Twin Adversarial Contrastive Learning for Underwater Image Enhancement and Beyond

Risheng Liu, *Member, IEEE*, Zhiying Jiang, Shuzhou Yang, and Xin Fan, *Senior Member, IEEE*

***Abstract*— Underwater images suffer from severe distortion, which degrades the accuracy of object detection performed in an underwater environment. Existing underwater image enhancement algorithms focus on the restoration of contrast and scene reflection. In practice, the enhanced images may not benefit the effectiveness of detection and even lead to a severe performance drop. In this paper, we propose an object-guided twin adversarial contrastive learning based underwater enhancement method to achieve both visual-friendly and task-orientated enhancement. Concretely, we first develop a bilateral constrained closed-loop adversarial enhancement module, which eases the requirement of paired data with the unsupervised manner and preserves more informative features by coupling with the twin inverse mapping. In addition, to confer the restored images with a more realistic appearance, we also adopt the contrastive cues in the training phase. To narrow the gap between visually-oriented and detection-favorable target images, a task-aware feedback module is embedded in the enhancement process, where the coherent gradient information of the detector is incorporated to guide the enhancement towards the detection-pleasing direction. To validate the performance, we allocate a series of prolific detectors into our framework. Extensive experiments demonstrate that the enhanced results of our method show remarkable amelioration in visual quality, the accuracy of different detectors conducted on our enhanced images has been promoted notably. Moreover, we also conduct a study on semantic segmentation to illustrate how object guidance improves high-level tasks. Code and models are available at https://github.com/Jzy2017/TACL**

***Index Terms*— Underwater image enhancement, object detection, contrastive learning, generative adversarial learning.**

I. Introduction

UNDERWATER image enhancement is a crucial and inverse restoration task, with the purpose of alleviating the turbid and color cast scenery caused by absorption and scattering. As we know, water has different absorption capabilities for different wavelengths. Red long waves are the most penetrating and easily absorbed, while blue and green short waves are weakly penetrated and easily diffused and scattered. Therefore, the images captured underwater usually appear blueish or greenish. Moreover, the suspended particles also disturb the light scattering, leading to the low contrast and muddy phenomenon. For underwater exploration, the degraded images tend to impair the effectiveness of subsequent applications, such as object detection [1], [2], recognition [3] and segmentation [4]. It is of practical significance to remedy the distortion to assist the exploitation of a complicated ocean environment.

Existing attempts adopt the underwater enhancement as a pre-process and feed the enhanced results as latent inputs to the successive detection algorithms. They formulate the degradation into a physical model and estimate the model parameters to obtain the reconstructed images. However, the fixed model is limited in characterizing diverse underwater factors. Recently, deep learning based enhancement methods have attracted a great deal of focus. They elaborate networks to perform the restoration in an end-to-end manner and have achieved a significant advance. It should be noted that underwater enhancement methods [5]–[7] concentrate on the restoration of scene radiance by intensifying the shaded details and authentic reflection. They enforce the results with a more vivid appearance and enriched minutia. However, these human-perceptual results may not facilitate the subsequent algorithms to understand the scene content. That is, the improvement of visual features plays a limited effect on the subsequent applications in principle.

In this paper, we propose an object-guided twin adversarial contrastive learning based underwater image enhancement. Since the lack of substantial paired distorted-clear underwater images and the synthetic data cannot simulate the diverse condition well, we employ a closed-loop constrained adversarial enhancement module to achieve the transfer between the distorted images and clear ones. What's more, the twin inverse mapping from in-air images is coupled with the forward procedure to promote the generalization on real-world data. Considering the opposite relationship among the restored images with the degraded and clear images, we introduce the contrastive principle, where the restored results are forced to approach the in-air images and far away from the distorted one in a certain representation. In this way, the contrastive prior is able to maximize the mutual information to confer the results with a more realistic appearance.

To benefit the enhanced results more suitable for object detection, we incorporate a task-aware feedback module to

Manuscript received 9 December 2021; revised 16 May 2022 and 22 June 2022; accepted 28 June 2022. Date of publication 18 July 2022; date of current version 22 July 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61922019 and in part by the Fundamental Research Funds for the Central Universities. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Junhui Hou. *(Corresponding author: Risheng Liu.)*

Risheng Liu is with the DUT-RU International School of Information Science and Engineering and the Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, Dalian University of Technology, Dalian 116024, China, also with the Peng Cheng Laboratory, Shenzhen 518066, China, and also with the Pazhou Laboratory (Huangpu), Guangzhou 510715, China (e-mail: rsliu@dlut.edu.cn).

Zhiying Jiang and Shuzhou Yang are with the School of Software Technology, Dalian University of Technology, Dalian 116024, China (e-mail: zyjiang0630@gmail.com; yszdyx@gmail.com).

Xin Fan is with the DUT-RU International School of Information Science and Engineering and the Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, Dalian University of Technology, Dalian 116024, China, and also with the Peng Cheng Laboratory, Shenzhen 518066, China (e-mail: xin.fan@dlut.edu.cn).

Digital Object Identifier 10.1109/TIP.2022.3190209

transmit the coherent detection information of a specific detector and guide the update of the enhancement module towards the task-favorable direction. Additionally, the task-aware feedback module is flexible for the various detector and can be substituted by any promising instrument to upgrade the whole effect, achieving a generic framework to bridge the low-level enhancement and subsequent high-level applications.

To summarize, our contributions are as follows:

- We develop a twin adversarial contrastive learning based underwater image enhancement, which utilizes the bilateral mapping between the underwater image and clear counterpart to relieve the highly under-constrained property of restoration. Compared with homogeneous supervised methods, the proposed method generalizes well in the real world.
- We introduce a contrastive principle in the training process, where the opposite relationship among anchors, positive and negative, is considered. It provides abundant mutual and perceptual information, improving the enhanced image with a more sensible and realistic feature.
- The task-aware feedback module is proposed, which employs the plugged detector to transmit the coherent gradient information via the localization and confidence, and constrains the update of the enhancement module towards the detection-favorable direction.
- Extensive experiments results of the proposed method show a great superiority against the others. The detection accuracy of prolific detectors proves that the proposed method is able to generate detection-favorable results. Overall, the proposed method advances the enhancement more suitable for detection.

The rest paper is organized as follows: Section II gives a brief review of the existing underwater image enhancement methods and the related techniques. Section III details the proposed method. Experiment and ablation studies are illustrated in Section IV, and the conclusion is presented in Section V.

II. Related Work

A. Underwater Image Enhancement

Recently, numerous underwater image enhancement methods have been proposed and can be divided into three categories, including physical model based, model-free, and deep learning based methods. In the first category, physical models are developed with the hand-crafted domain knowledge, in which the enhancement task refers to solving the inverse problem to recover the clear images. For example, dark channel prior (DCP) [8] reveals the pixel distribution of clear images. Chiang *et al.* [9] firstly combined the DCP with wavelength-dependent compensation to enhance the visual results. Peng *et al.* [10] proposed a generalized dark channel prior with the adaptive color correction into the formation to fit the complicated underwater environment. To solve the bluish and greenish phenomenon, a red channel prior [11] is developed. Liu *et al.* [12] proposed a regression-inspired medium transmission estimation and global light estimation algorithm to recover the distortion-free images. After that, Liu *et al.* [13] utilized the quadtree subdivision to maximize the image contrast with the optimal transmission map. Wang *et al.* [14] developed an adaptive attenuation-curve prior with non-local manner to stretch the statistical distribution of pixel values. Although the above-mentioned model based methods have achieved a great improvement on color rectification, they are effective for the specific distortion and weak in robustness due to the heavy reliance on hand-crafted priors.

Model-free methods refer to adjusting image pixel values directly to improve the color and visibility of underwater images. Zhuang *et al.* [15] proposed a Bayesian retinex based enhancement method. Ghani *et al.* [16] enhanced the underwater image by modifying the image histograms column wisely in accordance with Rayleigh distribution. Based on the fusion strategy, Ancuti *et al.* [5] derived the weight measures only from the degraded version of the image. After that, [17] developed a color-compensated and white-balanced version of the original degraded image and blended them in a multi-scale manner. Gao *et al.* [18] employed an adaptive retinal mechanism to enhance the underwater images. However, due to the lack of underwater imaging constraints, the enhanced results tend to appear over-exposure.

There are also productive attempts within deep learning. As the demand for sufficient training data, Li *et al.* [6] proposed the first real-world underwater image benchmark where the corresponding reference images are equipped from a professional perspective. Guo *et al.* [19] utilized weakly supervised learning and proposed a multi-scale dense generative network. More recently, Li *et al.* [20] investigated the features of different color spaces and proposed a multi-color integrated network to enforce more diverse feature presentation. Additionally, the transmission-guided decoder realized adaptive attention focus. Considering that the optimization of image quality separately may not achieve the intuitive improvement on the detection task, Chen *et al.* [21] proposed two perceptual enhancement models with a detection perceptor to generate patch level visually pleasing or detection favorable images. The aforementioned deep learning methods have obtained impressive enhancement results, but they rely on adequate, realistic data to learn the powerful transformation. However, existing data cannot fully overlay the complicated underwater environments, resulting in poor generalization in the real world.

B. Object Detection

Object detection is a worthy research direction in computer vision that aims at framing out the locations and specifying the categories of objects in the given images. In recent years, considerable object detection algorithms have been developed and made an impressive advance in accuracy. The existing methods can be divided into two categories: one-stage and two-stage methods. One-stage detectors represented by YOLO series [22]–[24] refer to regress the category probability and position of objects directly. While for the two-stage detectors, they generate a series of candidate boxes as samples and then classify the samples through the convolutional neural network, the representatives are RCNN series [1], [25]–[28]. More recently, the incorporation of a deep backbone promotes detection accuracy greatly, while the time-consuming and

computation cost restricts the application of these algorithms. Therefore, the development of highly efficient and lightweight detection networks raises research interest. It should be noted that in-air object detection has achieved significant performance, but there are few object detection algorithms designed for underwater images. At present, retraining the in-air detectors with underwater data has not obtained satisfactory results. Besides, existing attempts introduce underwater enhancement as pre-process and feed the enhanced results as the target, but the effectiveness of detection algorithms cannot be improved as they supposed, since the enhancement methods focus on the improvement of visibility and contrast, ignoring the intrinsical features detection concerned.

C. Generative Adversarial Network

As the important branch of deep learning, Generative Adversarial Learning (GAN) has achieved impressive progress in computer vision [29]–[31]. The critical point of GAN relies on the adversarial loss, which tries to confer the results with a more realistic appearance and deceive the judgment of the discriminator. The training of GAN plays a confrontational procedure. The effectiveness of the generator and discriminator are both improved alternatively until the equilibrium state. Recently, many variants of GAN have been explored to fit the different kinds of image data and been applied to various low-level tasks [32], [33] as the particular robustness on ill-posed restoration. For underwater image enhancement, the complicated distortion and variable condition impair the generic network to simulate the mapping from the source image to the target image. Therefore, it is exactly applicable to adopt the adversarial mechanism into this task, and the translated image cannot be distinguished from the target domain.

D. Contrastive Learning

Deep learning utilizes a deep convolution network to learn the implicit characteristic from the source domain to the target domain. In practice, abundant paired images are not available. Contrastive learning targets to ease the data requirement and focuses on self-supervised representation learning. It aims to enforce the results approach the positives and push them away from the negatives in a certain space. Nowadays, contrastive learning is widely applied in computer vision tasks [34], [35]. In low-level vision, Park *et al.* [36] demonstrated the improvement of contrastive learning in the unpaired image translation task. Wu *et al.* [37] proposed a compact dehazing network where it verified the improvement of contrastive learning performed on the autoencoder-like framework. Although the above mentioned methods have promoted the performance of self-supervised and semi-supervised learning, there is still a gap in the attempt to introduce contrast learning to underwater image enhancement. We notice that existing methods employed the learnable parameters and models to simulate the mapping from the degraded underwater images to the interference-free high quality counterpart, which ignored the potential value of the raw underwater data. We suppose that the undesirable image is able to constrain the restored image far away from the scattering and refraction appearance while the clear image forces them close to the desirable background, achieving the bilateral constraints on the enhanced results.

III. The Proposed Method

A. Framework Architecture

Given an observed underwater image $\mathbf{X}$, it suffers from severe degradation, such as color cast and muddy interference, leading to a performance drop of underwater object detection. Conventional underwater image enhancement methods only emphasize the stretch of visual and contrast on the results $\mathbf{Y}$. We found that the visual improvement of underwater images may not necessarily lead to the improvement of detection accuracy because the algorithms and the human eyes have different ways of perceiving the scene. Correcting the color and the contrast of the target scene simply will not promote the understanding of the scene, which shows the limited effect on the subsequent detection task. Therefore, we attempt to introduce the information from object detection to achieve both visual quality improvement and task-friendly enhancement. In this paper, we proposed an object-guided twin adversarial contrastive learning based underwater image enhancement, as shown in Fig. 1.

Due to the lack of sufficient real-world supervised data and the synthetic images present huge cap from real world, full supervised underwater enhancement methods [6], [38] meet with weak generalization and poor robustness. To ease the dependence on reliable data, we introduce the unsupervised and self-supervised manner into our method. Specifically, the whole framework consists of a twin adversarial and contrastive enhancement module. One performs the degradation to the in-air closed-loop generation process, while the other carries out the in-air to degradation loop generation. Through the proposed iterative generation procedure, i.e., X2Y and Y2X modules, we exploit the internal relationship between the underwater domain and in-air domain to realize the feature translation across them with more generalization. Besides, we also introduce the contrastive principle into our framework, where not only the consistency but also the discrepancy are taken into consideration. That is, the restored results are forced to approach the in-air clear images and far away from the real-world underwater images in a certain representation obtained by the Perceptual Feature Extractor (PFE) and vice versa, facilitating the results with more sensible features.

Compared with conducting detection on the underwater image directly, the performance of applying detection algorithms on the enhanced image did not show the expected improvement. Therefore, to bridge the gap between the visual-friendly and detection-favorable orientations, we embed a Task-Aware Feedback module (TAF) as an adapter to propagate the coherent information of detection and guide the update of enhancement towards the detection-favorable orientation. That is, the object information detected by algorithms is delivered to the enhancement module and supports the bilateral constrained network to learn the features of interest to the certain detector. And in the whole process, the task-aware module and the closed-loop constrained module behave interactively on each other, resulting in that the enhancement is

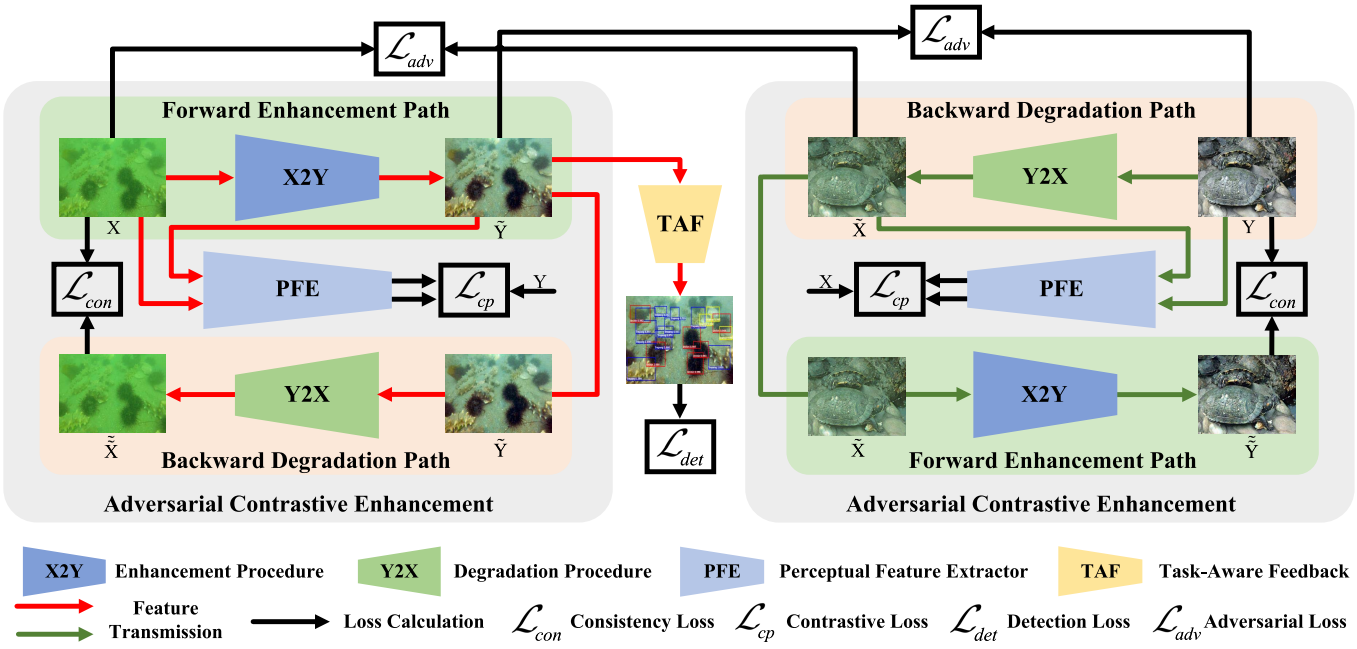

Fig. 1. Workflow of the proposed method. It presents a twin adversarial contrastive enhancement process. Each of them consists of a forward enhancement path and a backward degradation path. We embed a task-aware feedback module (TAF) to bridge the gap between the visual-friendly enhancement results and the detection-favorable images, achieving the collaborative improvement of image quality and detection accuracy.

more favorable for detection and the detection is also beneficial to enhancement. Note that the proposed framework is flexible to incorporate any other prolific detection algorithms, making it scalable to fit more detectors.

B. Twin Adversarial Enhancement Module

Compared with previous direct mapping, which translates underwater images to the clear counterpart, the twin adversarial enhancement consists of two closed-loop mappings, each of which is composed of a forward enhancement path and a backward degradation path. For the forward path, it aims at learning the translation map between two domains, i.e., underwater image denoted as $\mathbf{X}$ and in-air image denoted as $\mathbf{Y}$, and realizes the mapping $F : \mathbf{X} \rightarrow \mathbf{Y}$. On the contrary, the backward path utilizes the clear image $\mathbf{Y}$ to render the underwater image with the mapping $B : \mathbf{Y} \rightarrow \mathbf{X}$, which is the reverse process of F. In the proposed method, we employ these two mapping processes alternately to construct a twin closed-loop mechanism. Specifically, in the left part of Fig. 1, $\mathbf{X}$ is the observed underwater image, the loop mapping first translate the underwater image to the in-air domain, and then utilizes the constructed in-air image $\tilde{\mathbf{Y}}$ to the degraded version $\tilde{\tilde{\mathbf{X}}}$ again. Therefore, this forward-backward generation process can be formulated as:

$$\tilde{\tilde{\mathbf{X}}} = G_B(\tilde{\mathbf{Y}}) = G_B(G_F(\mathbf{X})), \tag{1}$$

where $G_F(\cdot)$ and $G_B(\cdot)$ mean the mapping function of F, B respectively. For the right part of Fig. 1, we feed the real-world clear image $\mathbf{Y}$ into the backward degradation network to generate the synthetic degraded image $\tilde{\mathbf{X}}$, it owns the characteristics of underwater images, such as color cast and turbidity. After that, the forward enhancement network is employed to recover the clear image $\tilde{\tilde{\mathbf{Y}}}$ based on the generated underwater data, expressed as:

$$\tilde{\tilde{\mathbf{Y}}} = G_F(\tilde{\mathbf{X}}) = G_F(G_B(\mathbf{Y}))). \tag{2}$$

In addition to the two reverse generators G_F, G_B, we also introduce two discriminators D_F and D_B in each closed-loop mapping, where D_F targets to distinguish the image $\tilde{\mathbf{Y}}$, $\tilde{\tilde{\mathbf{Y}}}$ with the real-world clear image $\mathbf{Y}$, and D_B targets to discriminate $\tilde{\mathbf{X}}, \tilde{\tilde{\mathbf{X}}}$ from $\mathbf{X}$. Two inverse mappings of the proposed bilateral constrain form the twin adversarial learning, which ensures the robustness and generalization of the enhancement module in the real-world underwater environment.

C. Contrastive Prior

For the lack of sufficient supervised real-world underwater paired, we introduce the contrastive principle into the framework to generate more realistic and plausible results. Contrastive learning divides the existing data into two categories, including positives and negatives, and learns a representation space to enforce the results closer to positive samples and far away from the negative samples. For underwater image enhancement, we suppose the observed underwater images $\mathbf{X}$ as negatives and the clear in-air images $\mathbf{Y}$ as positives. To build the reasonable representation space, we employ the VGG-19 [39] as Perceptual Feature Extractor (PFE), denoted as $\mathcal{V}$, and the intermediate features of it has been proved useful in perceptual space [40]–[43]. To make full use of the intermediate features, we extract the features from a series of hidden layers and adopt l_1 regularization to strengthen the

Contrastive Prior (CP), expressed as:

$$\mathcal{L}_{cp} = \sum_{i=1}^{n} \rho_i \cdot \left(\frac{\|\mathcal{V}_i(\mathbf{Y}) - \mathcal{V}_i(G_F(\mathbf{X}))\|_1}{\|\mathcal{V}_i(\mathbf{X}) - \mathcal{V}_i(G_F(\mathbf{X}))\|_1} + \frac{\|\mathcal{V}_i(\mathbf{X}) - \mathcal{V}_i(G_B(\mathbf{Y}))\|_1}{\|\mathcal{V}_i(\mathbf{Y}) - \mathcal{V}_i(G_B(\mathbf{Y}))\|_1} \right), \quad (3)$$

where i denotes the features from i-th layer and ρ_i means the weight parameter. In this paper, we set $i = 1, 3, 5, 9, 13$ and its corresponding weight $\rho_i = \frac{1}{32}, \frac{1}{16}, \frac{1}{8}, \frac{1}{4}, 1$, because the feature from the deeper layer contains more complicated perceptual information than the shallower layer. Bringing the aforementioned contrastive prior into twin adversarial learning, we alleviate the reliance on plenty of credible images, providing more flexibility on this task.

D. Task-Aware Feedback Module

In order to make the latent results of the enhancement module more suitable for detection and invest them with more detection-favoring feature information, we propose a Task-Aware Feedback module (TAF) to generate images of interest. Inspired by semantic loss [41], we suppose that the elaborate networks designed for high-level vision tasks own the ability to depict implicit characteristics and hold substantial semantic information. Therefore, the proposed feedback module aims at exploiting latent features and delivering the perception of detectors, which holds the object information in terms of detection accuracy. Specifically, a detector pre-trained with the enhancement data is employed to obtain the character of basic category information. Then, we adopt the underwater images with annotation to joint train the whole framework in an end-to-end manner to obtain the detection-favorable enhancement. In the whole framework, the degraded underwater images are fed into the enhancement module to generate the clear in-air like latent images, then the task-aware feedback module evaluates the susceptibility on detection task and transmits the coherent information to the enhancement module to guide visual improvement to be more conducive to the detector. It is worth mentioning that the proposed framework is flexible to be equipped with arbitrary detectors into the feedback module, making it expansible to adapt diverse detectors.

As mentioned above, existing object detection algorithms can be categorized into two parts, i.e., one-stage [23], [24], [44] and two-stage [1], [27], [28], [45]. The main difference between them is that two-stage methods adopt the corresponding region proposal algorithms to generate candidate regions from the inputs and classify the candidates with the classifier. At the same time, the one-stage methods output the bounding box and classification label directly. In this paper, both one-stage and two-stage methods can be applied as detection preceptors since the coherent loss of these detectors is designed from the same two perspectives, including confidence loss and localization loss, expressed as:

$$\mathcal{L}_{det} = \mathcal{L}_{conf} + \mathcal{L}_{loc}, \quad (4)$$

in which $\mathcal{L}_{conf}$ measures the classification deviation over multiple classes confidences and aims at minimizing the class discrepancy between the predicted and ground truth patches. In contrast, $\mathcal{L}_{loc}$ means the regression between the predicted and the ground truth box to minimize the location discrepancy. Specifically, in SSD detector [46], the smooth l_1 loss [26] is employed as localization loss:

$$\mathcal{L}_{smooth_{L1}}(x) = \begin{cases} 0.5x^2, & \text{if } |x| < 1, \\ |x| - 0.5, & \text{otherwise}, \end{cases} \quad (5)$$

where x denotes the center distance between the predicted box and the ground truth box. On the other hand, softmax loss is utilized as the specific measure of confidence:

$$\mathcal{L}_{conf} = -\sum_{i=1}^{C} pre_i \log(gt_i), \quad (6)$$

pre_i and gt_i denote the i-th element of the predicted and ground truth class vectors, respectively. For RetinaNet, they develop a focal loss [47] to substitute the cross-entropy loss in $\mathcal{L}_{conf}$, where the weight of a large number of simple negative samples in training can be reduced and the problem of serious imbalance in the ratio of positive and negative samples occurred in one-stage object detection can be alleviated. And in [48], the Generalized Intersection over Union (GIoU) is proposed to calculate the localization loss. For the two-stage detector [1], its detection feedback loss is also composed of localization and confidence regression losses that come from RPN [27] and RCNN [25].

From the perspective of optimization, all these detectors can be plugged into our framework to propagate the coherent object information in the form of detection-specific feedback loss $\mathcal{L}_{det}$ to benefit the update of the bilateral enhancement module to the detection-favorable direction.

E. Objective Function

We apply adversarial losses on each closed-loop mappings of the proposed twin enhancement module. For the mapping function $F : \mathbf{X} \to \mathbf{Y}$ and its discriminator D_F, the objective function can be expressed as:

$$\mathcal{L}_{adv}(G_F, D_F, \mathbf{X}, \mathbf{Y}) = \mathbb{E}_{\mathbf{Y} \sim p_{\text{data}}(\mathbf{Y})}[\log D_F(\mathbf{Y})] + \mathbb{E}_{\mathbf{X} \sim p_{\text{data}}(\mathbf{X})}[\log(1 - D_F(G_F(\mathbf{X})))], \quad (7)$$

where $G_F(\mathbf{X})$ attempts to generate the image which is similar to $\mathbf{Y}$, while D_F aims at distinguishing $G_F(\mathbf{X})$ from the real $\mathbf{Y}$. In Eq. 7, G_F tries to minimize the objective, while D_F tries to maximize it. Accordingly, the reverse mapping $B : \mathbf{Y} \to \mathbf{X}$ shares the similar objective with F, shown as:

$$\begin{aligned} &\mathcal{L}_{adv}(G_B, D_B, \mathbf{Y}, \mathbf{X}) \\ &= \mathbb{E}_{\mathbf{X} \sim p_{\text{data}}(\mathbf{X})}[\log D_B(\mathbf{X})] \\ &\quad + \mathbb{E}_{\mathbf{Y} \sim p_{\text{data}}(\mathbf{Y})}[\log(1 - D_B(G_B(\mathbf{Y})))]. \end{aligned} \quad (8)$$

As we know, adversarial training is able to produce the output with the same distribution as the target domain strictly. But in theory, the network translates the same set of input images to a variety of randomly arranged images in the target domain, in which all these mappings match the target distribution well. Therefore, the translated image may not be

the desired result corresponding to the input. In order to reduce the solution space that satisfies the target domain, we further introduce the consistency loss that ensures the correlation between the mapped result with the input image. For the left cycle in Fig. 1, the backward degradation should be able to restore the $\tilde{\tilde{\mathbf{X}}}$ approaching to $\mathbf{X}$, i.e., $\mathbf{X} \rightarrow G_F(\mathbf{X}) \rightarrow G_B(G_F(\mathbf{X})) \approx \mathbf{X}$. In the same way, the right cycle of Fig. 1 also should restore the results approximate to $\mathbf{Y}$, i.e., $\mathbf{Y} \rightarrow G_B(\mathbf{Y}) \rightarrow G_F(G_B(\mathbf{Y})) \approx \mathbf{Y}$. In this paper, we adopt the l_1-norm to measure the discrepancy, and the consistency loss of two loop mappings can be formulated as:

$$\mathcal{L}_{con} = \mathbb{E}_{\mathbf{X}\sim p_{\text{data}}(\mathbf{X})}[\|G_B(G_F(\mathbf{X})) - \mathbf{X}\|_1] + \mathbb{E}_{\mathbf{Y}\sim p_{\text{data}}(\mathbf{Y})}[\|G_F(G_B(\mathbf{Y})) - \mathbf{Y}\|_1]. \quad (9)$$

Taking contrastive prior into consideration, the objective of the twin enhancement module is:

$$\mathcal{L}_{en} = \mathcal{L}_{adv} + \lambda_1 \mathcal{L}_{con} + \lambda_2 \mathcal{L}_{cp}. \quad (10)$$

And the full objective of the proposed object-guided twin adversarial contrastive learning, expressed as:

$$\mathcal{L}_{total} = \mathcal{L}_{adv} + \lambda_1 \mathcal{L}_{con} + \lambda_2 \mathcal{L}_{cp} + \lambda_3 \mathcal{L}_{det}, \quad (11)$$

where $\lambda_1, \lambda_2, \lambda_3$ control the relative importance of the four parts, and are set as $10, 0.5, 0.1$ respectively. In section IV, we compare the proposed framework against the ablation of four objectives, including the single loop adversarial enhancement network, twin adversarial enhancement module, twin module equipped with contrastive prior as well as detector feedback. We observe that all these objectives play a positive role in the final performance.

IV. Experiments

A. Implement Details

The framework is implemented using PyTorch and trained on a Linux workstation with GTX 3090 GPU. We adopt the architecture of ResNet [49] as our forward enhancement and backward degradation networks of the generator. Inspired by [50], we adopted patchGAN [51] to construct the discriminator, which outputs a binary map rather than a binary value.

In the training process, the twin adversarial enhancement and the task-aware feedback are pre-trained firstly. Specifically, UIEBD [6] and BSD500 [52] are employed in the baseline twin adversarial enhancement module, in which UIEBD consists of 640 training images, and BSD500 contains 500 images. We randomly crop them into 512×512 patch as the input. The twin enhancement network is pre-trained for 400 epochs, the learning rate of the first 100 epochs is set as $2 \times e^{-4}$ and then is linearly decreased to zero during the last 300 epochs. The pre-training data used in the task-aware feedback module is the latent enhanced results on RUIE [53] dataset obtained from the baseline twin adversarial enhancement. For the joint training with the specific detector, original RUIE [53] containing 2070 training images is employed, and 200 epochs are conducted totally, in which the learning rate of the first 100 epochs is $2 \times e^{-5}$ and also decayed to zero in the left 100 epochs. It is worth mentioning that, in the joint training phase, the parameters of both enhancement and feedback modules are updated. What's more, the batch size is set as 1 in both two training phases, and Adam is adopted as the optimizer.

B. Datasets and Metrics

For underwater image enhancement, we adopted the widely used UIEBD [6], UCCS [53], SQUID [59] and U45 [60] datasets to evaluate the performance. Specifically, UIEBD [6] contains 177 testing image pairs with the corresponding reference images derived from prolific methods, while UCCS [53] is captured during underwater catching, which releases 400 real-world underwater images. The environment of UCCS is more complicated with diverse underwater organisms, which is practical to measure the effect of enhancement. SQUID [59] and U45 [60] collect 57 and 45 underwater raw images, respectively, in which color casts, low contrast, and haze-like effects of underwater degradation are all covered in them.

In order to assess the improvement of the enhanced underwater images on detection accuracy, the annotated underwater detection dataset RUIE [53] and Aquarium[1] are employed. RUIE [53] consists of 2070 training and 676 testing images. The annotation of marine life, including sea cucumber, scallop, and sea urchin, is tagged by a number of experienced researchers in computer vision. So far, RUIE is the largest underwater detection dataset with annotation information. At the same time, the Aquarium dataset contains 448 training and 63 testing data captured from two aquariums and are also labeled for object detection, including fish, jellyfish, penguins, sharks, puffins, stingrays, and starfish.

For the measure assessments, Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM) are adopted as the fully-reference evaluations, in which PSNR is based on the max and mean square error to illustrate the ratio of the peak signal to the average energy, while SSIM refers to the brightness, contrast, and structure to metric the similarity between the target and the ground-truth images. In UIEBD [6] dataset, we suppose the reference that picked from a series of results obtained by prolific enhancement methods as the ground-truth images. In contrast, for UCCS [53] dataset, we conduce the non-reference image evaluation metrics Underwater Color Image Quality Evaluation (UCIQE) [61], Underwater Image Quality Measure (UIQM) [62] as well as Underwater Image Sharpness Measure (UISM). Firstly, UCIQE is based on chroma, contrast, and saturation of CIELab since the human perception is closely related to the variance of chroma for the degraded underwater images. UISM stands for the sharpness measure via the gray-scale edges of the target images. UIQM appraises the quality on the HSV model and takes three attributes into account, i.e., colorfulness, sharpness, and contrast. Measure values of the aforementioned metrics are all positive to the image quality. In addition to the image quality measure, we also evaluate the enhanced results on detection accuracy. Therefore, the mean Average Precision (mAP) and mean Intersection over Union (mIoU)

[1] https://public.roboflow.com/object-detection/aquarium

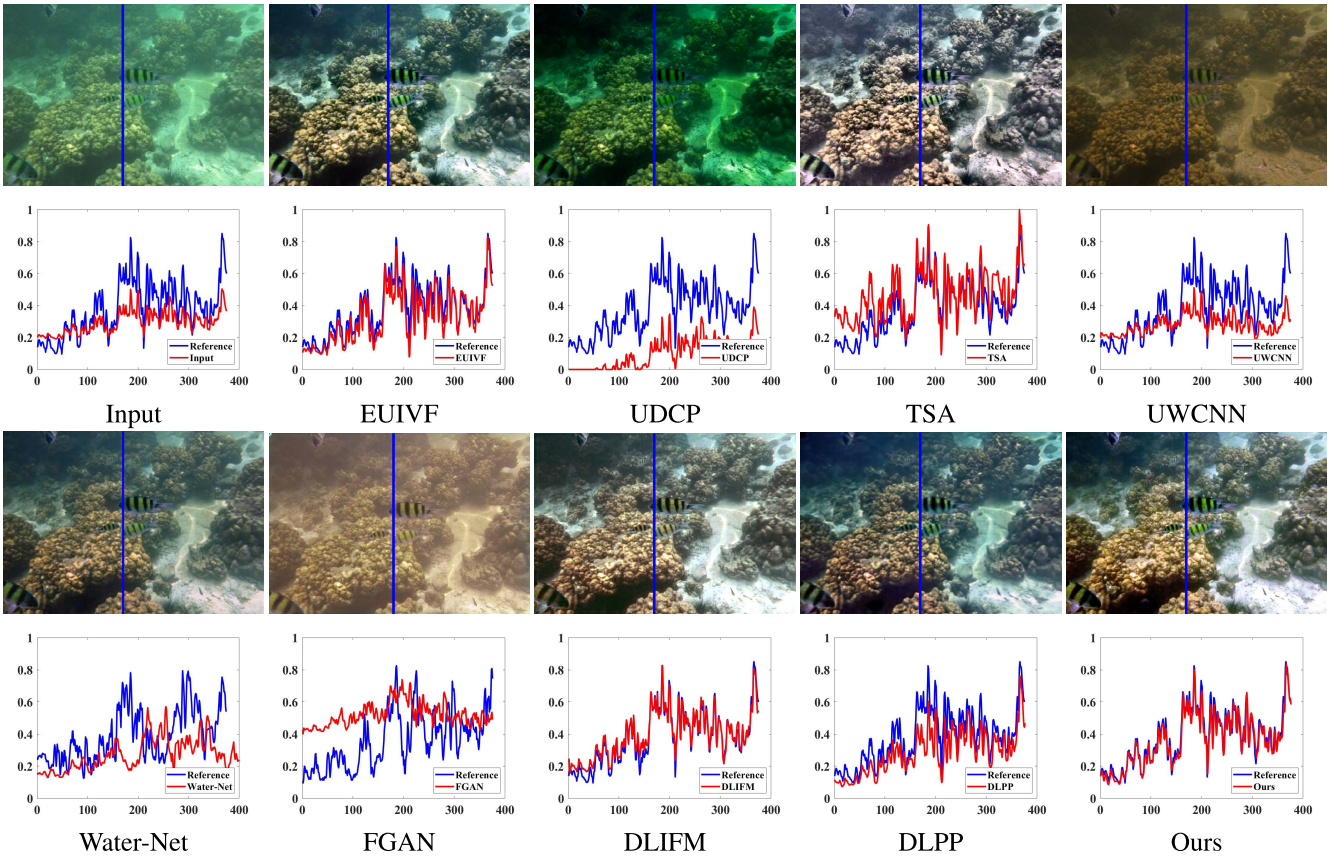

Fig. 2. Subjective comparisons on UIEBD datasets. Since the corresponding reference image is available, we analyze the data distribution between the reference and the enhanced results of different methods. It can be found that the data distribution of our result fits the reference better, revealing the superior performance of the proposed method on underwater image enhancement.

are used as the detection-specific evaluation metrics to assess the enhancement methods.

C. Evaluation on Underwater Image Enhancement

In this section, we conduct the experiments on different datasets and make comparison with diverse underwater image enhancement methods, including model based methods EUIVE [5], OCM [54], UDCP [55], TSA [56] and deep learning based methods UWCNN [57], Water-Net [6], FGAN [58], DLIFM [38] and DLPP [21]. Both subjective and objective results are adopted for analysis.

1) Qualitative Comparison: Fig. 2 presents the enhanced results on UIEBD [6] dataset. It is evident that the proposed method recovers the original reflection of the distorted scene, easing the undesirable greenish caused by light absorption and scattering. In contrast, the results of UDCP, UWCNN and FGAN fail to correct the intrinsic color. They present the deep green or dark brown in the reconstructed results. Water-Net and EUIVE show a limited effect on the enhancement, where the muddy phenomenon of distant scenes cannot be solved completely. The remaining methods TSA and DLIFM generate outstanding results, which is similar to the result obtained by our method. In order to compare the results of various methods further, we demonstrate the data distribution of a certain region and observe the fitting of different methods. In Fig 2, the graph below each subjective result delivers the corresponding distribution diagram. We can see that the red line of our result has the most overlapping areas with the blue reference line, indicating that the result of our method is closest to the reference.

For UCCS [53], subjective comparisons are presented in Fig. 3. Results of UDCP, TSA, UWCNN and FGAN suffer from severe color distortion, even though they adopt various light absorption models to simulate the color cast of the abysmal sea. EUIVE and OCM introduce extra noise interference, which would degrade the results inevitably. In contrast, Water-Net and DLIFM achieve relatively pleasant images, but the vivid appearance and informative details have not been restored. Nevertheless, our method generates images with not only bright reflections but also realistic properties without the artifacts introduction. To further compare the restoration of authentic reflection, we present the pixel value statistics on RGB space of the enhanced results in Fig. 4. We can see that the pixel distribution of our method is more symmetrical and reasonable than the others. Fig. 5 shows the comparison on the U45 dataset, where UDCP, TSA, and UWCNN introduce the degradation further. Water-Net shows a limited improvement. Since the page limitation, we do not illustrate the failure results obtained by FGAN. Compared with EUIVF, OCM, DLIFM, and DLPP, our method greatly restores the scene radiance and contrast, which is outstanding in the other methods.

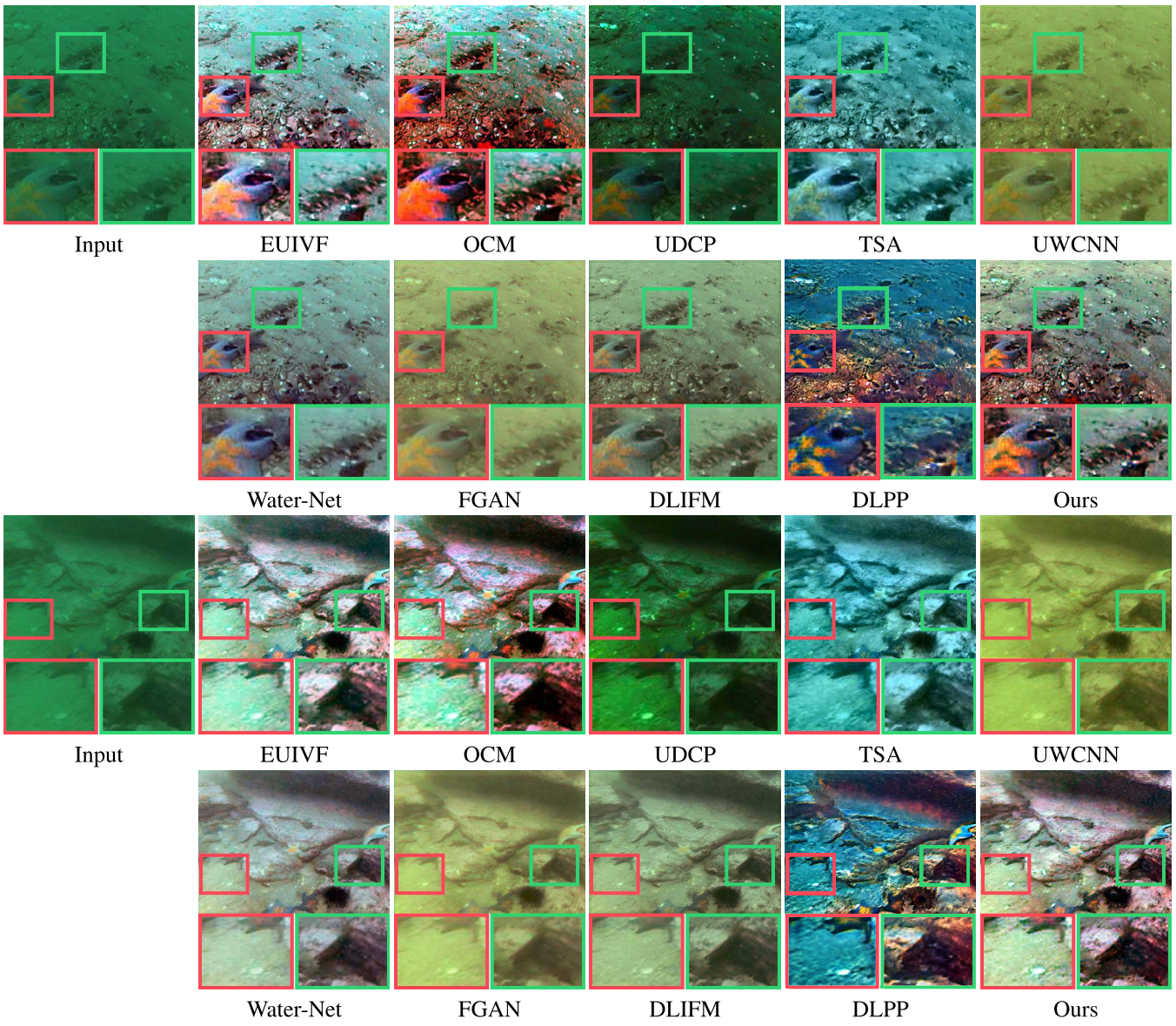

Fig. 3. Qualitative results of different methods on UCCS datasets. Obviously, the proposed method alleviates the muddy phenomenon and color cast to the greatest extent. In contrast, the other methods either remain greenish or introduce additional interference, which degrades the image quality further, especially the regions framed in red and green.

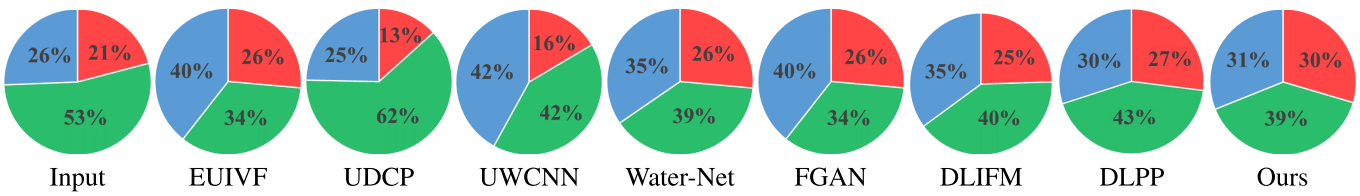

Fig. 4. Pixel value ration chart of the enhanced results on RGB color space, where the enhanced results are obtained from various underwater enhancement methods on UCCS datasets.

2) Quantitative Comparison: Table. I illustrates the quantitative results on four benchmarks. For the paired dataset UIEBD [6], fully-reference metrics PSNR and SSIM are adopted. Besides, the non-reference metrics UCIQE, UIQM and UISM are also taken into consideration on UIEBD [6], UCCS [53], SQUID [59] and U45 [60]. All these metrics values are positive to the quality. We can see that our method outperforms all competing methods except the UIQM. For UIEBD datasets, compared with the second best methods, our method achieves 2% improvement in terms of PSNR and SSIM, 10% improvement in UCIQE, and UISM. Although our method ranks second in UIQM, where we

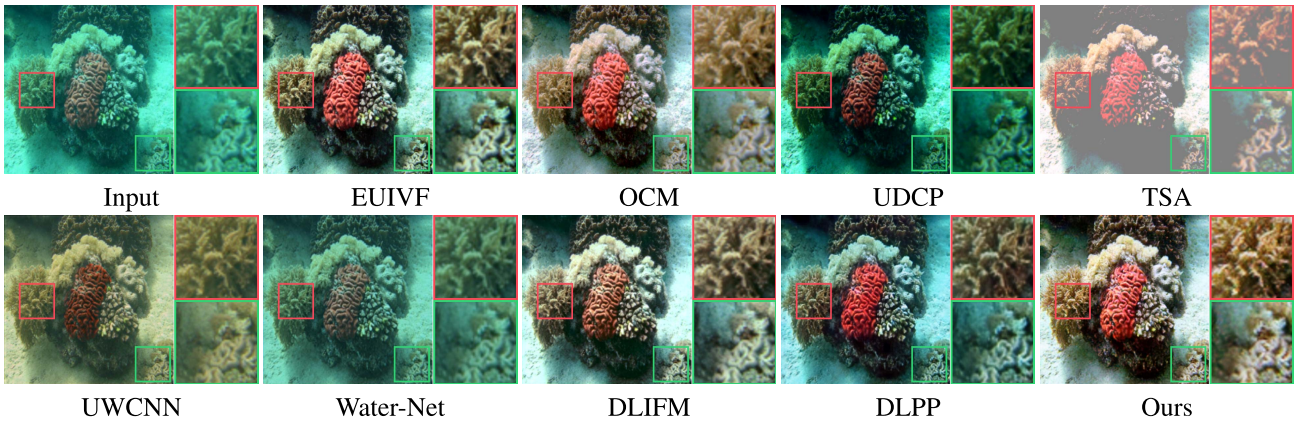

Fig. 5. Visual comparison on U45 dataset. Our method shows the superiority against the competitive methods on color correction and contrast improvement. Especially in the zoomed-in regions, the scene radiance and muddy phenomenon has been successfully restored and eased.

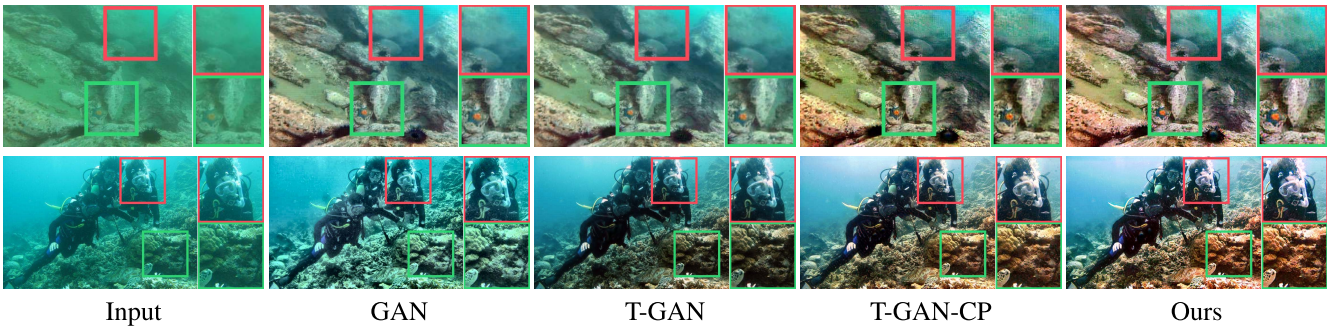

Fig. 6. Ablation study on the twin enhancement module. GAN is the onefold closed-loop network, T-GAN denotes the network with twin reverse cycle mapping, T-GAN-CP means adding the contrastive prior to the former, and Ours denotes the task-aware feedbacked enhanced result.

TABLE I

QUANTITATIVE COMPARISON OF THE ENHANCEMENT EFFECT WITH STATE-OF-THE-ART UNDERWATER IMAGE ENHANCEMENT METHODS ON UIEBD, UCCS, SQUID, AND U45 BENCHMARKS. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD

		Input	EUIVF [5]	OCM [54]	UDCP [55]	TSA [56]	UWCNN [57]	Water-Net [6]	FGAN [58]	DLIFM [38]	DLPP [21]	Ours
UIEBD	PSNR ↑	17.6024	21.9087	16.2096	11.7172	16.3445	13.3865	19.2890	15.5202	21.7325	20.5890	**22.3020**
	SSIM ↑	0.8143	0.8825	0.7950	0.5971	0.7953	0.7249	0.8645	0.5984	0.8796	0.8221	**0.8882**
	UCIQE ↑	0.5840	0.6151	0.6745	0.5852	0.5587	0.4730	0.5587	0.5379	0.6087	0.6196	**0.6835**
	UIQM ↑	2.4405	3.0803	3.8784	3.3724	**3.9100**	3.3992	3.1531	3.4710	3.1661	3.3863	3.5952
	UISM ↑	4.8178	5.3508	5.1003	4.6308	4.9002	4.6037	5.0808	5.0426	5.3308	5.6762	**5.7686**
UCCS	UCIQE ↑	0.4092	0.6180	0.6738	0.5255	0.5656	0.4530	0.5455	0.4928	0.5200	0.6083	**0.6883**
	UIQM ↑	0.1377	3.5596	**4.3793**	2.8851	3.9244	2.7100	3.1738	3.7565	2.9137	3.8538	3.9538
	UISM ↑	2.2957	4.8344	5.0064	3.1199	4.9677	3.0784	3.9393	4.9620	3.9925	4.8691	**5.0668**
SQUID	UCIQE ↑	0.3914	0.5760	0.5170	0.5394	0.5322	0.5228	0.3995	0.4456	0.5642	0.5723	**0.6383**
	UIQM ↑	1.9118	2.5499	2.3601	2.6668	2.2956	2.1850	1.0560	2.9156	2.8489	0.7077	**3.8520**
	UISM ↑	3.7806	7.0347	5.7638	5.2150	6.8584	6.8651	7.0261	6.6157	7.1037	3.1258	**7.1258**
U45	UCIQE ↑	0.4814	0.6427	0.4076	0.5999	0.5071	0.5190	0.4964	0.4452	0.588	0.6087	**0.6296**
	UIQM ↑	2.3472	4.0223	3.8770	3.8886	3.61906	3.5791	3.6457	4.1427	4.2764	3.1318	**4.4344**
	UISM ↑	6.3129	6.3651	6.9069	7.1157	4.2062	7.2078	7.2232	6.6498	7.2366	4.6645	**7.2564**

obtain 3.5952 and 3.9538 on UIEBD and UCCS datasets, respectively, the best performing methods, i.e., TSA and OCM, fail to obtain the best visual results since their best values on UIQM benefit from the amplified noise. In SUIQD and U45 datasets, the proposed method performs the best in terms of all three non-reference assessments, consistent with the visual results.

3) Study on Twin Enhancement Module: We conducted the ablation study to analyze how the twin adversarial contrastive learning based enhancement module influence the enhanced results, including the onefold closed-loop network, twin reverse networks, and contrastive principle. Fig. 6 illustrates the qualitative comparison of the enhancement module with different components. We can see that the onefold closed-loop network (denoted as GAN) alleviates the color cast phenomenon to a certain degree in the first sample while remaining bluish scene in the second one. The twin reverse network, which learns the reverse cycle mapping further (denoted as

Fig. 7. Visual comparisons among the proposed contrastive loss $\mathcal{L}_{cp}$ and the general $\mathcal{L}_1, \mathcal{L}_2$ loss, where our method achieves the most plausible and credible reconstruction against the others.

TABLE II

UNDERWATER IMAGE QUALITY EVALUATION ON THE ENHANCED RESULTS OBTAINED FROM DIFFERENT COMPONENTS. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD

Models	PSNR ↑	SSIM ↑	UCIQE ↑	UIQM ↑	UISM ↑
GAN	19.4925	0.7765	0.6324	3.0989	4.6364
T-GAN	20.8762	0.8305	0.6037	3.2135	5.6797
T-GAN-CP	21.9675	0.8832	0.6440	3.3287	5.7447
Ours	**22.3020**	**0.8882**	**0.6835**	**3.3952**	**5.7686**

TABLE III

QUANTITATIVE EVALUATION OF THE PROPOSED CONTRASTIVE PRIOR WITH $\mathcal{L}_1, \mathcal{L}_2$ LOSS ON UIEBD DATASET

Models	PSNR ↑	SSIM ↑	UCIQE ↑	UIQM ↑	UISM ↑
Input	17.6024	0.8143	0.5840	2.4405	4.8178
$\mathcal{L}_1$	20.4721	0.8436	0.6190	2.9668	5.3484
$\mathcal{L}_2$	20.0190	0.8326	0.5313	3.1560	4.9751
Ours	**22.3020**	**0.8882**	**0.6835**	**3.3952**	**5.7686**

T-GAN), performs better in the color correction. Considering the contrastive principle, we equip the contrastive prior into the twin network (denoted as T-GAN-CP). Vivid appearance and stretched details can be restored in the results. At last, the enhanced results trained with the task-aware feedback produce the clearest and the most reductive results of the original scene. In addition to the qualitative comparison, we also provide the quantitative results with two reference and three non-reference metrics in Table. II. We observe the superiority of the whole framework with task-aware feedback over the other latent models in all five metrics, indicating that the enhanced results of our method are the closest to the reference images. Removing any component in the whole framework attenuates the enhanced results evidently.

4) Study on Contrastive Prior: To evaluate the effectiveness of our proposed contrastive prior, i.e., $\mathcal{L}_{cp}$, we conduct the ablation comparison with $\mathcal{L}_1$ and $\mathcal{L}_2$ losses. Visual results are illustrated in Fig. 7 and the quantitative assessments are reported in Table. III. We can see that although $\mathcal{L}_1$, and $\mathcal{L}_2$ losses realize a certain degree of improvement on the degraded underwater images, they still remain slight chromatic aberration and muddy phenomenon compared with our method. Since the proposed contrastive prior takes both degraded negatives and the high-quality positives into consideration, the reconstructed results are more away from the underwater appearance, leading to the plausible and credible reconstruction against the others. According to the quantitative results in Table. III, a consistent conclusion can be found. Therefore, the contrast prior we proposed performs a significant improvement effect on the enhancement of underwater image quality.

D. Evaluation on Detection Task

1) Detection Accuracy Comparison: In order to confirm the effect of underwater image enhancement on the subsequent detection task, we apply the enhanced results to a series of detection algorithms, including one-stage methods SSD [46], RetinaNet [47], GIoU [48] and two-stage method S-R-CNN [1]. Table. IV and Table. V report the quantitative results on RUIE and Aquarium datasets respectively. We treat the enhanced results of prolific enhancement algorithms as an input image to conduct the various detection methods. It can be seen that for the four detectors we adopted, the proposed method outperforms the other competitive methods in terms of detection accuracy. The visualized detection results on the RUIE dataset are presented in Fig. 8, which is consistent with the objective results. Note that in Fig. 8, although the enhanced results of EUIVF, TSA, Water-Net as well as DLIFM appear the visual improvement, they meet with a performance drop on detection still. In contrast, our method performs the best since the proposed framework facilitates the enhancement results toward detection-favorable appearance. Fig. 9 delivers the visual comparison on the Aquarium dataset, where the proposed method beats the others in accuracy and quantity.

2) Study on Task-Aware Feedback: The task-aware feedback module aims at propagating the coherent information of the detector and forcing the enhancement module towards the detection-favorable direction. In this section, we compare the detection accuracy of diverse detectors on the enhanced results of baseline twin enhancement networks, denoted as T-GAN-CP, and the task-aware feedback equipped network denoted as Ours, respectively. Results on RUIE dataset are presented in Table. VI and the visual comparisons of SSD detector are demonstrated in Fig. 10. From Table. VI, it is obvious that the mAP and mIoU values of the proposed method are far ahead of the baseline enhancement method, revealing the effectiveness of the proposed task-aware feedback module to produce the enhanced images more suitable for detection. With the goal of understanding how the features of the three classes are distributed and whether the enhancement module affects the object features, we visualize the tSNE projection results of the detected features in different classes, including trepang, urchin and scallop. Fig. 11 shows the tSNE projection results. It can be seen that in the first two cases, features of different aquatics are projected in a neighbouring domain, sharing a communal region in a certain degree. While in the last case, projection of our method is more separable across different categories,

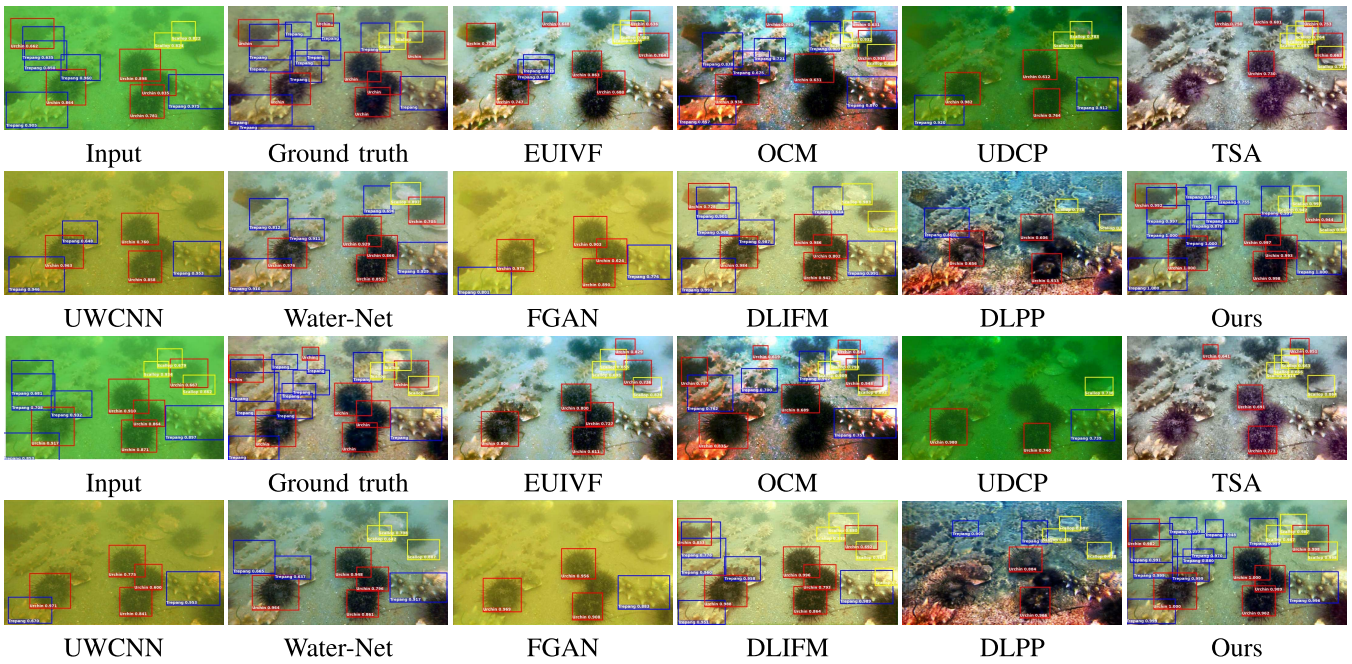

Fig. 8. The display of visualized accuracy about the results of applying different enhanced underwater images obtained by series enhancement methods to the S-R-CNN detection algorithms. Although EUIVF, TSA as well as Water-Net present obvious visual improvement, they fail to benefit the performance on detection. In contrast, the proposed method detects the most objects and has the highest confidence.

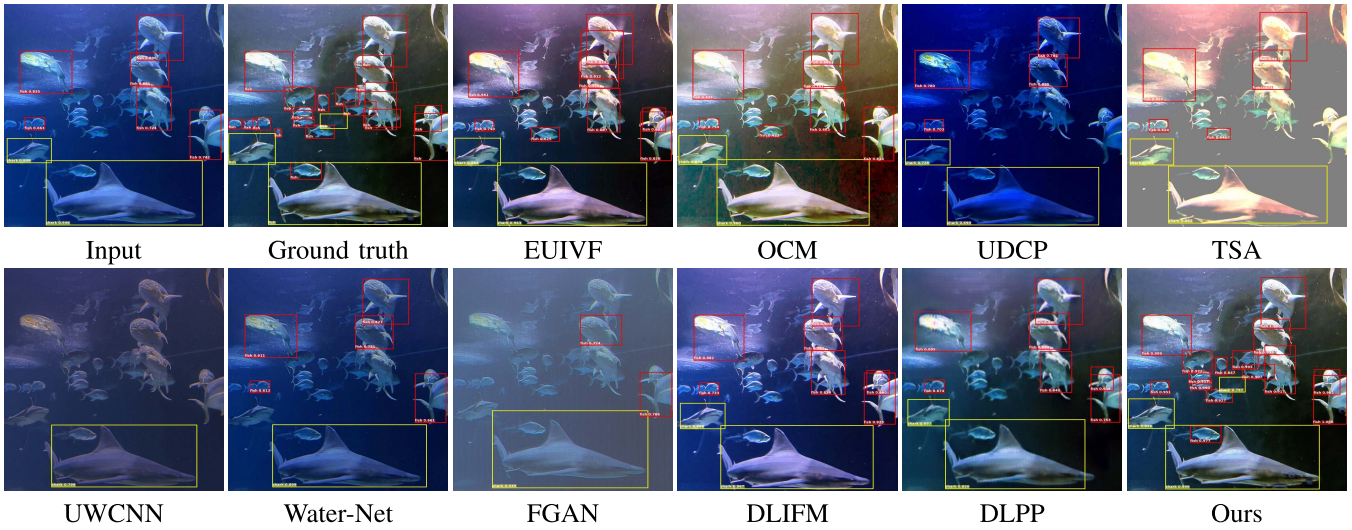

Fig. 9. Detection evaluation on Aquarium dataset with different enhancement results, where we employ the S-R-CNN detection algorithm. Note that the enhanced result of our method shows significant advantages in detection improvement against the others.

they share a huge difference between each clustering and the distinction is obvious. This might explain that our introduced task-aware feedback module transmit the object guidance to benefit the enhancement results more suitable for detection, where the object features are more separable than the original underwater images and the general enhancement images.

Additionally, the Precision-Recall (PR) and ROC curves of S-R-CNN [1] on RUIE datasets are illustrated in Fig. 12. Evidently, the proposed method performs the best across all categories, including trepang, urchin as well as scallop, indicating that the proposed task-aware feedback enforces the enhanced results with more detection-oriented information. In Fig. 10, we find that the visual effect of the whole framework with object guidance has also been ameliorated, showing the mutual improvement between the enhancement module and the feedback module.

3) Study on Detection Improvement: To validate the improvement of detection algorithms within our whole framework, where the detector is jointly-trained with the enhancement module, we enumerate the detection accuracy of using the original detector with degraded underwater images and the jointly-trained detector with the enhanced images. Table. VII presents the quantitative results on RUIE datasets. We adopt the aforementioned four detection methods respectively and

TABLE IV

DETECTION ACCURACY ON THE ENHANCED RESULTS FROM DIFFERENT UNDERWATER ENHANCEMENT METHODS ON RUIE DATASET

		EUIVF [5]	OCM [54]	UDCP [55]	TSA [56]	UWCNN [57]	Water-Net [6]	FGAN [58]	DLIFM [38]	DLPP [21]	Ours
SSD [46]	mAP (%) ↑	74.15	70.25	40.31	69.29	46.40	75.19	41.16	74.44	77.42	**90.73**
	mIoU (%)↑	74.27	73.52	71.05	73.35	72.54	75.58	71.19	75.03	74.51	**80.94**
RetinaNet [47]	mAP (%) ↑	81.69	80.94	51.46	80.14	76.21	82.14	70.59	77.92	80.33	**96.08**
	mIoU (%)↑	78.16	75.60	73.51	77.91	71.02	79.44	74.63	70.90	79.65	**88.92**
GIoU [48]	mAP (%) ↑	87.47	85.96	80.98	83.25	78.67	88.88	66.53	88.12	71.31	**89.59**
	mIoU (%)↑	81.24	79.56	78.87	79.68	78.43	81.92	75.55	80.92	75.24	**83.70**
S-R-CNN [1]	mAP (%) ↑	83.42	81.55	67.60	81.33	90.44	81.93	79.81	79.00	82.37	**96.09**
	mIoU (%)↑	80.61	78.93	79.31	79.88	86.21	75.60	72.65	81.04	80.55	**89.01**

TABLE V

DETECTION ACCURACY ON THE ENHANCED RESULTS FROM DIFFERENT UNDERWATER ENHANCEMENT METHODS ON AQUARIUM DATASET

		EUIVF [5]	OCM [54]	UDCP [55]	TSA [56]	UWCNN [57]	Water-Net [6]	FGAN [58]	DLIFM [38]	DLPP [21]	Ours
SSD [46]	mAP (%) ↑	71.12	70.04	61.46	69.62	34.75	74.33	31.45	73.36	72.60	**82.33**
	mIoU (%)↑	74.07	71.13	66.13	72.41	45.26	76.14	40.14	73.95	80.45	**86.60**
RetinaNet [47]	mAP (%) ↑	80.12	78.91	70.54	70.13	71.44	77.66	59.14	73.75	79.52	**81.46**
	mIoU (%)↑	81.10	81.65	71.05	72.42	75.96	81.01	61.12	81.12	81.52	**85.97**
GIoU [48]	mAP (%) ↑	72.51	70.25	65.92	66.55	72.31	70.09	40.78	70.28	71.44	**75.12**
	mIoU (%)↑	73.16	71.67	70.25	75.21	77.91	75.46	46.07	80.21	75.21	**81.61**
S-R-CNN [1]	mAP (%) ↑	80.25	82.62	69.81	77.14	79.42	80.16	60.15	80.14	81.51	**85.61**
	mIoU (%)↑	83.61	82.75	72.44	79.51	81.83	81.02	63.34	81.13	82.17	**86.21**

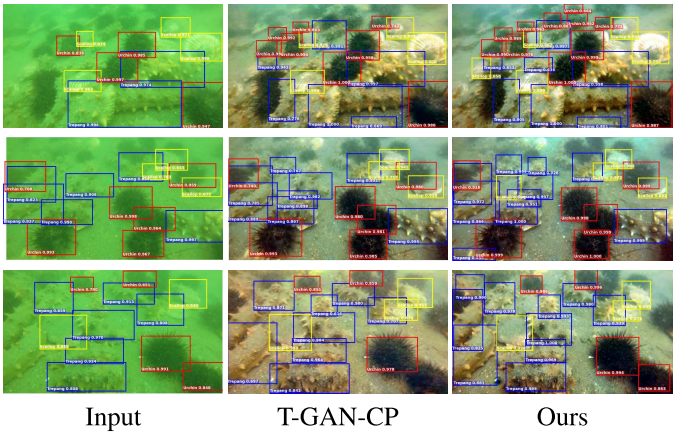

Fig. 10. Detection comparison between the baseline twin enhancement module (T-GAN-CP) with the proposed method (Ours). The proposed task-aware feedback module is able to generate enhanced results with detection-favorable features and promote detection accuracy greatly.

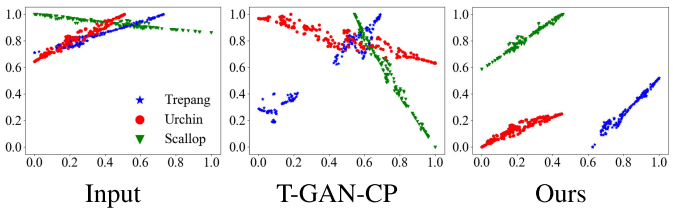

Fig. 11. The tSNE visualization comparison among the detection features of the 3 classes. Note that detected objections are more separable in our method.

asterisk the detector to distinguish the jointly-trained detectors. It is obvious that the detection accuracy of jointly-trained with the enhancement module outperforms the performance

TABLE VI

ANALYSIS OF THE TASK-AWARE FEEDBACK MODULE. T-GAN-CP MEANS THE BASELINE ENHANCEMENT MODULE, AND OURS DENOTES THE WHOLE NETWORK EQUIPPED WITH A FEEDBACK MODULE

		SSD [46]	RetinaNet [47]	GIoU [48]	S-R-CNN [1]
mAP (%)↑	T-GAN-CP	88.46	94.10	88.65	93.12
	Ours	**90.73**	**96.08**	**89.59**	**96.09**
mIoU(%)↑	T-GAN-CP	78.75	85.89	82.67	85.44
	Ours	**80.94**	**88.92**	**83.70**	**89.01**

of fitting the detector with raw degraded images. That is, the network re-trained with degraded underwater images is inferior to feeding the jointly-trained network with detection-specific enhanced data, revealing that the proposed framework promotes the performance improvement for detection algorithms.

E. Evaluation on Other Application

As analyzed above, the proposed task-aware feedback module transmits the high-level object information into the enhancement module, which benefits the performance of detection algorithms in the underwater environment. This motivates us to investigate the effect of the enhancement performance on other applications, such as semantic segmentation. Concretely, we adopt the underwater segmentation dataset SUIM [63] and test the segmentation accuracy among the original underwater image, our baseline enhancement model (denoted as T-GAN-CP) as well as the proposed object-guided twin adversarial contrastive enhancement network (denoted as Ours) within the SUIM-Net [63]. Results are demonstrated in Fig. 13 and its corresponding quantitative values on mIoU and mPA are reported in Table. VIII.

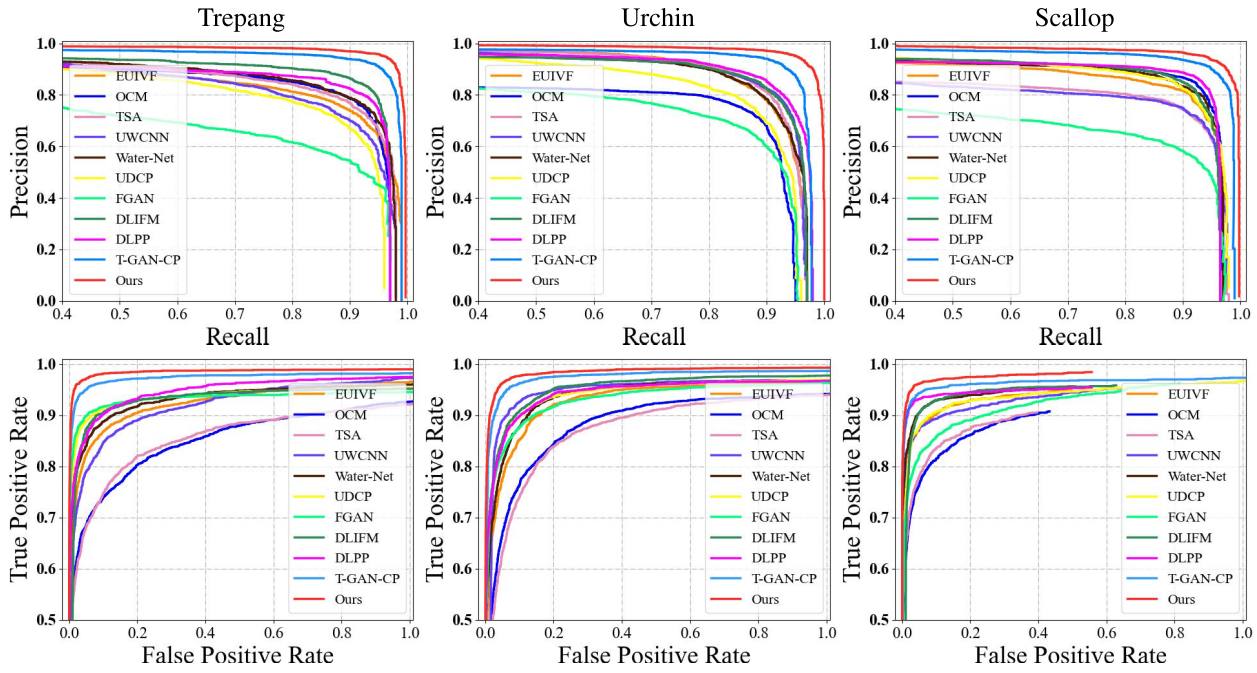

Fig. 12. PR and ROC curves of S-R-CNN performing on the enhanced images of different underwater image enhancement methods.

TABLE VII

DETECTION ACCURACY COMPARISON BETWEEN THE DETECTION MODELS RE-TRAINED WITH RAW UNDERWATER IMAGES AND THE JOINTLY-TRAINED MODELS PLUGGED IN OUR FRAMEWORK. THE ASTERISK INDICATES THAT THE CORRESPONDING DETECTION ALGORITHM IS JOINTLY-TRAINED WITHIN OUR FRAMEWORK AND ADOPTS THE ENHANCED IMAGE OBTAINED BY OUR METHOD AS INPUT. THE RE-TRAINED MODEL IS FED WITH THE RAW UNDERWATER IMAGES AS INPUT

	SSD [46]	SSD* [46]	RetinaNet [47]	RetinaNet* [47]	GIoU [48]	GIoU* [48]	S-R-CNN [1]	S-R-CNN* [1]
mAP (%) ↑	90.61	90.73	95.98	96.08	89.46	89.59	96.00	96.09
mIoU (%)↑	80.79	80.94	88.15	88.92	83.55	83.70	88.49	89.01

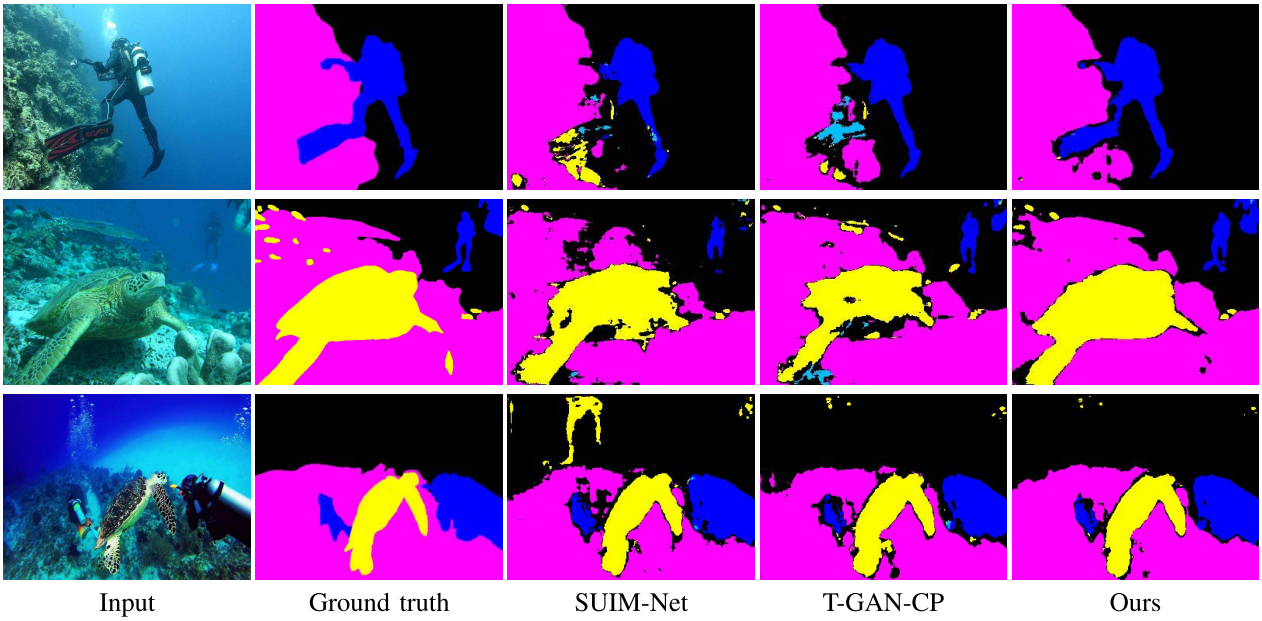

Fig. 13. Evaluation of semantic segmentation. The first and second columns denote the input underwater images and their corresponding label mask. The third column shows the segmentation results of SUIM. T-GAN-CP denotes the results obtained from the baseline twin adversarial contrastive enhancement model without the task-aware feedback compared to Ours.

Even though the SUIM-Net [63] is developed for underwater images specifically, the segmentation accuracy is inferior to the result of conducting on the enhanced image. In contrast, the whole framework performs best against the baseline

TABLE VIII

COMPARISON RESULTS OF SEMANTIC SEGMENTATION. T-GAN-CP DENOTES THE RESULTS OF THE TWIN ADVERSARIAL CONTRASTIVE ENHANCEMENT MODEL WITHOUT TASK-AWARE FEEDBACK

	SUIM-Net [63]	T-GAN-CP	Ours
mPA (%) ↑	56.0678	64.5897	**80.7543**
	56.0479	65.5406	**83.2925**
	76.5700	80.1180	**81.2412**
mIoU (%) ↑	67.8889	71.3695	**87.7071**
	61.0348	72.9302	**86.2925**
	77.6320	79.9876	**85.4668**

enhanced images, indicating that the task-aware feedback module confers the enhanced images with more high-level perceptual features and promotes the performance of subsequent applications.

V. CONCLUSION

This paper developed an object-guided twin adversarial contrastive learning based underwater enhancement method, where the enhancement achieved by this framework is more suitable for detection. Firstly, a twin adversarial constrained enhancement module is proposed to implement the transformation between the degraded underwater domain and high-quality clear domain, in which the reverse closed-loop mappings in a self-learning manner are developed to eliminate the reliance on paired training data. Besides, the contrastive principle is also introduced into the training process to enforce the results with more realistic appearance. To impart more detection-favorable features into the enhanced results, we equip the enhancement module with a task-aware feedback module to propagate the object guidance. Experiments show that our method is superior to other top-performing methods in terms of quality improvement and detection accuracy. What's more, the effect on other downstream vision task (i.e., semantic segmentation) has also been proven.

REFERENCES

[1] P. Sun *et al.*, "Sparse R-CNN: End-to-end object detection with learnable proposals," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 14454–14463.

[2] J. Liu *et al.*, "Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2022, pp. 5802–5811.

[3] A. Srinivas, T.-Y. Lin, N. Parmar, J. Shlens, P. Abbeel, and A. Vaswani, "Bottleneck transformers for visual recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 16519–16529.

[4] H. Wang, Y. Zhu, H. Adam, A. Yuille, and L.-C. Chen, "MaX-DeepLab: End-to-end panoptic segmentation with mask transformers," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 5463–5474.

[5] C. Ancuti, C. O. Ancuti, T. Haber, and P. Bekaert, "Enhancing underwater images and videos by fusion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 81–88.

[6] C. Li *et al.*, "An underwater image enhancement benchmark dataset and beyond," *IEEE Trans. Image Process.*, vol. 29, pp. 4376–4389, 2020.

[7] Z. Jiang, Z. Li, S. Yang, X. Fan, and R. Liu, "Target oriented perceptual adversarial fusion network for underwater image enhancement," *IEEE Trans. Circuits Syst. Video Technol.*, early access, May 13, 2022, doi: 10.1109/TCSVT.2022.3174817.

[8] 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, Sep. 2010.

[9] J. Y. Chiang and Y.-C. Chen, "Underwater image enhancement by wavelength compensation and dehazing," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1756–1769, Apr. 2011.

[10] Y.-T. Peng, K. Cao, and P. C. Cosman, "Generalization of the dark channel prior for single image restoration," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2856–2868, Jun. 2018.

[11] A. Galdran, D. Pardo, A. Picón, and A. Alvarez-Gila, "Automatic red-channel underwater image restoration," *J. Vis. Commun. Image Represent.*, vol. 26, pp. 132–145, Jan. 2015.

[12] C. Li, J. Guo, C. Guo, R. Cong, and J. Gong, "A hybrid method for underwater image correction," *Pattern Recognit. Lett.*, vol. 94, pp. 62–67, Jul. 2017.

[13] H. Liu and L.-P. Chau, "Underwater image restoration based on contrast enhancement," in *Proc. IEEE Int. Conf. Digit. Signal Process. (DSP)*, Oct. 2016, pp. 584–588.

[14] Y. Wang, H. Liu, and L.-P. Chau, "Single underwater image restoration using adaptive attenuation-curve prior," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 65, no. 3, pp. 992–1002, Mar. 2018.

[15] P. Zhuang, C. Li, and J. Wu, "Bayesian retinex underwater image enhancement," *Eng. Appl. Artif. Intell.*, vol. 101, May 2021, Art. no. 104171.

[16] A. S. A. Ghani and N. A. M. Isa, "Automatic system for improving underwater image contrast and color through recursive adaptive histogram modification," *Comput. Electron. Agric.*, vol. 141, pp. 181–195, Sep. 2017.

[17] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and P. Bekaert, "Color balance and fusion for underwater image enhancement," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 379–393, Jan. 2017.

[18] S.-B. Gao, M. Zhang, Q. Zhao, X.-S. Zhang, and Y.-J. Li, "Underwater image enhancement using adaptive retinal mechanisms," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5580–5595, Nov. 2019.

[19] 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.

[20] C. Li, S. Anwar, J. Hou, R. Cong, C. Guo, and W. Ren, "Underwater image enhancement via medium transmission-guided multi-color space embedding," *IEEE Trans. Image Process.*, vol. 30, pp. 4985–5000, 2021.

[21] L. Chen *et al.*, "Perceptual underwater image enhancement with deep learning and physical priors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 8, pp. 3078–3092, Aug. 2021.

[22] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[23] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 7263–7271.

[24] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.

[25] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.

[26] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[27] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Sys.*, vol. 28, 2015, pp. 91–99.

[28] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. ICCV*, 2017, pp. 2961–2969.

[29] X. Gong, S. Chang, Y. Jiang, and Z. Wang, "AutoGAN: Neural architecture search for generative adversarial networks," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 3224–3234.

[30] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Generative image inpainting with contextual attention," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5505–5514.

[31] J. Liu, J. Shang, R. Liu, and X. Fan, "Attention-guided global-local adversarial learning for detail-preserving multi-exposure image fusion," *IEEE Trans. Circuits Syst. Video Technol.*, early access, Jan. 18, 2022, doi: 10.1109/TCSVT.2022.3144455.

[32] R. Qian, R. T. Tan, W. Yang, J. Su, and J. Liu, "Attentive generative adversarial network for raindrop removal from a single image," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2482–2491.

[33] C. K. Sønderby, J. Caballero, L. Theis, W. Shi, and F. Huszár, "Amortised map inference for image super-resolution," in *Proc. IEEE Int. Conf. Learn. Represent.*, Jun. 2017, pp. 1–17.
[34] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. IEEE Int. Conf. Mach. Learn.*, Nov. 2020, pp. 1597–1607.
[35] O. Henaff, "Data-efficient image recognition with contrastive predictive coding," in *Proc. IEEE Int. Conf. Mach. Learn.*, Nov. 2020, pp. 4182–4192.
[36] T. Park, A. A. Efros, R. Zhang, and J.-Y. Zhu, "Contrastive learning for unpaired image-to-image translation," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 319–345.
[37] H. Wu *et al.*, "Contrastive learning for compact single image dehazing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10551–10560.
[38] X. Chen, P. Zhang, L. Quan, C. Yi, and C. Lu, "Underwater image enhancement based on deep learning and image formation model," 2021, *arXiv:2101.00991*.
[39] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. IEEE Int. Conf. Learn. Represent.*, 2015, pp. 1–14.
[40] M. S. Rad, B. Bozorgtabar, U.-V. Marti, M. Basler, H. K. Ekenel, and J.-P. Thiran, "SROBB: Targeted perceptual loss for single image super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2710–2719.
[41] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 694–711.
[42] J. Liu, X. Fan, J. Jiang, R. Liu, and Z. Luo, "Learning a deep multi-scale feature ensemble and an edge-attention guidance for image fusion," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 1, pp. 105–119, Jan. 2021.
[43] R. Liu, J. Liu, Z. Jiang, X. Fan, and Z. Luo, "A bilevel integrated model with data-driven layer ensemble for multi-modality image fusion," *IEEE Trans. Image Process.*, vol. 30, pp. 1261–1274, 2020.
[44] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
[45] H. Zhang, H. Chang, B. Ma, N. Wang, and X. Chen, "Dynamic R-CNN: Towards high quality object detection via dynamic training," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 260–275.
[46] W. Liu *et al.*, "SSD: Single shot MultiBox detector," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 21–37.
[47] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Trans. Pattern. Anal. Mach. Intell.*, no. 99, pp. 2999–3007, 2017.
[48] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 658–666.
[49] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[50] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2223–2232.
[51] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1125–1134.
[52] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 2, Jun. 2001, pp. 416–423.
[53] R. Liu, X. Fan, M. Zhu, M. Hou, and Z. Luo, "Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 12, pp. 4861–4875, Dec. 2020.
[54] C.-Y. Li, J.-C. Guo, R.-M. Cong, Y.-W. Pang, and B. Wang, "Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior," *IEEE Trans. Image Process.*, vol. 25, no. 12, pp. 5664–5677, Dec. 2016.
[55] P. Drews-Jr, E. R. Nascimento, S. S. C. Botelho, and M. F. M. Campos, "Underwater depth estimation and image restoration based on single images," *IEEE Comput. Graph. Appl.*, vol. 36, no. 2, pp. 24–35, Mar./Apr. 2016.
[56] X. Fu, Z. Fan, M. Ling, Y. Huang, and X. Ding, "Two-step approach for single underwater image enhancement," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst. (ISPACS)*, Nov. 2017, pp. 789–794.
[57] C. Li, S. Anwar, and F. Porikli, "Underwater scene prior inspired deep underwater image and video enhancement," *Pattern Recognit.*, vol. 98, Feb. 2020, Art. no. 107038.
[58] M. J. Islam, Y. Xia, and J. Sattar, "Fast underwater image enhancement for improved visual perception," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 3227–3234, Apr. 2020.
[59] 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. 2020.
[60] H. Li, J. Li, and W. Wang, "A fusion adversarial underwater image enhancement network with a public test dataset," 2019, *arXiv:1906.06819*.
[61] M. Yang and A. Sowmya, "An underwater color image quality evaluation metric," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 6062–6071, Dec. 2015.
[62] K. Panetta, C. Gao, and S. Agaian, "Human-visual-system-inspired underwater image quality measures," *IEEE J. Ocean. Eng.*, vol. 41, no. 3, pp. 541–551, Jul. 2015.
[63] M. J. Islam *et al.*, "Semantic segmentation of underwater imagery: Dataset and benchmark," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, Oct. 2020, pp. 1769–1776.

Risheng Liu (Member, IEEE) received the B.Sc. and Ph.D. degrees from the Dalian University of Technology, China, in 2007 and 2012, respectively. From 2010 to 2012, he was doing research as a joint Ph.D. graduate with the Robotics Institute, Carnegie Mellon University. From 2016 to 2018, he was doing research as a Hong Kong Scholar with The Hong Kong Polytechnic University. He is currently a Full Professor with the Digital Media Department, International School of Information Science and Engineering, Dalian University of Technology (DUT). His research interests include optimization, computer vision, and multimedia. He was awarded the "Outstanding Youth Science Foundation" of the National Natural Science Foundation of China. He serves as an Associate Editor for *Pattern Recognition*, *The Visual Computer*, and *IET Image Processing*; and a Senior Editor for the *Journal of Electronic Imaging*.

Zhiying Jiang received the B.E. degree in software engineering from Dalian Maritime University, China, in 2017, and the M.S. degree in software engineering from the Dalian University of Technology, China, in 2020, where she is currently pursuing the Ph.D. degree in software engineering with the Dalian University of Technology. She is with the Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, Dalian University of Technology. Her research interests include computer vision, image restoration, and image stitching.

Shuzhou Yang is currently pursuing the B.E. degree in software engineering. He is with the DUT-RU International School of Information Science and Engineering, Dalian University of Technology, Dalian, China. His research interests include image restoration, 3D reconstruction, and computer vision.

Xin Fan (Senior Member, IEEE) received the B.E. and Ph.D. degrees in information and communication engineering from Xi'an Jiaotong University, Xi'an, China, in 1998 and 2004, respectively. He was with Oklahoma State University, Stillwater, from 2006 to 2007, as a Postdoctoral Research Fellow. He joined the School of Software, Dalian University of Technology, Dalian, China, in 2009. His current research interests include computational geometry and machine learning, and their applications to low level image processing and DTI-MR image analysis.