarine environment," *J. Vis. Commun. Image Repres.*, vol. 29, pp. 226–238, 2016.
- [20] X. Chen, J. Yu, S. Kong, Z. Wu, X. Fang, and L. Wen, "Towards real-time advancement of underwater visual quality with GAN," *IEEE Trans. Ind. Electron.*, vol. 66, no. 12, pp. 9350–9359, Dec. 2019.
- [21] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, Springer, 2016, pp. 694–711.
- [22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [23] P. Isola, J. Zhu, and T. Zhou, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 5967–5976.
- [24] D. Berman, D. Levy, S. Avidan, and T. Treibitz, "Underwater single image color restoration using haze-lines and a new quantitative dataset," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 8, pp. 2822–2837, Aug. 2021.
- [25] C. Li, J. Guo, S. Chen, Y. Tang, Y. Pang, and J. Wang, "Underwater image restoration based on minimum information loss principle and optical properties of underwater imaging," in *Proc. IEEE Int. Conf. Image Process.*, 2016, pp. 1993–1997.
- [26] H. R. Gordon, "Can the lambert-beer law be applied to the diffuse attenuation coefficient of ocean water?" *Limnol. Oceanogr.*, vol. 34, no. 8, pp. 1389–1409, 1989.
- [27] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 248–255.
- [28] X. Mao, Q. Li, and H. Xie, "Least squares generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2813–2821.
- [29] J. Li, X. Liang, Y. Wei, T. Xu, J. Feng, and S. Yan, "Perceptual generative adversarial networks for small object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1951–1959.
- [30] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. 34th Int. Conf. Mach. Learn.* 2017, pp. 214–223.
- [31] Y. Cho, J. Jeong, and A. Kim, "Model-assisted multiband fusion for single image enhancement and applications to robot vision," *IEEE Robot. Automat. Lett.*, vol. 3, no. 4, pp. 2822–2829, Oct. 2018.
- [32] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.
- [33] C. G. Harris and M. J. Stephens, "A combined corner and edge detector," in *Proc. Alvey Vis. Conf.*, 1988, pp. 147–151.
- [34] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
- [35] W. Liu *et al.*, "Ssd: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, Springer, 2016, pp. 21–37.
- [36] E. Insafutdinov *et al.*, "Artrack: Articulated multi-person tracking in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 6457–6465.