Real-time GAN-based image enhancement for robust underwater monocular SLAM

Ziqiang Zheng[1†], Zhichao Xin[2†], Zhibin Yu[2,3*] and Sai-Kit Yeung[1]

[1]Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong, Hong Kong SAR, China, [2]Faculty of Information Science and Engineering, Ocean University of China, Qingdao, China, [3]Sanya Oceanographic Institution, Sanya, China

Underwater monocular visual simultaneous localization and mapping (SLAM) plays a vital role in underwater computer vision and robotic perception fields. Unlike the autonomous driving or aerial environment, performing robust and accurate underwater monocular SLAM is tough and challenging due to the complex aquatic environment and the collected critically degraded image quality. The underwater images' poor visibility, low contrast, and color distortion result in ineffective and insufficient feature matching, leading to the poor or even failure of the existing SLAM algorithms. To address this issue, we propose introducing the generative adversarial network (GAN) to perform effective underwater image enhancement before conducting SLAM. Considering the inherent real-time requirement of SLAM, we conduct knowledge distillation to achieve GAN compression to reduce the inference cost, while achieving high-fidelity underwater image enhancement and real-time inference. The real-time underwater image enhancement acts as the image pre-processing to build a robust and accurate underwater monocular SLAM system. With the introduction of real-time underwater image enhancement, we can significantly promote underwater SLAM performance. The proposed method is a generic framework, which could be extended to various SLAM systems and achieve various scales of performance gain.

KEYWORDS

generative adversarial networks, SLAM, knowledge distillation, underwater image enhancement, real-time, underwater SLAM

1 Introduction

Recently, many vision-based state estimation algorithms have been developed based on the monocular, stereo, or multi-camera systems in indoor (García et al., 2016), outdoor (Mur-Artal and Tardós, 2017; Campos et al., 2021), and underwater environments Rahman et al. (2018); Rahman et al. (2019b). Underwater SLAM (Simultaneous Localization and Mapping) is an autonomous navigation technique used by underwater robots to build a map of an unknown environment and localize the robot within the map. Underwater SLAM provides a safe, efficient, and cost-effective way to explore and survey unknown

underwater environments. Specifically, monocular visual SLAM provides an effective solution to many navigation applications Bresson et al. (2017), detecting unknown environments and assisting in decision-making, planning, and obstacle avoidance based on only a single camera. Monocular cameras are the most common vision sensors, which are inexpensive and ubiquitous mobile agents, making them a popular choice of sensor for SLAM.

There has been increasing attention on using an autonomous underwater vehicle (AUV) or remotely operated underwater vehicle (ROV) to conduct the monitoring of marine species migration Buscher et al. (2020) and coral reefs Hoegh-Guldberg et al. (2007), the inspection of submarine cables and wreckage Carreras et al. (2018), deep ocean exploration Huvenne et al. (2018) and underwater cave exploration Rahman et al. (2018); Rahman et al. (2019b). Unlike atmospheric imaging, the captured underwater images have issues with low contrast and color distortion due to the strong scattering and absorption phenomena. In detail, underwater pictures are usually critically degraded due to large suspended particles, poor visibility, and under-exposure. Thus it is complex and challenging to detect robust features to track for visual SLAM systems. As a result, directly performing the current available vision-based SLAM usually cannot obtain a satisfactory and robust result.

To address this issue, Cho et al. Cho and Kim (2017) combined Contrast-limited Adaptive Histogram Equalization (CLAHE) Reza (2004) to conduct real-time underwater image enhancement to promote the underwater SLAM performance. Furthermore, Huang et al. Huang et al. (2019) performed underwater image enhancement by converting RGB images to HSV space and then performing color correction based on Retinex theory. Then the enhanced outputs were applied for downstream underwater SLAM. However, these methods only achieved marginal improvement and could not work in highly turbid conditions.

Generative adversarial networks (GANs) Goodfellow et al. (2014) had been adopted for underwater image enhancement Anwar and Li (2020); Islam et al. (2020a) to boost underwater vision perception. Compared 48 with the model-free enhancement methods Drews et al. (2013); Huang et al. (2019), GAN-based image-to image (I2I) translation algorithms could enhance textile and content representations and generate realistic images with clear and plausible features Ledig et al. (2017), especially in highly turbid conditions Han et al. (2020); Islam et al. (2020c). This line of research has mostly taken place in the computer vision fields, with the main focus on underwater single image restoration Akkaynak and Treibitz (2019); Islam et al. (2020b). Benefiting from the superior performance of GAN-based approaches, some researchers attempted to use CycleGAN to boost the performance of ORB-SLAM in an underwater environment Chen et al. (2019). The experimental results have shown that CycleGAN-based underwater image enhancement can lead to more matching points in a turbid environment. However, CycleGAN Zhu et al. (2017) could not meet the real-time requirement. Nevertheless, CycleGAN-based underwater image enhancement may also increase the risk of incorrect matching pairs. Besides, the feature matching analysis and detailed quantitative SLAM results are missing in Chen et al. (2019) for discussing the potential of adopting the underwater enhancement for promoting underwater SLAM performance in real-world underwater environments. To address these issues, we aim to comprehensively analyze this point.

In this paper, we target to perform a lightweight GAN-based image enhancement framework for underwater monocular SLAM to promote performance. The proposed GAN-based image enhancement can promote the feature matching performance (Please refer to section 4.3.2 and Figure 1 for more details), which can further lead to better and more robust SLAM results. To speed

FIGURE 1
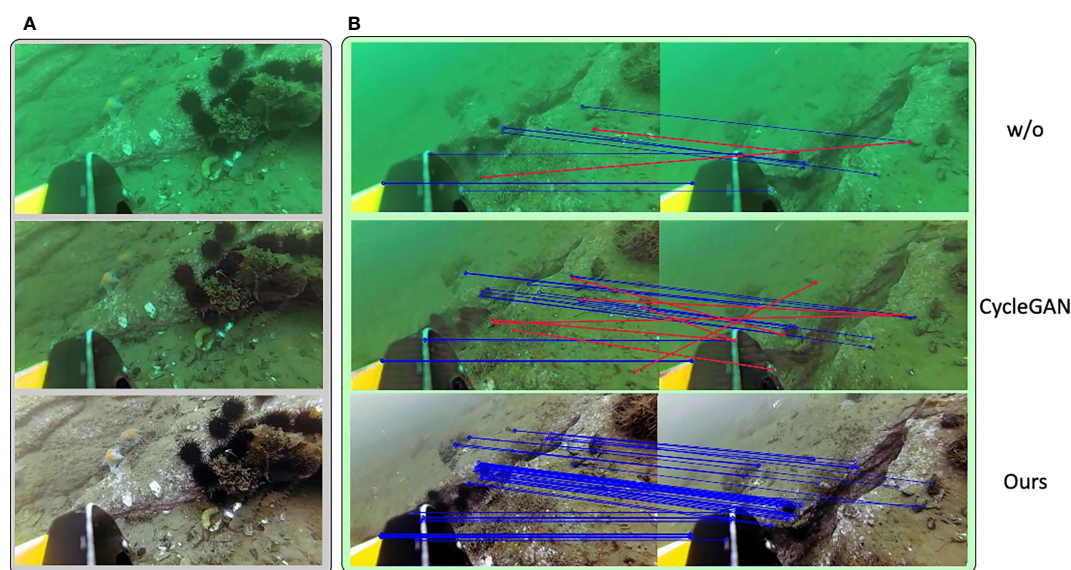The illustrations of **(A)** underwater images and **(B)** the ORB [41] feature-matching under three settings: without any enhancement, with CycleGAN, and with our method. Blue lines represent correct feature matching pairs and Red lines represent incorrect feature matching pairs. The proposed underwater enhancement can significantly promote feature-matching performance. Best viewed in color.

up the underwater image enhancement progress and reduce the risk of incorrect matching pairs, we propose to perform GAN compression Li et al. (2020b) to accelerate underwater image enhancement inference. The knowledge distillation Aguinaldo et al. (2019) is adopted to reduce the computational costs and the inference time. We propose a generic robust underwater SLAM framework shown in Figure 2, which could be extended to various SLAM systems (e.g., ORB-SLAM2 Mur-Artal and Tardós (2017), Dual-SLAM Huang et al. (2020) and ORB-SLAM3 Campos et al. (2021)) and achieve a performance gain with the real-time GAN-based underwater image enhancement module. The proposed method performs favorably against state-of-the-art methods in both position estimation and system stability. To sum up, our main contributions are listed as follows:

- We introduce a generic robust underwater monocular SLAM system, which can be extended to different SLAM algorithms and achieve a large performance gain.

- To accelerate GAN-based image enhancement, we perform GAN compression through knowledge distillation for performing real-time underwater image enhancement as a compelling image pre-processing module. As a result, we can obtain more robust, stable, and accurate state estimation outputs.

- Our method can achieve current state-of-the-art performance and tailored analysis about 1) underwater image enhancement, 2) robust and accurate feature matching, and 3) SLAM performance is included in our paper.

2 Related work

2.1 Underwater image enhancement

The underwater image enhancement algorithms could mainly fall into three categories: 1) model free Asmare et al. (2015); 2) model-based Akkaynak and Treibitz (2019) and 3) data-driven Li et al. (2018); Islam et al. (2020b); Islam et al. (2020c) algorithms. The representative model-free CLAHE Reza (2004) method could enhance an underwater image without the image formation process. Asmare et al. Asmare et al. (2015) converted the images into the frequency domain and proposed to enhance the high-frequency component to promote the image quality. Though these model-free methods could perform image enhancement with a very high speed, they still heavily suffered from over-enhancement, color distortion, and low contrast Li et al. (2020a), and they only achieved slight improvement under highly turbid conditions. Model-based methods considered the physical parameters and formulated an explicit image formation process. Drews et al. Drews et al. (2013) proposed to apply dark channel prior He et al. (2010) in the underwater setting to perform underwater dehazing. The Sea-Thru method Akkaynak and Treibitz (2019) firstly proposed to estimate the backscattering coefficient and then recover the color information with the known range based on RGB-D images. However, collecting a large-scale underwater RGB-D image dataset is expensive and time-consuming. The latter data-driven underwater image enhancement algorithms Li et al. (2018); Han et al. (2020); Islam et al. (2020b); Islam et al. (2020c) combined deep CNNs to conduct underwater image restoration based on large-scale paired or unpaired data. UWGAN Li et al. (2018) proposed to combine multi-style underwater image synthesis for the underwater depth estimation. SpiralGAN Han et al. (2020) proposed a spiral training strategy to promote image enhancement performance. FUnIE-GAN Islam et al. (2020c) could perform real-time underwater image enhancement for underwater object detection. Unlike this object-level enhancement algorithm, we target to perform real-time GAN-based underwater image enhancement for a more challenging underwater SLAM, which requires high-fidelity pixel correspondences.

2.2 Underwater SLAM

The popular ORB-SLAM Mur-Artal and Tardós (2017); Elvira et al. (2019) introduced an efficient visual SLAM solution based on
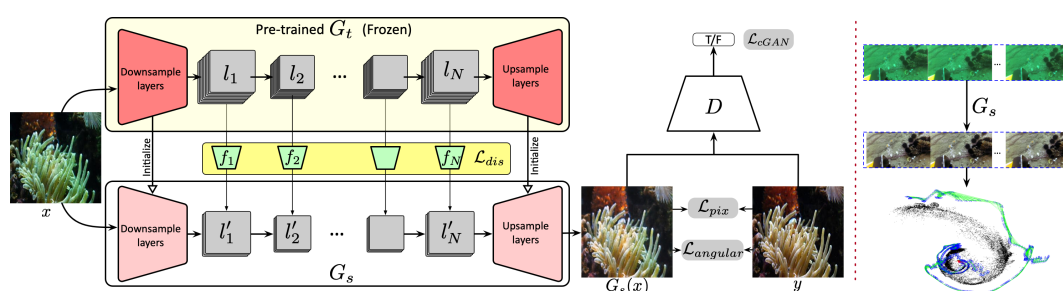

FIGURE 2
The overview framework of the proposed method. The left part on the red dotted line illustrates the GAN-based underwater image enhancement module. In contrast, the right part shows the downstream underwater monocular SLAM with learned G_s for real-time inference. G_t and G_s indicate the teacher and student generators, respectively. Given a pre-trained G_t, we aim to conduct GAN compression through knowledge distillation. $\{l_n\}_1^N$ represent the chosen N layers of G_t to compress and $\{l_n'\}_1^N$ indicate the compressed layers of G_s. f_n indicates the additional convolutional layer to achieve channel reduction to achieve shape matching between the intermediate layers of G_t and G_s. $\mathcal{L}_{dis}$ is computed to transfer the learned knowledge of G_t to G_s. We compute pixel-wise loss $\mathcal{L}_{pix}$, angular loss $\mathcal{L}_{angular}$ and conditional GAN loss $\mathcal{L}_{cGAN}$ between $G_s(x)$ and y.

ORB feature descriptor Rublee et al. (2011). VINS Qin et al. (2018); Qin and Shen (2018) proposed a general monocular framework with the IMU information. Unlike the aerial setting, underwater is a typical global positioning system (GPS) denied environment, where visual information provides valuable navigation queues for robot navigation. Currently, without the GPS for camera pose ground truth generation, a recent work Ferrera et al. (2019a) adopted Colmap Schönberger and Frahm (2016); Schönberger et al. (2016) to generate relatively precise camera trajectory based on structure-from-motion (SFM). UW-VO Ferrera et al. (2019b) further adopted the generated trajectory as ground truth to evaluate the underwater SLAM performance. Because of the good properties of sound 116 propagation in the water, some sonar-based methods Rahman et al. (2018); Rahman et al. (2019a); Rahman et al. (2019b) (e.g., SVIN Rahman et al. (2018) and SVin2 Rahman et al. (2019b)) combined the additional sparse depth information from the sonar sensor to perform more accurate position estimation. However, these are more suited for long-range underwater missions rather than close-range ones. Besides, the sonar sensor is still expensive, and we target to propose a general underwater SLAM framework based on the visual information.

3 Methodology

3.1 Overall framework

We aim to propose a generic robust underwater monocular SLAM framework, which contains two main procedures: **Real-time GAN-based Underwater Image Enhancement** and **Downstream Underwater SLAM** based on the enhanced underwater images generated from the former stage. First, we refer the readers to check the overall framework in Figure 2. To perform real-time GAN-based I2I translation for underwater image enhancement, we adopt the knowledge distillation Aguinaldo et al. (2019)for GAN compression to achieve better performance-speed tradeoff. The network parameters and computational costs could be heavily reduced after compression while achieving comparable or even better underwater enhancement performance.

3.2 GAN-based underwater image enhancement

To achieve underwater image enhancement from a source domain $\mathbb{X}$ to a target domain $\mathbb{Y}$ (*e.g.*, the source turbid underwater image domain and another target clear underwater image domain). The conditional GAN pipeline Mirza and Osindero (2014) is chosen in our work since it could generate more natural and realistic image outputs based on full supervision. For generating reasonable image outputs in the target domain, the adversarial loss is applied:

$$\mathcal{L}_{cGAN} = \mathbb{E}_{x,y}[\log \ D(x,y)] + \mathbb{E}_x[\log \ (1 - D(x, G_s(x)))], \quad (1)$$

where x and y are image samples from $\mathbb{X}$ and $\mathbb{Y}$, respectively. The adversarial loss $\mathcal{L}_{cGAN}$ Isola et al. 137 (2017) could reduce the distance between the generated sample distribution and the real sample distribution. Besides the adversarial loss, the pixel-wise $\mathcal{L}_{pix}$ is also included to measure the pixel difference (1-norm) between the generated image output and the corresponding real clear image:

$$\mathcal{L}_{pix} = \| \ G_s(x) - y \ \|_1; \quad (2)$$

please note that we compute both $\mathcal{L}_{pix}$ and $\mathcal{L}_{cGAN}$ based on the output of G_s rather than G_t.

Angular loss. To further promote the naturalness of synthesized outputs, we adopt the angular loss $\mathcal{L}_{angular}$ Han et al. (2020) to obtain better image synthesis:

$$\mathcal{L}_{augular} = \mathbb{E}_{\mathbb{X},\mathbb{Y}}[\angle \ (G_s(x), y)], \quad (3)$$

where $\angle$ indicates the angular distance between $G_s(x)$ and y in RGB space. It is observed that the used $\mathcal{L}_{angular}$ could lead to better robustness and enhancement outputs to some critical over-under exposure problems in the underwater images. The color distortion could be effectively alleviated by $\mathcal{L}_{angular}$. Through the integration of the above-mentioned loss functions, we could achieve effective and reasonable underwater image enhancement. However, it cannot meet the real-time inference requirement in Isola et al. (2017); Han et al. (2020); Zhu et al. (2017).

3.3 GAN compression through knowledge distillation

We perform GAN compression through knowledge distillation to save computational costs and achieve the tradeoff between enhancement performance and inference speed. The detailed design of the proposed GAN compression module is shown in Figure 2, which contains the teacher generator G_t, the student generator G_s, and the discriminator D. In detail, we transfer the learned knowledge learned from G_t to G_s by matching the distribution of the feature representations. We initialize the teacher network G_t with a pre-trained underwater enhancement model and G_t is frozen during the whole training procedure. The optimized teacher network could guide the student network on extracting effective feature representations and achieving better enhancement performance. The distillation objective can be formulated as:

$$\mathcal{L}_{dis} = \sum_{n=1}^{N} \| f_n(G_s^n(x)) - G_t^n(x) \|_2, \quad (4)$$

where $G_s^n(x)$ and $G_t^n(x)$ (*with channel number c = 16*) are the intermediate feature representations of the n-th selected feature layer in G_s and G_t, and N denotes the number of selected layers. f_n is the convolutional layer with 1×1 kernel to achieve channel reduction, which will not introduce many training parameters. In our experiments, we set $N = 6$ and select the middle intermediate feature representations. Different from Li et al. (2020b), we do not perform a neural architecture search (NAS) considering its huge time consumption. The channel number in G_s is set to 16. More

ablation studies about the channel number c selection can be found in Sec. 4.4.

3.4 Full objective function

We update the final objective function of the proposed method as:

$$\mathcal{L} = \mathcal{L}_{cGAN} + \mathcal{L}_{dis} + \mathcal{L}_{angular} + \lambda \mathcal{L}_{pix}, \quad (5)$$

where λ is a hyper-parameter to balance the loss component. We set $\lambda = 10$ in our experiments following the setup in Isola et al. (2017) to better balance the contribution of pixel-wise supervision and other components in the proposed method.

3.5 Downstream underwater SLAM

For the downstream monocular SLAM module, we have explored different in-air SLAM systems: **ORB-SLAM2**, **Dual-SLAM** and **ORB-SLAM3** to perform state estimation based on the enhanced underwater images after the image resizing for obtaining the approximate image inputs. To be noted, the two modules are optimized separately and the SLAM system is running in a hardcore engineering manner. The in-air visual SLAM algorithms underperform in the aquatic environment as the critical image degradation. With the real-time GAN-based underwater image enhancement module, the model could better model the complex marine environment and find robust features to track from the enhanced underwater images, which leads to more stable and continuous SLAM results. Besides, the proposed framework could be extended to various SLAM systems to achieve performance gain.

4 Experiments

In this section, we first provide the implementation details of the proposed method and review the experimental setup. Then we report the inference speed comparison of different underwater image enhancement algorithms, followed by the detailed performance comparison of different algorithms. Next, we target to analyze the underwater SLAM performance from three aspects: 1) underwater image enhancement performance, 2) feature matching analysis, and 3) qualitative and quantitative underwater SLAM performance of different SLAM baselines under various settings. Finally, we provide ablation studies to explore the tradeoff between the underwater enhancement performance and the inference speed.

4.1 Implementation details and experimental setup

4.1.1 Implementation details

To obtain a lightweight and practical underwater enhancement module, we perform the knowledge distillation to compress the enhancement module to meet the real-time inference requirement. The trained SpiralGAN model (also other GAN models) is chosen as the teacher model G_t to stabilize the whole training procedure and speed up the convergence. It is worth noting that G_t is frozen when performing the knowledge distillation. The image resolution of underwater enhancement is set to 256×256, and we perform upsampling to resize the enhanced image outputs to 640×480 based on bilinear interpolation for further SLAM. The hyperparameter c for selected feature layers is set to 16, and we include the discussion about choices of c in our ablation studies. For optimizer, we choose Adam optimizer Kingma and Ba (2014) in all our experiments and set the initial learning rate to 0.0002.

4.1.2 Datasets
4.1.2.1 Training datasets for underwater image enhancement

We adopt the training dataset from the previous work Fabbri et al. (2018), which contains 6,128 paired underwater turbid-clear images synthesized from CycleGAN Zhu et al. (2017). The proposed method has been only trained with one underwater dataset and can be extended to different unseen underwater image sequences for performing underwater image enhancement.

4.1.2.2 Datasets for underwater SLAM

URPC dataset contains a monocular video sequence collected by the ROV in a real aquaculture farm. The ROV navigates at a water depth of about 5 meters. Operating ROV collected a total of 190 seconds of a video sequence with an acquisition frequency 24Hz. A total of 4,538 frames of RGB images (640×352 image resolution) were obtained. The collected video sequence has large scene changes and low water turbidity. The image suffers severe distortion, and the watercolor is bluish-green. Considering the first 2,000 consecutive images do not contain meaningful objects, we remove them and only choose the last 2,538 images for experimental testing. We choose the open-source offline SFM library Colmap Schönberger and Frahm (2016); Schönberger et al. (2016) to generate the camera pose trajectory for evaluating the underwater SLAM performance. OUC fisheye Zhang et al. (2020) dataset is a monocular dataset collected by the fisheye camera in a highly turbid underwater environment. It provides 10 image sequences from three water turbidity: 1) slight water turbidity with about 6m visibility; 2) middle water turbidity with about 4m visibility and 3) high water turbidity with about 2m visibility. The image sequences are collected with the acquisition frequency of 30Hz and each sequence lasts about 45 seconds. In our experiments, we evaluate an image sequence containing 1,316 frames ($1,920 \times 1,080$ image resolution) with high water turbidity. The severe distortion and heavy backscattering lead to a significant influence on feature tracking. The trajectory is also generated from Colmap Schönberger and Frahm (2016); Schönberger et al. (2016).

4.1.3 SLAM baselines

We chose three baselines: ORB-SLAM2, Dual-SLAM and ORB-SLAM3 for comparison:

- ORB-SLAM2 Mur-Artal and Tardós (2017) is a complete SLAM system for monocular, stereo, and RGB-D cameras. The adopted ORB-SLAM2 system has various applications for indoor and outdoor environments. We choose it as the baseline for performing underwater mapping and reconstruction.
- Dual-SLAM Huang et al. (2020) extended ORB-SLAM2 to save the current map and activate two new SLAM threads: one is to process the incoming frames for creating a new map and another is to link the created new map and older maps together for building a robust and accurate system.
- ORB-SLAM3 Campos et al. (2021) perform visual, visual-inertial, and multi-map SLAM based on monocular, stereo, and RGB-D cameras, which has achieved current state-of-the-art performance and provided a more comprehensive analysis system.

4.1.4 Evaluation metric for SLAM

To measure the SLAM performance, we choose 1) Absolute Trajectory Error (ATE), 2) Root Mean Square Error (RMSE), and 3) Initialization performance for evaluation. ATE directly calculates the difference between the camera pose ground truth and the estimated trajectory from SLAM. RMSE can describe the rotation and translation errors of the two trajectories. The smaller the RMSE is, the better the system trajectory fits. The initialization performance indicates the number of frames to perform the underwater SLAM initialization. The lower the initialization frames, the better SLAM performance, and more stable and continuous outputs. To make a fair comparison, we repeat the underwater SLAM experiments 5 times to obtain the best result for all methods.

4.2 Inference speed comparison

In this section, we target to provide the inference speed comparison of different underwater image enhancement methods under the same experimental setting. For underwater image enhancement methods, we choose CLAHE Reza (2004), UDCP Drews et al. (2013) and FUnIE-GAN Islam et al. (2020c) for underwater image enhancement comparison. To measure the frames per second (FPS) for different methods, we test the speed of

various methods on the practical Jetson AGX Xavier, which is widely equipped on underwater ROVs and AUVs. The detailed FPS and memory access cost (MAC) comparison is shown in Table 1 (the testing image resolution is set to 640×480 (default image resolution of ORB-SLAM2 and ORB-SLAM3) for all methods to make a fair comparison). Compared with the default image resolution (256×256) adopted in FUnIE-GAN, the proposed FPS computation setting is more practical and can lead to more reasonable and accurate translated outputs. As reported, UDCP has a very low underwater image enhancement speed, and it costs several seconds to process only one image. Besides, SpiralGAN and FUnIE-GAN cannot perform real-time (e.g., ≥ 30) underwater image enhancement. Our method has fewer network parameters and can achieve real-time GAN-based underwater image enhancement.

4.3 Performance comparison

4.3.1 Underwater image enhancement results

Firstly, we target to demonstrate that the proposed method could generate high-quality image synthesis outputs after the underwater image enhancement module. We have provided a direct comparison with the model-free image enhancement algorithms (CLAHE and UDCP) and GAN-based image enhancement method (FUnIE-GAN) in Figure 3 on the URPC dataset. Compared with the previous model-free image enhancement methods, the proposed method could enhance the content representations of the objects. The synthesis image by FUnIE-GAN has visible visual artifacts. In contrast, the proposed GAN-based image enhancement method could achieve better results with more reasonable outputs. To be noted, the proposed method has been only trained on one underwater dataset and can be extended to different unseen underwater image sequences for testing. The strong generalization ability could alleviate the efforts of the model-based algorithms to change the physical parameters, which is also time-consuming. The GAN-based image enhancement module has shown powerful effectiveness and achieved better results. We provide more underwater image enhancement result comparisons in our supplementary.

4.3.2 Feature matching analysis

We have designed comprehensive feature-matching experiments to reveal whether the proposed underwater image enhancement

TABLE 1   Quantitative FPS, MACs(G) and Parameter(M) comparison of various methods.

Method	FPS ↑	MACs (G) ↓	Parameter(M) ↓
CLAHE	**260.0**	–	–
UDCP	0.041	–	–
FUnIE-GAN	7.36	47.96	7.02
CycleGAN	3.25	266.40	11.38
SpiralGAN	17.63	34.75	4.99
Proposed	31.83	**8.81**	**1.28**

The image resolution is set to 640×480 for evaluation based on practical Jetson AGX Xaviver. ↑ (↓) indicates that the larger (smaller) the value is, the better the performance. The best results are in bold. "-" means not applicable.
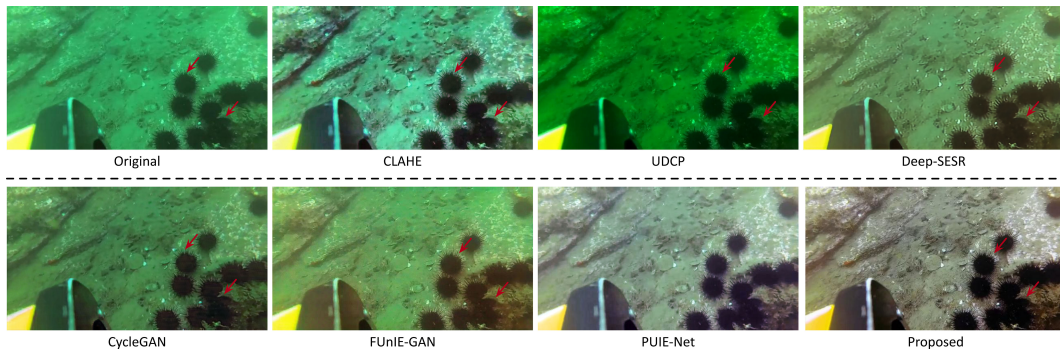
FIGURE 3
The qualitative results of different underwater image enhancement methods. Best viewed in color.

could promote the feature-matching performance for SLAM. First, following the experimental setup in Cho and Kim (2017), we report the ORB, SIFT, and SURF feature-matching results. For a fair comparison, 500 same image pairs are chosen for performing feature matching based on various feature descriptors with two different frame intervals: 20 and 30. If the matching points number is larger than 50, we regard the matching as successful and report the successful matching rate. The detailed results are illustrated in Figure 4. Besides, we also provide the average number of matching points of different feature descriptors. Compared with feature matching performance conducted on the original images, UDCP Drews et al. (2013) could only lead to marginal improvement or slight degradation. FUnIE-GAN Islam et al. (2020c) failed to generate reasonable enhanced image outputs with plausible textile

information. There is an observable performance degradation compared with the "original" setting. In contrast, the proposed method can improve performance under all settings.

Furthermore, to verify that the yielded feature matching points are valid interior points, we conduct feature point matching evaluation through reprojection. In detail, the feature points extracted from the current frame are reprojected to the previous 20th image frame. We obtain the ground truth feature matching based on Structure-from-Motion. For defining accurate feature matching points; we choose a 3×3 pixel area:

- When the distance between the projected point (computed based on the estimated transformation matrix H and the intrinsic camera parameter K) and the detected feature
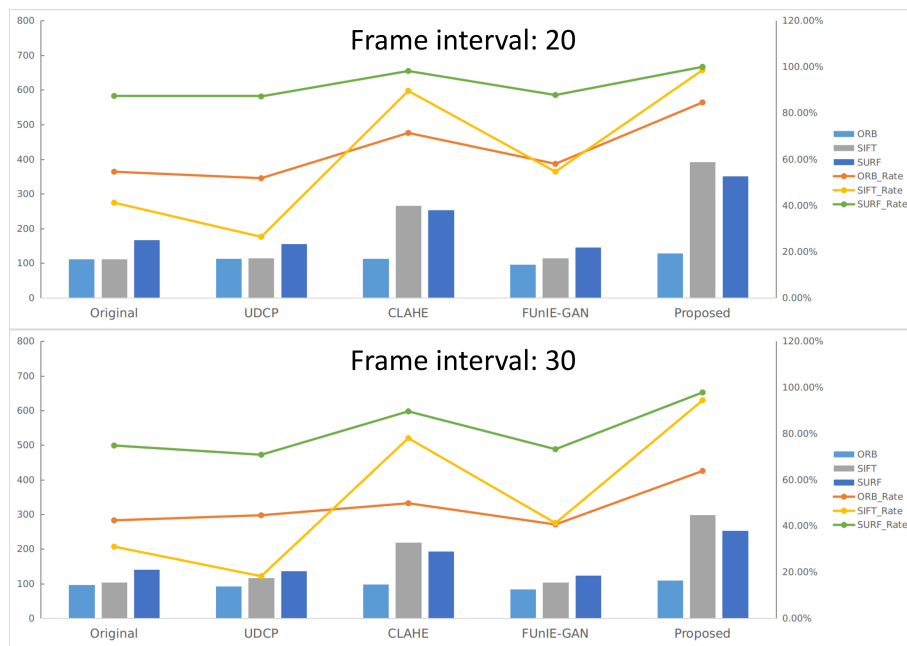

FIGURE 4
The qualitative feature matching results of various methods based on different feature descriptors. The lines and the bars indicate the feature matching success rates and average matching points based on various feature descriptors, respectively.

point is less than 3, such detected feature matching points are marked as tbfInner points (denoted as P_i),

- Other detected feature points are falsely matched as Outlier points (denoted as P_s).

We have provided the qualitative feature matching performance under three settings: 1) w/o underwater enhancement, 2) enhancement by CycleGAN, and 3) our method in Figure 1. As reported, CycleGAN adopted in Chen et al. (2019) a as a pre-processing module could increase the number of correct matching pairs. However, the number of incorrect matching pairs also increased. The proposed method can significantly increase the number of correct matching pairs with few errors.

For the quantitative comparison, we compute the error rate statistically based on 100 pairs as follows:

$$\text{Err.} = \frac{P_o}{P_i + P_o}. \tag{6}$$

The proposed method could achieve a matching error rate of 1.2%, significantly outperforming the error rate of 11.5% achieved by CycleGAN. The error rate of 10.1% under the setting without underwater enhancement is also reported for better comparison. As reported, the proposed method could effectively promote the feature matching performance. Finally, it is worth noting that Chen et al. (2019) did not conduct feature matching accuracy analysis.

4.3.3 Qualitative and quantitative results

In this section, we aim to provide both qualitative and quantitative underwater SLAM performance comparisons using real-world underwater datasets. Similarly, the qualitative image enhancement results on both URPC and OUC fisheye datasets are reported in Figure 5. Our method could effectively alleviate the over-under exposure problem and increases contrast and brightness. Besides, our enhancement module could render more details and utilize previous content representations from the original input images. We combine different image enhancement methods with ORB-SLAM2 to explore the improvement of underwater SLAM performance on the URPC dataset. Due to the fact that it is time-consuming to perform UDCP, we do not perform UDCP for the downstream underwater SLAM. The quantitative SLAM performance comparison can be found in Table 2. The proposed GAN-based underwater image enhancement method could heavily promote underwater SLAM performance with a real-time processing inference time. On the other hand, the FUnIE-GAN cannot synthesize enhanced outputs and there is a performance degradation compared to the SLAM performance conducted on the original underwater images.

Furthermore, we combine the GAN-based underwater image enhancement module with two SLAM systems: Dual-SLAM and ORB-SLAM3. The quantitative results are shown in Table 3. The estimated camera pose trajectory is more stable and the initial performance has been promoted heavily. The reasonable image
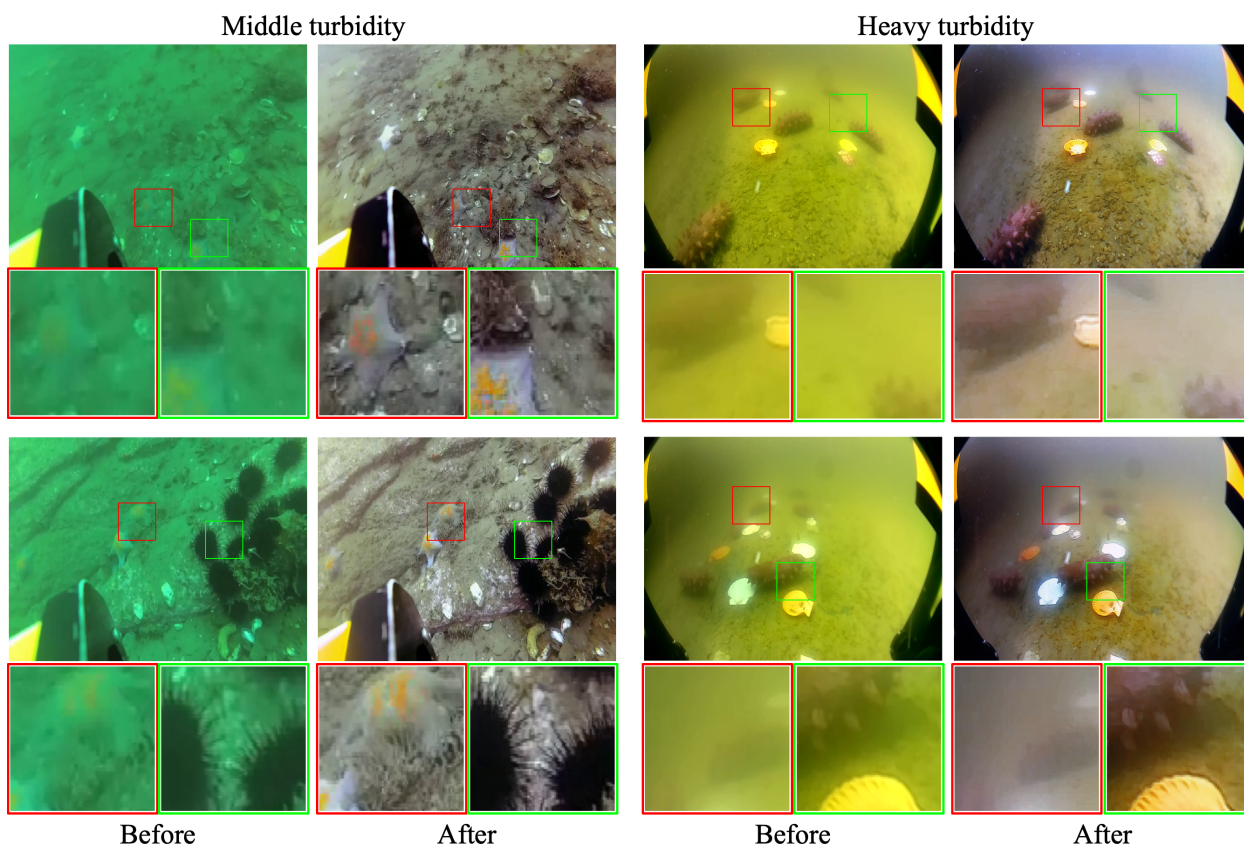

FIGURE 5
The qualitative results of our GAN-based underwater image enhancement on **(A)** URPC dataset and **(B)** OUC fisheye dataset. Best viewed in color.

TABLE 2 Quantization error ORB-SLAM2 baseline with different enhancement methods on the URPC dataset.

Method		ATE ↓	RMSE ↓	Initialization ↓
ORB-SLAM2	w/o	1.418	1.484	84
ORB-SLAM2	CLAHE	1.397	1.468	61
ORB-SLAM2	FUnIE-GAN	1.474	1.505	136
ORB-SLAM2	CycleGAN	1.501	1.565	159
ORB-SLAM2	SpiralGAN	1.348	**1.446**	24
ORB-SLAM2	Proposed	**1.344**	1.447	**23**

The best results are in bold.

enhancement could result in more reliable feature matching so that our method could achieve more stable and accurate outputs. The image enhancement module could promote the underwater SLAM performance in all metrics. Besides, the qualitative trajectory results are also included in Figure 6. The proposed framework outperforms current SLAM methods in both qualitative and quantitative evaluations.

4.3.4 Highly turbid setting

Comprehensive experiments have demonstrated that the proposed method can generate realistic enhanced images with high fidelity and image quality, which can be applied to promote underwater monocular SLAM performance. To further demonstrate the effectiveness and the generalization performance of the proposed framework, we perform experiments on OUC fisheye dataset Zhang et al. (2020). For better illustration, we provide the original underwater image and the enhanced output image in Figure 5B. Similarly, the quantitative and qualitative results under various settings are reported in Table 4 and Figure 7, respectively. The proposed framework can also promote SLAM performance under various challenging settings.

4.4 Ablation studies

4.4.1 Tradeoff between enhancement performance and inference speed

To better explore the performance-computation tradeoff, we have conducted experiments using different values of c in G_s. We

report the computational costs, inference time, and SLAM results in Table 5. SpiralGAN Han et al. (2020) sets $c = 32$ and the proposed compressed method ($c = 16$) has achieved comparable or even better performance with higher speed. When $c = 8$, though it could perform real-time underwater image enhancement with a very high inference speed (FPS=47.54), there is a noticeable enhancement performance drop compared with the proposed method ($c = 16$).

5 Discussions

In this work, the GAN-based image enhancement module and the downstream visual SLAM are optimized separately. The image enhancement is only adopted as an effective image pre-processing module. We assume that the enhanced image could have higher image quality. However, if the GAN-based module cannot generate reasonable images, there would be performance degradation for the SLAM system. The wrong enhanced underwater outputs could lead to error accumulation. We target to optimize the two modules in a multi-task learning manner. The two modules could be mutually beneficial. Besides, we target to build a general open-source underwater SLAM framework which is robust to various underwater conditions. Furthermore, we also target integrating visual-inertial global odometry to combine the scale information into our system. We leave these as our future work.

Furthermore, we adopted the camera pose estimation results from the 3D reconstruction as the pseudo ground truth to evaluate

TABLE 3 Quantization error of different SLAM methods under two settings: 1) without and 2) with the proposed GAN-based underwater image enhancement on the URPC dataset.

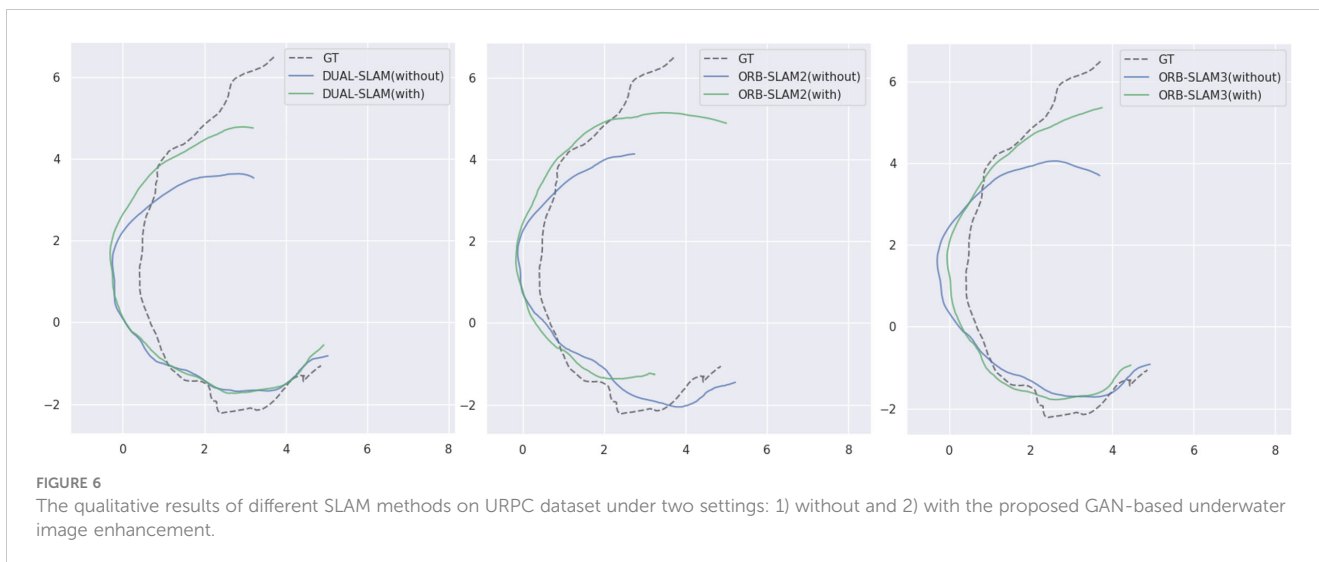| Method		ATE ↓	RMSE ↓	Initialization ↓
ORB-SLAM2	w/o	1.418	1.484	84
ORB-SLAM2	w/	1.344	1.447	23
Dual-SLAM	w/o	1.438	1.502	49
Dual-SLAM	w/	1.350	1.444	6
ORB-SLAM3	w/o	1.405	1.472	69
ORB-SLAM3	w/	**1.332**	**1.433**	**3**

The best results are in bold.

FIGURE 6
The qualitative results of different SLAM methods on URPC dataset under two settings: 1) without and 2) with the proposed GAN-based underwater image enhancement.

TABLE 4 Quantization error of different SLAM methods under two settings: 1) without and 2) with the proposed GAN-based underwater image enhancement on the OUC fisheye dataset.

Method		ATE ↓	RMSE ↓	Initialization ↓
ORB-SLAM2	w/o	2.655	2.700	153
ORB-SLAM2	w/	**2.410**	**2.450**	10
Dual-SLAM	w/o	2.676	2.688	59
Dual-SLAM	w/	2.586	2.520	**1**
ORB-SLAM3	w/o	2.654	2.667	33
ORB-SLAM3	w/	2.559	2.561	2

The best results are in bold.

SLAM performance since it is very challenging and difficult to obtain absolutely accurate ground truth in the underwater setting. To alleviate the ground truth acquisition, we utilize the Structure-from-Motion technique for more robust pose estimation Schönberger and Frahm (2016) in an offline manner since it combines the global bundle adjustment (BA) and pose-graph optimization for more effective and accurate state estimation. The SIFT feature point adopted in Schönberger and Frahm (2016) ¨ 354; Schönberger et al. (2016) is also more accurate than ORB which is widely used in SLAM systems. However, the reconstructed camera poses through 3D reconstruction may still have errors and cannot work under some adverse underwater

FIGURE 7
The qualitative results of different SLAM methods on the OUC fisheye dataset under two settings: 1) without and 2) with the proposed GAN-based underwater image enhancement.

TABLE 5 The ablation studies of various settings on the URPC dataset.

Settings	FPS ↑	MACs (G) ↓	Para. (M) ↓	ATE ↓	RMSE ↓	Initialization ↓
$c = 8$	**47.54**	**2.26**	**0.321**	1.371	1.466	60
$c = 16$	31.99	8.81	1.28	**1.344**	1.447	**23**
$c = 32$	17.63	34.75	4.99	1.348	**1.446**	24
w/o	–	–	–	1.418	1.484	84

We choose ORB-SLAM2 to evaluate the SLAM performance and the last column "w/o" represents the SLAM performance of the original ORB-SLAM2.
The best results are in bold. "-" means not applicable.

environments (*e.g.*, motion blur, camera shaking, an extensive range of rotation, and *etc.*).

6 Conclusion

This paper has proposed a generic and practical framework to perform robust and accurate underwater SLAM. We have designed a real-time GAN-based image enhancement module through knowledge distillation to promote underwater SLAM performance. With the adaptation of an effective underwater image enhancement as a pre-processing image module, we could synthesize enhanced underwater images with high fidelity for further underwater SLAM, leading to observable performance gains. The proposed framework can work effectively in an extensible way, in which external modifications can plug in the underwater monocular SLAM algorithms.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

ZZ and ZX did most of the work and contributed equally in this paper. ZY handled the work of revising the article, and S-KY provided guidance and funding for this research. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2023.1161399/full#supplementary-material

References

Aguinaldo, A., Chiang, P.-Y., Gain, A., Patil, A., Pearson, K., and Feizi, S. (2019). Compressing gans using knowledge distillation. doi: 10.48550/arXiv.1902.00159

Akkaynak, D., and Treibitz, T. (2019). "Sea-Thru: a method for removing water from underwater images," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA. (New York City, USA: IEEE), 1682–1691.

Anwar, S., and Li, C. (2020). Diving deeper into underwater image enhancement: a survey. *Signal Process. Image Commun.* 89, 115978. doi: 10.1016/j.image.2020.115978

Asmare, M. H., Asirvadam, V. S., and Hani, A. F. M. (2015). Image enhancement based on contourlet transform. *Signal Image Video Process* 9, 1679–1690. doi: 10.1007/s11760-014-0626-7

Bresson, G., Alsayed, Z., Yu, L., and Glaser, S. (2017). Simultaneous localization and mapping: a survey of current trends in autonomous driving. *IEEE Trans. Intelligent Vehicles* 2, 194–220. doi: 10.1109/TIV.2017.2749181

Buscher, E., Mathews, D. L., Bryce, C., Bryce, K., Joseph, D., and Ban, N. C. (2020). Applying a low cost, mini remotely operated vehicle (rov) to assess an ecological baseline of an indigenous seascape in canada. *Front. Mar. Sci.* 7, 669. doi: 10.3389/fmars.2020.00669

Campos, C., Elvira, R., Rodríguez, J. J. G., Montiel, J. M., and Tardós, J. D. (2021). Orb-slam3: an accurate open-source library for visual, visual–inertial, and multimap slam. *IEEE Trans. Robotics* 37 (6), 1874–1890. doi: 10.1109/TRO.2021.3075644

Carreras, M., Hernández, J. D., Vidal, E., Palomeras, N., Ribas, D., and Ridao, P. (2018). Sparus ii auv–a hovering vehicle for seabed inspection. *IEEE J. Oceanic Eng.* 43, 344–355. doi: 10.1109/JOE.2018.2792278

Chen, W., Rahmati, M., Sadhu, V., and Pompili, D. (2019). "Real-time image enhancement for vision-based autonomous underwater vehicle navigation in murky waters," in *WUWNet '19: Proceedings of the 14th International Conference on Underwater Networks & Systems*, Atlanta, GA, USA. (USA: ACM Digital Library), 1–8.

Cho, Y., and Kim, A. (2017). "Visibility enhancement for underwater visual slam based on underwater light scattering model," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. (New York City, USA: IEEE), 710–717.

Drews, P., Nascimento, E., Moraes, F., Botelho, S., and Campos, M. (2013). "Transmission estimation in underwater single images," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops*, Sydney, NSW, Australia. (New York City, USA: IEEE), 825–830.

Elvira, R., Tardós, J. D., and Montiel, J. M. (2019). "Orbslam-atlas: a robust and accurate multi-map system," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Macau, China. (New York City, USA: IEEE), 6253–6259.

Fabbri, C., Islam, M. J., and Sattar, J. (2018). "Enhancing underwater imagery using generative adversarial networks," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, Brisbane, QLD, Australia. (New York City, USA: IEEE), 7159–7165.

Ferrera, M., Creuze, V., Moras, J., and Trouvé-Peloux, P. (2019a). Aqualoc: an underwater dataset for visual–inertial–pressure localization. *Int. J. Robotics Res.* 38, 1549–1559. doi: 10.1177/0278364919883346

Ferrera, M., Moras, J., Trouvé-Peloux, P., and Creuze, V. (2019b). Real-time monocular visual odometry for turbid and dynamic underwater environments. *Sensors* 19, 687. doi: 10.3390/s19030687

García, S., López, M. E., Barea, R., Bergasa, L. M., Gómez, A., and Molinos, E. J. (2016). "Indoor slam for micro aerial vehicles control using monocular camera and sensor fusion," in *2016 international conference on autonomous robot systems and competitions (ICARSC)*, Bragança, Portugal. (New York City, USA: IEEE), 205–210.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative adversarial networks. *Commun ACM* (USA: ACM Digital Library) 63 (11), 139–144.

Han, R., Guan, Y., Yu, Z., Liu, P., and Zheng, H. (2020). *Underwater image enhancement based on a spiral generative adversarial framework* (IEEE Access), 8, 218838–218852.

He, K., Sun, J., and Tang, X. (2010). Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* 33, 2341–2353. doi: 10.1109/TPAMI.2010.168

Hoegh-Guldberg, O., Mumby, P. J., Hooten, A. J., Steneck, R. S., Greenfield, P., Gomez, E., et al. (2007). Coral reefs under rapid climate change and ocean acidification. *Science* 318, 1737–1742. doi: 10.1126/science.1152509

Huang, H., Lin, W.-Y., Liu, S., Zhang, D., and Yeung, S.-K. (2020). "Dual-slam: a framework for robust single camera navigation," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Las Vegas, NV, USA. (New York City, USA: IEEE), 4942–4949.

Huang, Z., Wan, L., Sheng, M., Zou, J., and Song, J. (2019). "An underwater image enhancement method for simultaneous localization and mapping of autonomous underwater vehicle," in *2019 3rd International Conference on Robotics and Automation Sciences (ICRAS)*, Wuhan, China. (New York City, USA: IEEE), 137–142.

Huvenne, V. A., Robert, K., Marsh, L., Iacono, C. L., Le Bas, T., and Wynn, R. B. (2018). Rovs and auvs. *Submarine Geomorphology*, 93–108.

Islam, M. J., Enan, S. S., Luo, P., and Sattar, J. (2020a). "Underwater image super-resolution using deep residual multipliers," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, Paris, France. (New York City, USA: IEEE), 900–906. doi: 10.1109/ICRA40945.2020.9197213

Islam, M. J., Luo, P., and Sattar, J. (2020b). Simultaneous enhancement and super-resolution of underwater imagery for improved visual perception. doi: 10.15607/RSS.2020.XVI.018

Islam, M. J., Xia, Y., and Sattar, J. (2020c). Fast underwater image enhancement for improved visual perception. *IEEE Robotics Automation Lett.* 5, 3227–3234. doi: 10.1109/LRA.2020.2974710

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). "Image-to-image translation with conditional adversarial networks," in *2017 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA. (New York City, USA: IEEE), 1125–1134.

Kingma, D. P., and Ba, J. (2014). Adam: A method for stochastic optimization. doi: 10.48550/arXiv.1412.6980

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., et al. (2017). "Photo-realistic single image super-resolution using a generative adversarial network," in *2017 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA. (New York City, USA: IEEE), 4681–4690.

Li, C., Anwar, S., and Porikli, F. (2020a). Underwater scene prior inspired deep underwater image and video enhancement. *Pattern Recognition* 98, 107038. doi: 10.1016/j.patcog.2019.107038

Li, M., Lin, J., Ding, Y., Liu, Z., Zhu, J.-Y., and Han, S. (2020b). "Gan compression: efficient architectures for interactive conditional gans," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (New York City, USA: IEEE) 44 (12), 9331–9346. doi: 10.1109/TPAMI.2021.3126742

Li, N., Zheng, Z., Zhang, S., Yu, Z., Zheng, H., and Zheng, B. (2018). *The synthesis of unpaired underwater images using a multistyle generative adversarial network* (IEEE Access), 54241–54257.

Mirza, M., and Osindero, S. (2014). Conditional generative adversarial nets. doi: 10.48550/arXiv.1411.1784

Mur-Artal, R., and Tardós, J. D. (2017). Orb-slam2: an open-source slam system for monocular, stereo, and rgb-d cameras. *IEEE Trans. Robotics* 33, 1255–1262. doi: 10.1109/TRO.2017.2705103

Qin, T., Li, P., and Shen, S. (2018). Vins-mono: a robust and versatile monocular visual-inertial state estimator. *IEEE Trans. Robotics* 34, 1004–1020. doi: 10.1109/TRO.2018.2853729

Qin, T., and Shen, S. (2018). "Online temporal calibration for monocular visual-inertial systems," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Madrid, Spain. (New York City, USA: IEEE), 3662–3669.

Rahman, S., Li, A. Q., and Rekleitis, I. (2018). "Sonar visual inertial slam of underwater structures," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, Brisbane, QLD, Australia. (New York City, USA: IEEE), 5190–5196.

Rahman, S., Li, A. Q., and Rekleitis, I. (2019a). "Contour based reconstruction of underwater structures using sonar, visual, inertial, and depth sensor," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Macau, China. (New York City, USA: IEEE), 8054–8059.

Rahman, S., Li, A. Q., and Rekleitis, I. (2019b). "Svin2: an underwater slam system using sonar, visual, inertial, and depth sensor," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Macau, China. (New York City, USA: IEEE), 1861–1868.

Reza, A. M. (2004). Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement. *J. VLSI Signal Process. Syst. signal image video Technol.* 38, 35–44. doi: 10.1023/B:VLSI.0000028532.53893.82

Rublee, E., Rabaud, V., Konolige, K., and Bradski, G. (2011). "Orb: an efficient alternative to sift or surf," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, Brisbane, QLD, Australia. (New York City, USA: IEEE), 2564–2571.

Schönberger, J. L., and Frahm, J.-M. (2016). "Structure-from-motion revisited," in *2016 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA. (New York City, USA: IEEE), 4104–4113.

Schönberger, J. L., Zheng, E., Frahm, J.-M., and Pollefeys, M. (2016). "Pixelwise view selection for unstructured multi-view stereo," in *European Conference on computer vision* (Springer), 501–518.

Zhang, X., Zeng, H., Liu, X., Yu, Z., Zheng, H., and Zheng, B. (2020). *In situ holothurian noncontact counting system: a general framework for holothurian counting* (IEEE Access) 8, 210041–210053.

Zhu, J., Park, T., Isola, P., and Efros, A. A. (2017). "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy. (New York City, USA: IEEE). 2242–2251. doi: 10.1109/ICCV.2017.244