

# Domain Adaptation for Underwater Image Enhancement

Zhengyong Wang<sup>ID</sup>, Liquan Shen<sup>ID</sup>, Mai Xu<sup>ID</sup>, Senior Member, IEEE, Mei Yu<sup>ID</sup>, Kun Wang<sup>ID</sup>, and Yufei Lin<sup>ID</sup>

**Abstract**— Recently, learning-based algorithms have shown impressive performance in underwater image enhancement. Most of them resort to training on synthetic data and obtain outstanding performance. However, these deep methods ignore the significant domain gap between the synthetic and real data (i.e., inter-domain gap), and thus the models trained on synthetic data often fail to generalize well to real-world underwater scenarios. Moreover, the complex and changeable underwater environment also causes a great distribution gap among the real data itself (i.e., intra-domain gap). However, almost no research focuses on this problem and thus their techniques often produce visually unpleasing artifacts and color distortions on various real images. Motivated by these observations, we propose a novel Two-phase Underwater Domain Adaptation network (TUDA) to simultaneously minimize the inter-domain and intra-domain gap. Concretely, in the first phase, a new triple-alignment network is designed, including a translation part for enhancing realism of input images, followed by a task-oriented enhancement part. With performing image-level, feature-level and output-level adaptation in these two parts through jointly adversarial learning, the network can better build invariance across domains and thus bridging the inter-domain gap. In the second phase, an easy-hard classification of real data according to the assessed quality of enhanced images is performed, in which a new rank-based underwater quality assessment method is embedded. By leveraging implicit quality information learned from rankings, this method can more accurately assess the perceptual quality of enhanced images. Using pseudo labels from the easy part, an easy-hard adaptation technique is then conducted to effectively decrease the intra-domain gap between easy and hard samples. Extensive experimental results demonstrate that the proposed

Manuscript received 25 August 2021; revised 24 October 2022; accepted 4 February 2023. Date of publication 17 February 2023; date of current version 28 February 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 61931022, Grant 62271301, Grant 62271276, and Grant 62071266; in part by the Shanghai Science and Technology Program under Grant 22511105200; in part by the Natural Science Foundation of Shandong Province under Grant ZR2022ZD38; and in part by the Open Fund of Key Laboratory of Advanced Display and System Applications of Ministry of Education, Shanghai University. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Wangmeng Zuo. (Corresponding author: Liquan Shen.)

Zhengyong Wang, Kun Wang, and Yufei Lin are with the School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China (e-mail: zywang@shu.edu.cn; fangfang\_tu@163.com; anzhanglyf@163.com).

Liquan Shen is with the Key Laboratory of Specialty Fiber Optics and Optical Access Networks, Joint International Research Laboratory of Specialty Fiber Optics and Advanced Communication, Shanghai University, Shanghai 200444, China (e-mail: jsslq@163.com).

Mai Xu is with the School of Electronic and Information Engineering, Beihang University, Beijing 100191, China (e-mail: maiyu@buaa.edu.cn).

Mei Yu is with the Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China (e-mail: yumei@nbu.edu.cn).

This article has supplementary downloadable material available at <https://doi.org/10.1109/TIP.2023.3244647>, provided by the authors

Digital Object Identifier 10.1109/TIP.2023.3244647

TUDA is significantly superior to existing works in terms of both visual quality and quantitative metrics.

**Index Terms**— Underwater image enhancement, inter-domain adaptation, intra-domain adaptation, rank-based underwater image quality assessment.

## I. INTRODUCTION

IN THE underwater, the captured images always suffer from several kinds of degradation, including blurriness, color casts and low contrast. As light travels in the water, red light, which has longer wavelength than green and blue light, is absorbed faster, and thus underwater images often appear in a typical bluish or greenish tone. Furthermore, large amounts of suspended particles often change the direction of light in the water, resulting in dim and fuzzy effects. Therefore, excellent underwater image enhancement (UIE) methods are expected to improve low visibility, eliminate color casts and stretch low contrast, which can effectively enhance visual quality of input images. Meanwhile, the enhanced visibility can make scenes and objects more highlighted, providing a better starting point for various high-level computer vision tasks such as underwater image object detection and recognition.

In the past decades, many algorithms have been proposed to enhance underwater images, ranging from traditional methods (image-based [1], [2], [3], [4], [5], [6], [7], [8], [9], [10] and physical-based [11], [12], [13], [14], [15], [16], [17], [18], [19]) to learning-based methods [20], [21], [22], [23], [24], [25], [26], [27]. Among them, learning-based UIE methods using large training data have shown dramatic improvements over traditional methods. Unlike other low-level vision tasks such as deraining and dehazing, since it is difficult to obtain clean dewatered images of the same underwater scene, existing deep UIE methods usually adopt pseudo references or synthetic data for network training. The most commonly used pseudo references dataset is UIEB [24], which contains 890 paired real underwater images. All reference images of UIEB are generated by 12 UIE algorithms and scored by 50 volunteers to choose the best result as the pseudo label, which is not the actual ground truth. More importantly, the number of paired real underwater images of UIEB is very limited, resulting in poor generalization ability of the trained model [26], [28]. In comparison, synthetic data has accurate labels and sufficient samples, which is much easier to be obtained. Existing deep methods tend to exploit synthetic data to train the proposed networks, achieving relatively promising performance. Nevertheless, these methods ignore the domain

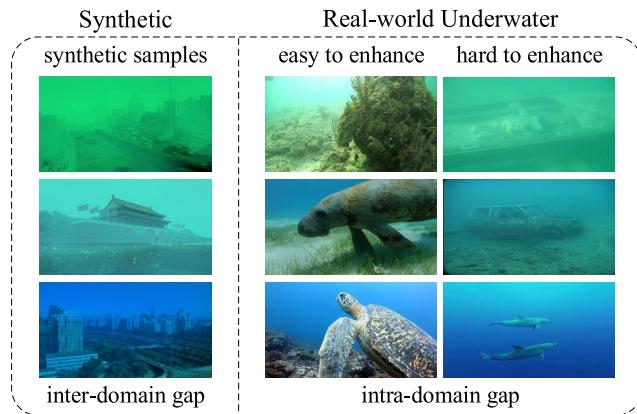


Fig. 1. Illustration of two challenges for underwater image enhancement. (1) Inter-domain gap challenge: the domain shift between the synthetic images and real images is often ignored, and thus the deep models trained on synthetic data often suffer a great performance drop in some real underwater images with different distortion distributions; (2) Intra-domain gap challenge: the complex and changeable underwater environment causes a large gap in the real-world data itself. Without considering it, deep models often produce visually unpleasing artifacts and color distortions on various real images.

shift problem from synthetic to real data, i.e., inter-domain gap, and thus they often suffer a severe performance drop when facing some real-world underwater images with different distortion distributions, as shown in Fig.1.

Apart from this, another challenging problem in underwater image enhancement is diversity of real image distributions. Generally, the quality of images captured in the water is severely affected by many factors such as illumination conditions, water bodies, water depth, seasonal and weather changes, etc [24], [28]. As presented in Fig.2 (a), these factors lead to various kinds of degradation and a large gap among real images itself, i.e., intra-domain gap. There have been rarely studies proposed to address the challenge of underwater real image itself distribution diversity. Four representative real examples and their corresponding results made by a deep model are shown in Fig.2 (b). The model shows satisfactory performance on some images (good results). However, it cannot perform well on some images (poor results), introducing obvious local artifacts, noises and over-enhancement, etc. Obviously, without considering the intra-domain gap, it is hard for a deep model to effectively handle real underwater images with such various degradation distributions.

Motivated by the above analysis, this paper proposes a novel Two-phase Underwater Domain Adaptation network (TUDA) for underwater image enhancement to jointly bridge the inter-domain gap and the intra-domain gap, which consists of two phases: inter-domain adaptation and intra-domain adaptation. Concretely, in the first phase, a new triple-alignment network is designed, including a translation part and a task-oriented enhancement part, one for the synthetic-to-real translation and another for image enhancement. Coupled with image-level, feature-level and output-level adaptations in an end-to-end manner, triple alignment parts can cooperate with each other for learning more domain-invariant representations to better minimize the inter-domain discrepancy.

In the second phase, a simple yet efficient rank-based underwater quality assessment algorithm (RUIQA) is proposed, which can better evaluate the perceptual quality of enhanced images by learning to rank. The proposed RUIQA is strongly sensitive to various artifacts and can be easily plugged in both the training and testing pipeline. Based on the assessed quality of enhanced images, we divide the real data into two categories: easy and hard samples, and get a trustworthy real image set with pseudo labels. Subsequently, using the easy-pseudo pairs and unpaired hard samples, an easy/hard domain adaptation technique is performed to close the intra-domain gap between easy and hard samples. The overview of our TUDA is shown in Fig.3. To the best of our knowledge, this is the first work that jointly explores the inter-domain and intra-domain adaptation in the underwater image enhancement community. The main contributions of this paper can be summarized as follows:

- 1) A novel two-phase underwater domain adaptation network is presented, termed as TUDA, to simultaneously reduce the inter-domain and intra-domain gap, which successfully sheds new light on future direction for enhancing underwater images.
- 2) A novel triple-alignment architecture is designed in the inter-domain adaptation phase, which effectively performs image-level, feature-level and output-level adaptations using jointly adversarial learning. Triple alignment parts can improve each other, and the combination of them can better build invariance across domains and thus bridging the inter-domain gap.
- 3) A rank-based underwater quality assessment method is developed in the intra-domain adaptation phase, which can effectively assess the perceptual quality of enhanced images with the help of learning to rank. From this method, we successfully perform an easy-hard classification and an easy/hard adaptation technique to significantly reduce the intra-domain gap.

## II. RELATED WORK

### A. Underwater Image Enhancement (UIE)

Recently, numerous UIE approaches have been developed and can be roughly categorized into three groups: image-based, physical-based and learning-based methods.

1) *Image-Based Methods*: [1], [2], [3], [4], [5], [6], [7], [8], [9], [10] mainly modify pixel values of underwater images to improve visual quality, including pixel value adjustment [1], [4], [5], [10], retinex decomposition [7], [8] and image fusion [6], [9], etc. For example, Zhang et al. [8] propose an extended multi-scale retinex-based underwater image enhancement method, in which the input image is processed by three steps: color correction, layer decomposition and enhancement. Ancuti et al. [9] propose a novel multi-scale fusion strategy, which blends a color-compensated and white-balanced version of the given image to generate a better result. Recently, based on the characteristics of severely non-uniform color spectrum distribution in underwater images, Ancuti et al. [10] propose a new color-channel-compensation pre-processing step in the opponent color channel to better

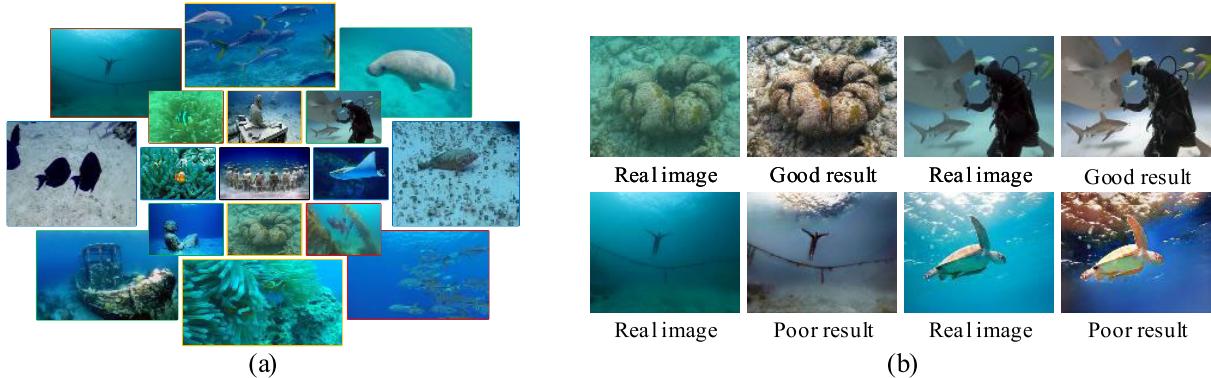


Fig. 2. (a) Examples of real-world underwater images, which have obvious different characteristics of underwater image quality degradation, e.g., color casts and blurred details. (b) Some results come from our inter-domain adaptation phase. Obviously, the results of some samples have higher perceptual quality, whereas the results of some samples suffer from local artifacts, noise, color casts and over-enhancement, etc.

overcome artifacts. These image-based methods can improve visual effects of degraded underwater images to some extent. However, due to ignoring the domain knowledge of underwater imaging, they are often unable to provide high quality results in some complex underwater scenarios.

Most **physical-based methods** [11], [12], [13], [14], [15], [16], [17], [18], [19] are based on the underwater image formation model [29], in which the background light and transmission map are estimated by some priors. The priors include underwater dark channel prior [12], minimum information prior [13], blurriness prior [14] and color-line prior [15], etc. For instance, built on underwater image blurriness and light absorption, Peng et al. [14] propose an underwater image restoration method combined with a blurriness prior to estimate more accurate scene depth. Inspired by the minimal information loss principal, Li et al. [13] estimate an optimal transmission map to restore underwater images, and exploit a histogram distribution prior to effectively improve the contrast and brightness. Berman et al. [15] incorporate the color-line prior and multiple spectral profiles information of different water types into the physical model, and employ the gray-world assumption theory to choose the best result, showing great performance on image dehazing. These methods restore underwater images well in some cases. However, when the priors are invalid, unstable and visually unpleasing artifacts still inevitably appear in some regions.

Recently, deep learning has made remarkable progresses in low-level vision problems, which benefits from the large data and powerful computational ability. However, for the problem of UIE, the raw underwater images are usually lack of the clean counterparts. There are many methods to improve performance by training their models on pseudo references data. As a pioneering work, Li et al. [24] construct the first underwater pseudo references dataset, termed as UIEB, including totally 890 underwater raw images and corresponding manually selected pseudo references. With these images, Li et al. [24] design a gate fusion network, where three confidence maps are learned to fuse three pre-processing versions into a decent result. Recently, Li et al. [26] develop an underwater image enhancement network in medium transmission-guided multi-color space for more robust enhancement. The pseudo references-based

methods can produce visually pleasing results. However, they cannot restore the color and structure of objects well and tend to generate inauthentic results since the reference images are not the actual ground truths. Moreover, pseudo reference datasets usually have monotonous content, limited scenes and insufficient data, which leads to the poor generalization performance of the learned model.

There are also many algorithms to train their networks using synthetic data since underwater image synthesis has accurate labels and sufficient data. As a pioneering work, combined with the knowledge of underwater imaging, Li et al. [21] design a generative adversarial network for generating realistic underwater-like images from in-air images and depth maps, and then utilize these generated data to correct color casts in a supervised manner. Fabbri et al. [22] directly employ a CycleGAN to generate paired training data, and then a fully convolutional encoder-decoder is trained to improve the underwater image quality. Recently, Li et al. [20] propose to synthesize ten types of underwater images based on an underwater image formation model and some scene parameters, and train an enhancement model for each type of water. Duhane et al. [30] improve the work of [20] by introducing the object blurriness and color shift components to synthesize more accurate underwater-like data.

Synthesis data can simulate different underwater types and degradation levels, and has the corresponding reference images as guidance for network training. However, due to the certain domain discrepancy between synthetic and real-world data, deep models trained on synthetic data often fail to generalize well on some real underwater scenarios.

### B. Domain Adaptation

Domain adaptation has been extensively explored recently, which aims to reduce the distribution gap between two different domains, and can be performed at the image level or feature level. To the best of our knowledge, domain adaptation is seldom systematically studied in underwater image enhancement field. However, it has a wide range of applications in other fields such as image hazing [31], semantic segmentation [32], [33] and depth prediction [34], [35], etc. For example, Shao et al. [31] propose a domain adaptation for

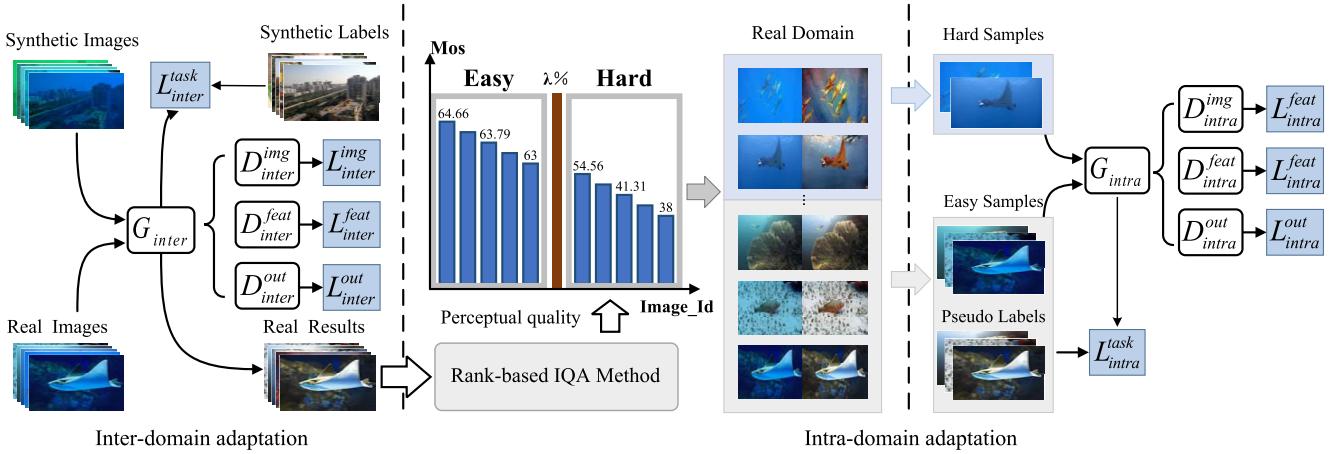


Fig. 3. Illustration of our proposed TUDA, which consists of two phases, inter-domain adaptation and intra-domain adaptation. In the inter-domain adaptation phase,  $G_{inter}$  can effectively reduce the distribution discrepancy between synthetic and real images using the image-level, feature-level and output-level discriminator  $D_{inter}^{img}$ ,  $D_{inter}^{feat}$  and  $D_{inter}^{out}$ . Details are introduced in Section III. In the intra-domain adaptation phase, a rank-based underwater image quality assessment method (IQA) is first presented to separate all real data into easy and hard samples, where  $\lambda$  is the ratio of real-world images assigned into the easy samples. Then, using the trustworthy easy set with generated precise pseudo labels, we can powerfully close the intra-domain gap between easy and hard samples with the help of  $G_{intra}$ ,  $D_{intra}^{img}$ ,  $D_{intra}^{feat}$  and  $D_{intra}^{out}$ .

single image dehazing based on CycleGAN, in which a new bidirectional translation network is designed to reduce the gap between synthetic and real images by jointly synthetic-to-real and real-to-synthetic image-level adaptations. Zhao et al. [35] propose a novel geometry-aware symmetric domain adaptation framework to explore the labels in the synthetic data and epipolar geometry in the real data jointly for better bridging the gap between synthetic and real domains, and thus generating high-quality depth maps.

More recently, Pan et al. [32] propose an unsupervised intra-domain adaptation through self-supervision for semantic segmentation. To obtain extra performance gains, the authors first train the model using the inter-domain adaptation from existing approaches, and decompose the target domain in two small subdomains based on the mean value of entropy maps from the predicted segmentation maps. Then an alignment on entropy maps for both subdomains is conducted to further reduce the intra-domain gap. Inspired by this work, the concepts of inter- and intra-domain are introduced to underwater image enhancement. In this paper, we propose a different domain adaptation method, in which a new triple-alignment network used for inter-domain adaptation and a novel underwater image quality assessment algorithm used for intra-domain adaptation are proposed. The detailed architectures of the proposed method are introduced in the following sections.

### III. PROPOSED METHOD

Given a set of synthetic images  $X_S = \{x_s, y_s\}$  and a real underwater image set  $X_R = \{x_r\}$ , we aim to reduce the inter-domain gap between the synthetic and real data and the intra-domain gap among the real data itself. A novel two-phase underwater domain adaptation architecture is proposed, which consists of two parts: inter-domain adaptation and intra-domain adaptation. As shown in Fig.3, in the inter-domain phase, a new triple-alignment network  $G_{inter}$  is developed

to jointly perform image-level, feature-level and output-level alignment, including an image translation part  $G_{inter}^T$  and an image enhancement part  $G_{inter}^E$ . The former is used for learning a more robust transformation from synthetic to real-world underwater images, and the latter is used for performing image enhancement using both translated and real images. Details are introduced in Section III-A. From this adaptation, a rank-based underwater quality assessment method (i.e., RUIQIA) is designed to evaluate the perceptual quality of the enhanced images. Based on these predicted quality scores, we separate the real underwater raw images into easy and hard samples ( $X_E = \{x_e, y_e\}$  and  $X_H = \{x_h\}$ ), and then conduct the intra-domain adaptation similar to inter-domain adaptation. Details of this phase are described in Section III-B.

#### A. 1<sup>st</sup> phase: Inter-Domain Adaptation

The proposed triple-alignment network aims to reduce the inter-domain adaptation gap between the synthetic and real data domain in the image level, feature level and output level, as shown in Fig.4. The proposed network is composed of two parts: an image translation module  $G_{inter}^T$  for enhancing realism of input images, followed by an enhancement module  $G_{inter}^E$ .  $G_{inter}^T$  takes synthetic samples and their corresponding ground truth labels ( $x_s, y_s$ ) as inputs, and generates translated images  $x_{st}$ , i.e.,  $x_{st} = G_{inter}^T(x_s)$ . The translated images  $x_{st}$  are expected as possible with similar distribution of real images  $x_r$ . Meanwhile, the discriminator  $D_{inter}^{img}$  is encouraged to identify the difference between  $x_{st}$  and  $x_r$ . To stabilize the gradients and improve performance, the WGAN-GP adversarial loss [36] is adopted to perform image-level alignment, set as:

$$L_{inter}^{img} = \mathbb{E}_{x_{st}} \left[ D_{inter}^{img}(x_{st}) \right] - \mathbb{E}_{x_r} \left[ D_{inter}^{img}(x_r) \right] + \lambda_{img} \mathbb{E}_{\hat{I}} \left( \left\| \nabla_{\hat{I}} D_{inter}^{img}(\hat{I}) \right\|_2 - 1 \right)^2 \quad (1)$$

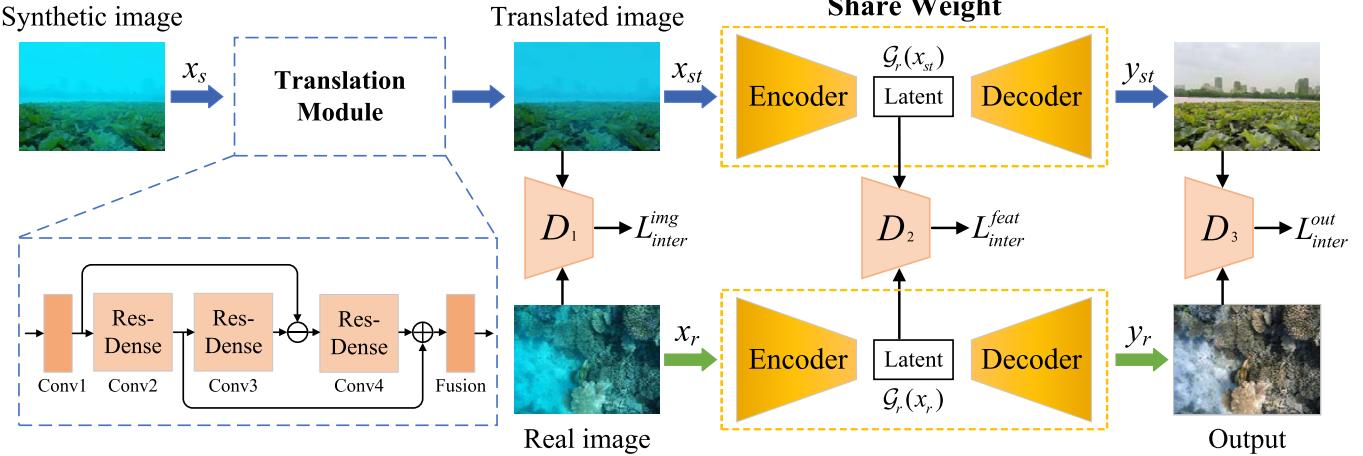


Fig. 4. Illustration of our triple-alignment network proposed in the inter-domain adaptation phase, trained on synthetic underwater image pairs and unpaired real images, which consists of an image translation part for enhancing realism of input images, followed by an image enhancement part. They are cooperatively performed image-level, feature-level and output-level alignments and trained end-to-end in an adversarial learning manner.

where  $\hat{I}$  represents the sampling distribution which is sampled uniformly from  $x_r$  and  $x_{st}$ , and  $\lambda_{img}$  is the penalty parameter, in our works,  $\lambda_{img} = 10$ .

Color cast is one of the main characteristics of underwater images, which generally can be divided into three tones: blue, green and blue-green [28]. Inspired by this, the synthetic and real images are divided into three color tone subsets according to the average value of the blue (b) channel in the CIElab color space. When the synthetic images and the real images are in the same color tone, the synthetic-to-realistic translation can be accomplished, which greatly speeds up the convergence of the model. In addition, intuitively, the gap between the synthetic and real data mainly comes from low-level differences, such as color and texture. Thus, the translated images  $x_{st}$  should be retained with the same semantic content as  $x_s$ , but with a different appearance. Thus, a semantic content loss component is incorporated along with the adversarial loss, set as:

$$L_{inter}^{cont} = w_k \sum_{k \in L_c} \|\phi_k(x_s) - \phi_k(x_{st})\|_1 \quad (2)$$

where  $\phi_k(\cdot)$  is the  $k$ th-layer feature extractor of the VGG-19 network pretrained on ImageNet,  $L_c$  is the set of layers, including conv1-1, conv2-1, conv3-1, conv4-1 and conv5-1.  $w_k$  denotes the weight of the  $k$ th-layer, set as  $\frac{1}{32}, \frac{1}{16}, \frac{1}{8}, \frac{1}{4}, 1.0$  in our experiments.

After the synthetic images  $x_s$  are translated, the generated realistic images  $x_{st}$  can be obtained. The paired translated data  $(x_{st}, y_s)$  is utilized to train the enhancement network  $G_{inter}^E$ .  $G_{inter}^E$  is trained in a supervised way, including a content loss and a perceptual loss, set as:

$$L_{inter}^{task} = a * \|y_s - y_{st}\|_1 + b * \sum_{k \in L_c} \|\phi_k(y_s) - \phi_k(y_{st})\|_1 \quad (3)$$

where  $y_{st}$  is the output of the enhancement network  $G_{inter}^E$ , i.e.,  $y_{st} = G_{inter}^E(x_{st})$ . The two parameters  $a$  and  $b$  are the weights of different loss components, set as 0.8 and 0.2, respectively.

To better minimize the inter-domain gap, a feature-level adversarial loss is also introduced into the enhancement part,

set as:

$$\begin{aligned} L_{inter}^{feat} = & \mathbb{E}_{x_{st}} [D_{inter}^{feat}(\mathcal{G}_r(x_{st}))] - \mathbb{E}_{x_r} [D_{inter}^{feat}(\mathcal{G}_r(x_r))] \\ & + \lambda_{feat} \mathbb{E}_{\hat{I}} (\|\nabla_{\hat{I}} D_{inter}^{feat}(\hat{I})\|_2 - 1)^2 \end{aligned} \quad (4)$$

where  $G_{inter}^E$  shares identical weights in both real and translated input pipelines and  $\mathcal{G}_r$  is the encoder of  $G_{inter}^E$ .  $\hat{I}$  denotes the sampling distribution sampled uniformly from  $\mathcal{G}_r(x_{st})$  and  $\mathcal{G}_r(x_r)$ .  $\lambda_{feat}$  is the penalty parameter, set as 10 in our experiments.

For the enhancement result  $y_r$  of the real image, its distribution should be consistent with the enhancement result  $y_{st}$  of the synthetic image as much as possible. Therefore, a output-level adversarial loss is further introduced to improve the effect of inter-domain adaptation, set as:

$$\begin{aligned} L_{inter}^{out} = & \mathbb{E}_{y_{st}} [D_{inter}^{out}(y_{st})] - \mathbb{E}_{y_r} [D_{inter}^{out}(y_r)] \\ & + \lambda_{out} \mathbb{E}_{\hat{I}} (\|\nabla_{\hat{I}} D_{inter}^{out}(\hat{I})\|_2 - 1)^2 \end{aligned} \quad (5)$$

in which  $\hat{I}$  denotes the sampling distribution which is sampled uniformly from  $y_r$  and  $y_{st}$ , and  $\lambda_{out}$  is the penalty parameter, which is set as 10, i.e.,  $\lambda_{out} = 10$ .

With all image-level, feature-level and output-level alignments in an end-to-end manner, the proposed triple-alignment network can better build invariance between domains and thus closing the inter-domain gap. The overall loss function for the inter-domain adaptation phase is expressed as follows:

$$\begin{aligned} L_{inter} = & \lambda_1 L_{inter}^{img} + \lambda_2 L_{inter}^{cont} + \lambda_3 L_{inter}^{task} \\ & + \lambda_4 L_{inter}^{feat} + \lambda_5 L_{inter}^{out} \end{aligned} \quad (6)$$

where  $\lambda_1, \lambda_2, \lambda_3, \lambda_4$  and  $\lambda_5$  are trade-off weights, set as 1, 100, 10, 0.0005 and 0.0005, respectively.

## B. 2<sup>nd</sup> phase: Intra-Domain Adaptation

As mentioned above, the intra-domain gap exists among real underwater images itself, so a straightforward method

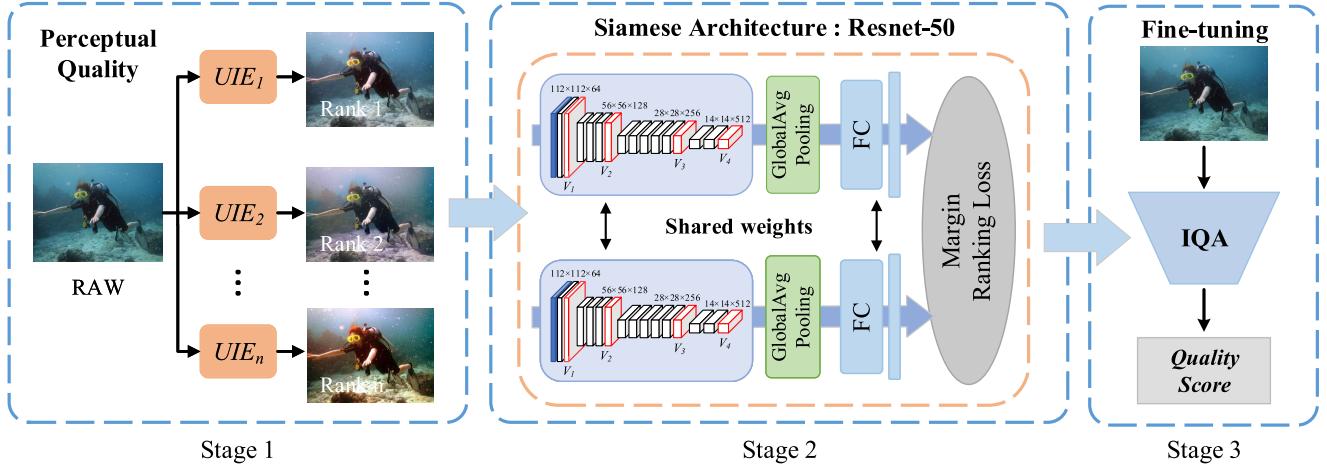


Fig. 5. The proposed RUIQA consists of three stages, namely stage 1: build a real-world underwater ranking dataset based on their perceptual quality; stage 2: train the Siamese architecture ResNet-50 using the ranking dataset; stage 3: perform a fine-tuning technique to predict the image quality score.



Fig. 6. Illustration of easy and hard samples. Their results come from the same inter-domain adaptation network. Obviously, the results of easy samples have higher perceptual quality, whereas the results of hard samples suffer from local artifacts, noise, color casts and over-enhancement, etc.

is the divide-and-conquer strategy. Some images containing a similar distribution with the training data are easy to be enhanced, called easy samples, and vice versa. Therefore, real underwater images can be separated into easy samples and hard samples according to the assessed quality of enhanced images. Enhanced results of easy samples are trustworthy, which can be used as pseudo-labels. By using easy samples and their corresponding pseudo-labels, an unsupervised way is conducted to learn easy/hard adaptation to close the intra-domain gap between easy and hard samples. To reasonably separate real underwater into easy and hard parts based on the quality of enhanced images, an effective method is required. One may attempt to use existing underwater image quality assessment methods for separating, such as UCIQE [37] and UIQM [38]. However, the experimental results in [24] and [26] show that these methods cannot accurately evaluate image quality in some cases. Notably, this paper presents a novel and effective underwater quality assessment method with the help of rank information learned from rankings, which can effectively assess the quality of enhanced images, named rank-based underwater image quality assessment (RUIQA).

*1) Rank-Based Underwater Image Quality Assessment (RUIQA):* Existing deep IQA methods usually initialize

their model parameters using the pre-trained models on the ImageNet dataset [39], [40]. Although these metrics achieve good results on ground images to some extent, the performance is still unsatisfactory when facing images with various underwater distortions. In our opinion, this is mainly caused by the fact that pre-trained models capture information that is conducive to ground image processing instead of the unique prior information of underwater images, and thus they cannot easily adapt the characteristics of underwater image quality assessment tasks.

Inspired by [41] in image super-resolution, this paper utilizes an underwater ranking dataset to train a large network to extract some ranking information by learning to rank, which is closely related to the perceptual quality. Then we fine-tune it to more accurately predict the perceptual quality of enhanced images. Differently, in [41], a Ranker is trained to learn the behavior of perceptual metrics and then a novel rank-content loss is introduced to optimize the perceptual quality, while our method trains an underwater ranker and makes it as model initialization parameters to help assess perceptual quality. As shown in Fig.5, RUIQA consists of three stages: generating rank images, training ranker and fine-tuning network.

2) *Generating Rank Images*: A large number of underwater images are first collected from online and some public datasets [24], [28], [42], and then carefully selected and refined. Most of the collected images are weeded out, and about 3900 candidate images are remained. We randomly choose 800 pictures to construct an underwater ranking dataset. With the candidate underwater images, the enhanced images are generated by 8 image enhancement methods, including Unet, Fusion-12 [6], Fusion-18 [9], Two-step-based [43], Histogram prior [13], Blurriness-based [14], WaterNet [24], FUIE-GAN [42] and a commercial application for enhancing underwater images (i.e., dive+). Each enhanced image is assessed with a continuous quality scale, ranging from 1 to 5. After then, the quality scale is mapped to a continuous score between 1 to 100. 20 volunteers are invited to conduct the evaluation in the same monitor environment. Following the work of [44], the raw scores are refined by means of some standardized settings [45], [46] and the Mean Opinion Score (MOS) is calculated [47], [48], obtaining reliable subjective rating results. Our dataset and code will be publicly released on <https://github.com/UnderwaterLab/TUDA>.

3) *Training Ranker*: With the obtained MOS values, the pair-wise images and the corresponding ranking order labels can be obtained. Meanwhile, ResNet-50 [49] is employed as the Siamese network architecture to extract ranking information. The Siamese network is trained by a margin-ranking loss proposed in [41], which is beneficial for the model to learn the ranking information. After training, a single branch of the Siamese network, i.e., the pre-trained ResNet-50 model parameters on the ranking images, is extracted to initialize our backbone network.

4) *Fine-Tuning Network*: In our RUIQA, the last global average pooling (GAP) and fully connected (FC) layer of the pre-trained ResNet-50 model are removed. To better handle distortion diversity, multi-scale features extracted from four layers (conv2-10, conv3-12, conv4-18 and the last layer) are treated as the input of four blocks. The block is composed of a  $1 \times 1$  convolution, a GAP and a FC layer, mapping the multi-scale features into the corresponding perceptual quality vectors. Finally, these predicted quality vectors are regressed into a quality score. In the training phase, the network is fine-tuned by minimizing the  $L_1$  loss between the predicted score and the MOS value label.

Using the proposed RUIQA, the quality score of each enhanced image is predicted. The higher the value, the model is more confident with the real-world image (i.e., easy sample). This step can be named as an easy-hard classification. Some classification results are shown in Fig. 6, it can be observed that the enhanced results of easy samples have higher perceptual quality and are near to the human perception. In practice,  $\lambda$  is introduced as ratio to help the separation, which means the ratio of easy samples to total samples. The corresponding MOS value of the specified  $\lambda$  is set as a threshold to pick up easy samples and the rest images are considered as hard samples for the training. In Section IV-D.4, how to obtain the best  $\lambda$  is explored. It is very important for the intra-domain training and finally TUDA testing pipeline.

5) *Easy/Hard Adaptation*: For easy samples  $x_e$ , the enhanced results  $y_e$  are set as pseudo labels to obtain some real underwater pair data  $(x_e, y_e)$ . By using the pair data  $(x_e, y_e)$ , we aim to adopt an easy/hard adaptation technique to close the intra-domain gap between easy and hard samples, which is composed of an intra-domain translation part  $G_{intra}^T$  and an intra-domain enhancement part  $G_{intra}^E$ .  $G_{intra}^T$  tries to translate the easy sample  $x_e$  to be indistinguishable from the hard images  $x_h$ . Meanwhile, a discriminator  $D_{intra}^{img}$  aims to differentiate between the translated image  $x_{et}$  and hard images  $x_h$ . This minimax game can be modeled using an adversarial loss as follows:

$$L_{intra}^{img} = \mathbb{E}_{x_{et}} \left[ D_{intra}^{img}(x_{et}) \right] - \mathbb{E}_{x_h} \left[ D_{intra}^{img}(x_h) \right] + \lambda_{img} \mathbb{E}_{\hat{I}} \left( \left\| \nabla_{\hat{I}} D_{intra}^{img}(\hat{I}) \right\|_2 - 1 \right)^2 \quad (7)$$

where the parameter  $\lambda_{img} = 10$ ,  $\hat{I}$  represents the sampling distribution which is sampled uniformly from  $x_h$  and  $x_{et}$ .

Similar to  $G_{inter}^T$ , an excellent translation  $G_{intra}^T$  should keep the translated image  $x_{et}$  “similar” in content to the original easy image  $x_e$ . Thus, semantic content loss is incorporated to better achieve content preservation, set as:

$$L_{intra}^{cont} = w_k \sum_{k \in L_c} \|\phi_k(x_e) - \phi_k(x_{et})\|_1 \quad (8)$$

where  $L_c$  is the set of layers (conv1-1, conv2-1, conv3-1, conv4-1 and conv5-1) and  $\phi_k(\cdot)$  is the corresponding  $k$ th-layer feature map in pre-trained VGG-19 model.  $w_k$  denotes the weight of the  $k$ th-layer, in our work, set as  $\frac{1}{32}, \frac{1}{16}, \frac{1}{8}, \frac{1}{4}, 1.0$  respectively.

Then, the translated image  $x_{et}$  is input to the intra-domain enhancement part  $G_{intra}^E$ , and the enhanced image  $y_{et}$  is obtained.  $G_{intra}^E$  is trained in a supervised manner, including a content loss and a perceptual loss, set as:

$$L_{intra}^{task} = c * \|y_e - y_{et}\|_1 + d * \sum_{k \in L_c} \|\phi_k(y_e) - \phi_k(y_{et})\|_1 \quad (9)$$

where  $c$  and  $d$  are trade-off weights, set as 0.8 and 0.2 respectively.

To better minimize the intra-domain gap between easy and hard samples in the real domain, a feature-level adaptation is also performed, where a discriminator  $D_{intra}^{feat}$  is introduced to align the distributions between the feature map of  $x_{et}$  and  $x_h$ . The loss is defined as:

$$L_{intra}^{feat} = \mathbb{E}_{x_{et}} \left[ D_{intra}^{feat}(\mathcal{G}_h(x_{et})) \right] - \mathbb{E}_{x_h} \left[ D_{intra}^{feat}(\mathcal{G}_h(x_h)) \right] + \lambda_{feat} \mathbb{E}_{\hat{I}} \left( \left\| \nabla_{\hat{I}} D_{intra}^{feat}(\hat{I}) \right\|_2 - 1 \right)^2 \quad (10)$$

where  $G_{intra}^E$  shares the same weight in both translated input and hard images pipelines and  $\mathcal{G}_h$  is the encoder of  $G_{intra}^E$ .  $\hat{I}$  denotes the sampling distribution sampled uniformly from  $\mathcal{G}_h(x_{et})$  and  $\mathcal{G}_h(x_h)$ .  $\lambda_{feat}$  is the penalty factor, set as 10 in this work. At the same time, a output-level discriminator between the enhancement results of the easy sample and the hard

sample is introduced to further improve the effect of intra-domain adaptation, defined as:

$$L_{intra}^{out} = \mathbb{E}_{x_{et}} [D_{intra}^{out}(x_{et})] - \mathbb{E}_{x_h} [D_{intra}^{out}(x_h)] + \lambda_{out} \mathbb{E}_{\hat{I}} \left[ \left( \left\| \nabla_{\hat{I}} D_{intra}^{out}(\hat{I}) \right\|_2 - 1 \right)^2 \right] \quad (11)$$

where the parameter  $\hat{I}$  refers to the sampling distribution that is sampled uniformly from  $y_h$  and  $y_{et}$ , and  $\lambda_{img} = 10$ .

In summary,  $G_{intra}^T$  and  $G_{intra}^E$  are trained in an end-to-end manner, and thus the full loss function is as follow:

$$L_{intra} = \lambda_a L_{intra}^{img} + \lambda_b L_{intra}^{cont} + \lambda_c L_{intra}^{task} + \lambda_d L_{intra}^{feat} + \lambda_e L_{intra}^{out} \quad (12)$$

in which  $\lambda_a, \lambda_b, \lambda_c, \lambda_d$  and  $\lambda_e$  are trade-off weights, set as 1, 100, 10, 0.0005 and 0.0005, respectively.

### C. Architecture Details

The detailed architecture of two transform modules ( $G_{inter}^T, G_{intra}^T$ ) is shown in Fig.7, in which the down-sampling layer is not employed in the translator for avoiding valuable information loss. For image-level discriminators ( $D_{inter}^{img}, D_{intra}^{img}$ ), feature-level discriminators ( $D_{inter}^{feat}, D_{intra}^{feat}$ ) and output-level discriminators ( $D_{inter}^{out}, D_{intra}^{out}$ ), PatchGANs [50] is employed, which can better locally discriminate whether image patches are real or fake. A simple network architecture (stack the dense block under the U-Net structure)<sup>1</sup> is used as our enhancement parts ( $G_{inter}^E, G_{intra}^E$ ). It's worth mentioning that our test pipeline only needs the enhancement parts ( $G_{inter}^E, G_{intra}^E$ ) and the proposed rank-based IQA method, as shown in Fig.8.

## IV. EXPERIMENTS

In this section, the implementation details and experiment settings are first described. Then, quantitative and visual quality analyses are carried out to evaluate the effectiveness of the proposed TUDA. Finally, a series of ablation studies are provided to verify the advantages of each component, and the model complexity and running time are analyzed.

### A. Implementation Details

For training, a synthetic underwater image dataset is generated following the physical model proposed in the project page of ANA-SYN. The synthetic dataset contains 9 water types (Type I, II, III, IA and IB for open ocean water and type 1C, 3C, 5C and 7C for coastal water) defined in [29], and each type has 1000 images which are randomly chosen from RTTS dataset [51]. The constructed dataset is divided into two parts, 7200 ( $800 \times 9$ ) images for training, denoted as **Train-S7200** and 1800 ( $200 \times 9$ ) images for testing, denoted as **Test-S1800**. For real underwater images, as mentioned above, a large real-world underwater database including 3900 images is proposed. The database is divided into two parts, 2900 images for training, denoted as **Train-R2900** and 1000 images for testing, denoted as **Test-R1000**. All images are resized to

$256 \times 256$  and the pixel values are normalized to  $[-1, 1]$ . Several data augmentation techniques are performed in the training phase including random rotating  $90^\circ, 180^\circ, 270^\circ$  and horizontal flipping.

The proposed TUDA and RUIQA are implemented in Pytorch framework and all experiments are carried out on two NVIDIA Titan V GPUs. Adam optimizer with a learning rate of  $1 \times 10^{-4}$  is utilized to train all generators. For all discriminators, we adopt an Adam optimizer with learning rate of  $2 \times 10^{-4}$  as the optimization method. Default values of  $\beta_1$  and  $\beta_2$  are set as 0.5 and 0.999, respectively. The batch size is set to 4. Models are trained for 200 epochs, and their learning rates decay linearly to zero in the next 100 epochs.

### B. Experiment Settings

For testing, we conduct comprehensive experiments on five publicly available real-world underwater image benchmarks: **SQUID** [15], **UIEB** [24], **UCCS** [28], **EUVP** [42] and **UFO-120** [52]. The compared UIE algorithms include **UIEM** [18], **UNTV** [17], **UGAN** [22], **UWCNN** [20], **FUIEGAN** [42], **LCNet** [27], **Water-Net** [24] and **Ucolor** [26]. The first two algorithms are traditional methods, while the remaining are deep-learning methods. Note that for all the above-mentioned methods, we utilize the publicly released test models and parameters to produce their results.

For results on real image benchmarks, performances are measured by three no-reference underwater quality assessment metrics: UCIQE, UIQM and the proposed RUIQA. For the three metrics, a higher score denotes a better human visual perception. It should be pointed out that UCIQE and UIQM are not sufficient to reflect the performance of various underwater image enhancement methods in some cases [24], [26]. In our study, we only present the scores of UCIQE and UIQM as the reference for the following research.

To more accurately evaluate the visual quality of the results, a user study is also conducted, in which 50 images are randomly selected from each testing dataset to be scored. 15 volunteers are invited in this evaluation, and the scoring range is 0 to 5 levels, referring Bad, Poor, Fair, Good and Excellent. We also calculate the average angular reproduction error [15] on the 16 representative examples presented in the project page<sup>2</sup> of SQUID (denoted as SQUID-16) to evaluate the color restoration accuracy. The SQUID-16 dataset contains four dive sites (Katzaa, Michmoret, Nachsholim, Satil, denoted as Set A, Set B, Set C and Set D, respectively), four representative samples are selected from each dive site. The smaller color error, the better color correction performance.

### C. Comparison With State-of-the-Art UIE Methods

In this subsection, we conduct quantitative and visual comparisons on diverse testing datasets to evaluate the effectiveness of the proposed method. Moreover, the robustness and accuracy of color restoration are analyzed. Due to the limited space, more results are given in the supplementary material.

<sup>1</sup><https://github.com/Underwater-Lab-SHU/ANA-SYN>

<sup>2</sup>[http://csms.haifa.ac.il/profiles/treibitz/datasets/ambient\\_forwardlooking/index.html](http://csms.haifa.ac.il/profiles/treibitz/datasets/ambient_forwardlooking/index.html)

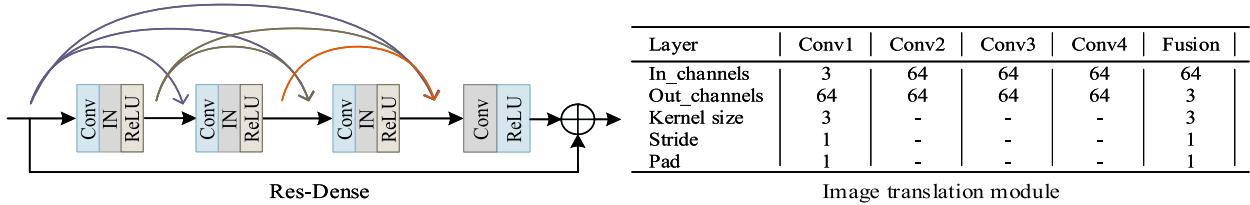


Fig. 7. Configurations of image translation module. “Conv1” is combined by a convolutional layer, a IN layer and a ReLU activation function. “Conv2”, “Conv3”, “Conv4” denote Res-Dense block. “Fusion” is combined by a convolution layer and a Tanh activation function.

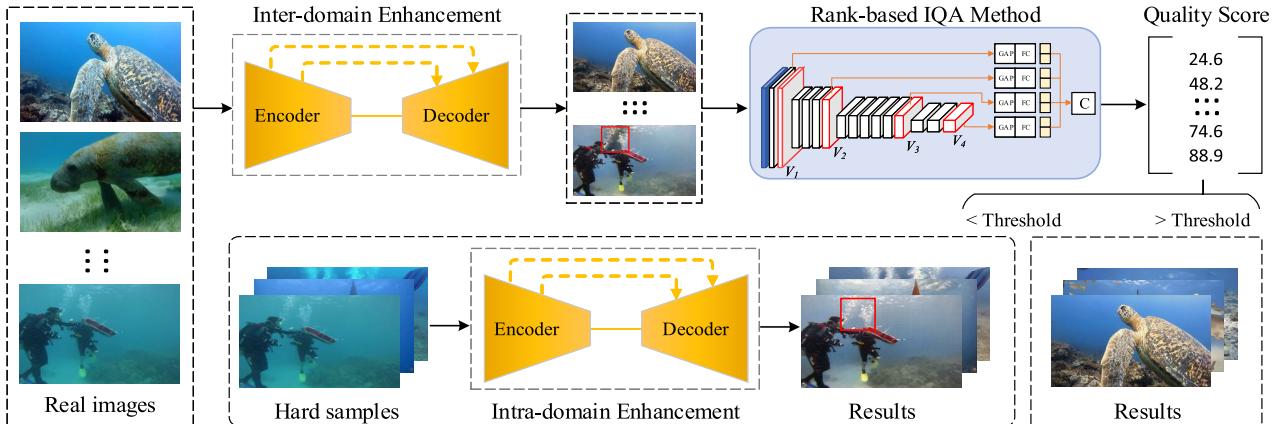


Fig. 8. An overview of our testing pipeline. The inter-domain enhancement part first takes real underwater images as input and outputs the corresponding inter-domain enhancement results. Then, our proposed rank-based IQA method evaluates the perceived quality of the enhancement result. When the score is less than the threshold, the corresponding raw image is regarded as a hard sample, and intra-domain enhancement is performed. When the score is greater than the threshold, the result is trustworthy and output directly.

TABLE I  
QUANTITATIVE RESULTS (AVERAGE UCIQE/UIQM) OF DIFFERENT METHODS ON SIX REAL BENCHMARKS (TEST-R1000, EUVP, UIEB, UCCS, UFO-120 AND SQUID). THE TOP THREE RESULTS ARE MARKED IN RED, BLUE AND GREEN

Methods	UCIQE ↑						UIQM ↑					
	Test-R1000	EUVP	UIEB	UCCS	UFO-120	SQUID	Test-R1000	EUVP	UIEB	UCCS	UFO-120	SQUID
UIEM [18]	0.606	0.614	0.613	0.535	0.661	0.504	1.932	1.825	2.198	2.528	1.768	0.868
UNTV [17]	0.623	0.628	0.627	0.603	0.637	0.590	1.787	1.242	1.892	2.573	1.616	1.717
UGAN [22]	0.611	0.604	0.621	0.573	0.624	0.584	3.127	3.237	3.180	3.213	3.152	2.780
UWCNN [20]	0.493	0.485	0.493	0.458	0.512	0.446	2.646	2.954	2.736	2.819	2.783	2.112
FUIE-GAN [43]	0.548	0.556	0.563	0.502	0.582	0.488	2.763	3.079	2.981	3.086	2.954	1.791
LCNet [27]	0.558	0.569	0.572	0.492	0.604	0.505	2.567	2.623	2.792	2.979	2.477	2.457
Water-Net [24]	0.553	0.550	0.564	0.536	0.568	0.510	2.741	2.992	2.953	3.126	2.828	2.031
Ucolor [26]	0.560	0.540	0.580	0.524	0.579	0.511	2.881	3.171	3.075	3.074	3.000	2.212
Our TUDA	0.588	0.585	0.593	0.560	0.593	0.582	3.077	3.287	3.183	3.175	3.135	2.601

TABLE II  
QUANTITATIVE RESULTS (AVERAGE RUIQA/PERCEPTUAL SCORES) OF DIFFERENT METHODS ON SIX REAL BENCHMARKS (TEST-R1000, EUVP, UIEB, UCCS, UFO-120 AND SQUID). THE TOP THREE RESULTS ARE MARKED IN RED, BLUE AND GREEN

Methods	RUIQA ↑						Perceptual Scores ↑					
	Test-R1000	EUVP	UIEB	UCCS	UFO-120	SQUID	Test-R1000	EUVP	UIEB	UCCS	UFO-120	SQUID
UIEM [18]	42.800	41.523	43.736	39.875	39.598	40.969	2.414	2.500	2.657	2.514	2.438	2.219
UNTV [17]	47.567	45.485	48.956	50.493	47.812	42.501	2.014	1.471	1.990	2.028	1.828	1.442
UGAN [22]	47.006	47.412	48.614	47.572	48.218	46.461	2.042	2.323	2.128	1.919	2.095	1.823
UWCNN [20]	41.539	41.175	38.255	39.922	41.088	43.804	2.542	2.385	2.238	2.557	2.542	2.390
FUIE-GAN [43]	50.750	49.150	52.363	47.947	52.871	46.060	2.942	2.957	3.309	2.980	3.214	2.600
LCNet [27]	49.669	48.480	51.631	47.790	51.194	49.800	2.690	2.909	3.052	2.471	2.909	2.485
Water-Net [24]	52.200	50.667	53.490	51.176	52.720	46.818	3.033	2.833	3.185	3.119	2.909	2.728
Ucolor [26]	53.756	52.130	55.478	49.105	53.890	47.636	3.157	3.204	3.500	3.004	3.219	2.971
Our TUDA	63.674	62.761	64.021	61.721	63.681	64.145	4.033	4.061	4.066	3.947	4.000	4.123

1) *Quantitative Comparisons:* The quantitative results of different methods on real challenging sets are reported in Table I and Table II. As presented, UNTV achieves the highest scores in terms of UCIQE, while the proposed method ranks

the third or fourth best on all challenging set. For the UIQM scores, the proposed method achieves the best performance on the EUVP and UIEB testing set. On the Test-R1000, UCCS, UFO-120 and SQUID, our method is only inferior to



Fig. 9. Visual comparisons on challenging underwater images sampled from **Test-R1000**. From left to right are raw underwater images, and the results of UIEM [18], UNTV [17], UGAN [22], UWCNN [20], FUIE-GAN [42], LCNet [27], Water-Net [24], Ucolor [26] and our proposed TUDA.

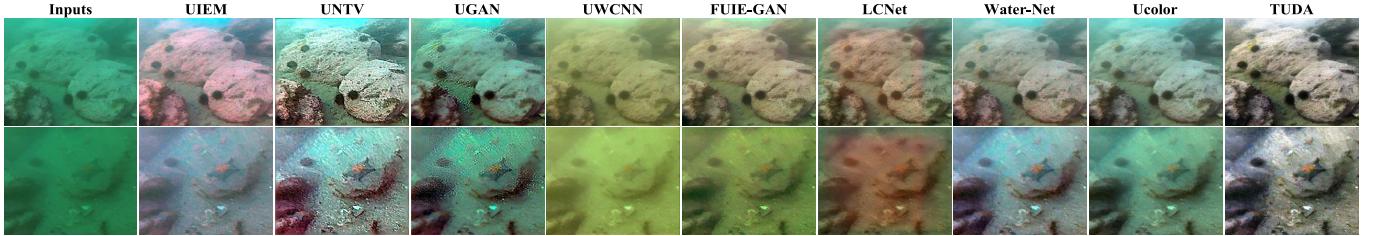


Fig. 10. Visual comparisons on challenging underwater images sampled from **UCCS**. From left to right are raw underwater images, and the results of UIEM [18], UNTV [17], UGAN [22], UWCNN [20], FUIE-GAN [42], LCNet [27], Water-Net [24], Ucolor [26] and our proposed TUDA.

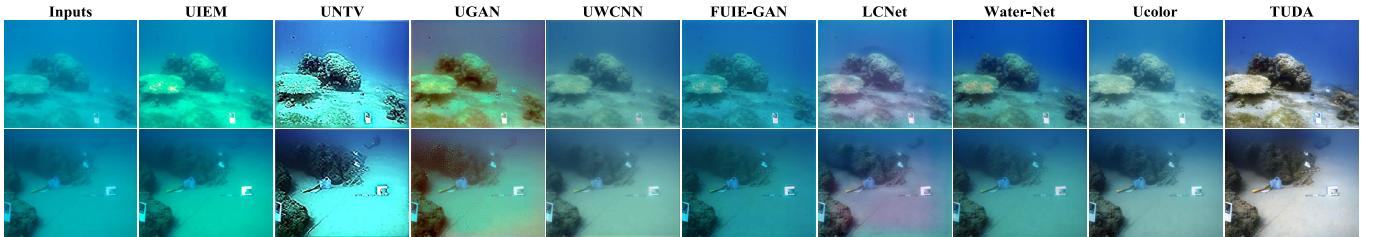


Fig. 11. Visual comparisons on challenging underwater images sampled from **SQUID**. From left to right are raw underwater images, and the results of UIEM [18], UNTV [17], UGAN [22], UWCNN [20], FUIE-GAN [42], LCNet [27], Water-Net [24], Ucolor [26] and our proposed TUDA.

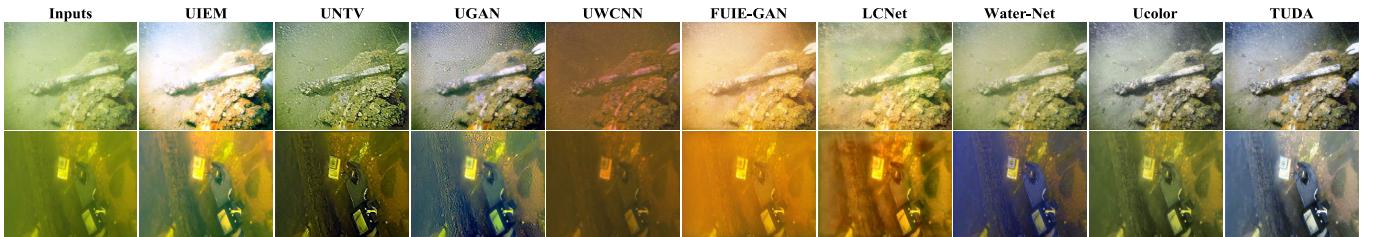


Fig. 12. Visual comparisons on challenging underwater images sampled from **UIEB**. From left to right are raw underwater images, and the results of UIEM [18], UNTV [17], UGAN [22], UWCNN [20], FUIE-GAN [42], LCNet [27], Water-Net [24], Ucolor [26] and our proposed TUDA.



Fig. 13. Visual comparisons on challenging underwater images sampled from **EUVP**. From left to right are raw underwater images, and the results of UIEM [18], UNTV [17], UGAN [22], UWCNN [20], FUIE-GAN [42], LCNet [27], Water-Net [24], Ucolor [26] and our proposed TUDA.

UGAN, ranking second. Observing the values of RUIQA and perceptual metrics, the deep methods achieve relatively higher scores. Among them, our method is obviously superior to the

other competing methods. Water-Net and Ucolor which trained based on pseudo references data perform relatively well, but they cannot restore green or some excessively distorted

TABLE III

THE AVERAGE ANGULAR REPRODUCTION ERROR (AE) ON SQUID-16.  
THE TOP THIRD RESULTS ARE MARKED IN RED, BLUE AND GREEN

Method	Angular Reproduction Error (AE)				
	Set A	Set B	Set C	Set D	Avg
UIEM [18]	32.531	32.027	33.405	35.280	33.310
UNTV [17]	19.943	26.304	20.675	22.541	22.366
UGAN [22]	10.275	11.875	9.486	6.633	9.567
UWCNN [20]	12.042	13.061	15.641	18.229	14.744
FUIE-GAN [43]	24.849	25.804	26.445	30.661	26.940
LCNet [27]	6.884	11.620	12.151	17.860	12.130
Water-Net [24]	21.112	21.547	19.703	23.198	21.390
Ucolor [26]	22.543	18.109	14.731	18.110	18.370
Our TUDA	10.750	8.829	9.389	6.491	8.865

images due to ignoring the intra-domain gap among real underwater images itself, and thus their performance is limited. UGAN trains the model using the synthetic data generated by Cycle-GAN. Since the inter-domain gap is not effectively reduced, the results often contain various artifacts, and thus the subjective score is relatively low.

There is an interesting finding from the quantitative results. UNTV and UGAN almost obtain the highest UCIQE and UIQM scores on all real datasets, respectively. However, their perceptual scores are relatively low, which means that they have poor subjective quality. The main reason behind this observation is that these metrics pay too much attention to some characteristics (not entire image) and do not consider the color shift and artifacts, and thus they are not sensitive to artifacts generated on the boundary of objects. They are also inconsistent with human visual perception in some cases, especially when the enhanced image is under-/over-enhanced (please refer to Fig.16) [24], [26].

2) *Visual Comparisons*: Fig.9 presents some enhanced results on **Test-R1000**, it can be observed that our method not only corrects the intrinsic appearance but also enhances details. All the compared methods cannot obtain satisfactory results. Some of them even introduce undesirable color artifacts in their enhanced results to some extent, such as UIEM and UWCNN. FUIE-GAN, Water-Net and Ucolor under-enhance these images, remaining a certain degree of color distortion in the results. Most methods fail to restore the structural details of underwater scene, in which UGAN and UNTV even introduce serious artifacts at the boundary of objects.

Results of the **UCCS** and **SQUID** dataset are shown in Fig.10 and Fig.11. As shown, for images with blueish or greenish tones, the proposed method significantly removes the haze and color casts, and effectively recovers details, producing visually pleasing results. All the comparison methods cannot restore the realistic color. Most of them suffer from obvious under-enhancement, such as UIEM, Water-Net and Ucolor. UNTV and UGAN still show a limited effect on the detail recovery. UWCNN, FUIE-GAN produce extra color deviations, and LCNet even introduces severe artificial colors.

Results of the **UIEB** and **EUVP** dataset are delivered in Fig.12 and Fig.13. For the image with light yellowish tone in

shallow water areas or the image with dark yellowish tone in deep water areas, the proposed method still achieves notable superiority in both color correction and detail preservation. All the methods under comparison fail to correct the intrinsic color. UIEM over-enhances the brightness that results, which seems unrealistic in the real world. For these low-light underwater images, most methods generate unrealistic results with color artifacts and cannot effectively improve the visibility of objects, and often amplify noise in their enhanced results. Our method not only effectively increases the brightness of images but also refines the object edges, producing realistic results with correct color from extremely noisy.

Results of the **UFO-120** dataset are shown in Fig.14. For these low or high quality underwater images, compared to most existing methods, the proposed method significantly reduces color distortion and satisfactorily removes blurriness. It can be seen that the images enhanced by UWCNN, UNTV and UIEM have obvious reddish color shift and artifacts in some regions. Besides, UGAN often introduces undesirable artifacts at the boundary of objects. Most methods cannot correct the colors well and even amplify color deviation (e.g., the color of background). FUIE-GAN and Ucolor can produce relatively good results. However, they still contain numerous noises and color distortion. All the quantitative and visual comparisons suggest that the proposed method can produce visually pleasing results and have more robust performance in handling images taken in a diversity of underwater scenes.

3) *Color Restoration Evaluation*: To analyze the accuracy of color restoration, we conduct the comparisons on SQUID-16. The average values of the color restoration accuracy are reported in Table III. As reported, the proposed method achieves the lowest mean angular error on Set B, Set C and Set D, which is greatly improved compared with other algorithms. Moreover, the proposed method obtains the lowest average score across all test images. Fig. 15 presents some visual comparisons on images of SQUID-16, it can be observed that our method not only obtains a pleasing visual perception, but also achieves a quite good color restoration accuracy. Such results further demonstrate the superiority of the proposed method in underwater image color correction.

#### D. Ablation Studies and Analysis

In this subsection, we first evaluate the performance of our proposed RUIQA and analyze its superiority. Subsequently, a series of ablation studies are conducted to analyze the contribution of each proposed component. Moreover, we study the influence of different  $\lambda$  on intra-domain adaptation training and TUDA testing.

1) *Effectiveness of the Proposed RUIQA Method*: As mentioned above, 90% image pairs (i.e., 7200) of the underwater ranking data are randomly selected as training data, and the other 10% image pairs (i.e., 800) are used for IQA testing. To validate the effectiveness of the proposed metric, we compare it with eight state-of-the-art IQA methods, including non-learning methods (UIQM [38], UCIQE [37], NIQE [53], CCF [54] and NUIQ [55]) and learning-based



Fig. 14. Visual comparisons on challenging underwater images sampled from **UFO-120**. From left to right are raw underwater images, and the results of UIEM [18], UNTV [17], UGAN [22], UWCNN [20], FUIE-GAN [42], LCNet [27], Water-Net [24], Ucolor [26] and our proposed TUDA.

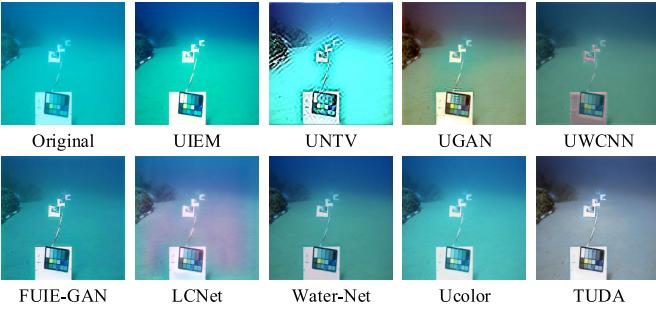


Fig. 15. Visual comparisons on the **SQUID-16** dataset. Obviously, our results have a pleasing visual perception and good color restoration accuracy.



Fig. 16. Visual comparisons in terms of UCIQE, UIQM and our proposed RUIQA metrics. It is obvious that our quantitative scores can better represent subjective quality.

TABLE IV  
EXPERIMENTS ON THE SAME RANKING TESTING DATASET

Methods	PLCC $\uparrow$	SROCC $\uparrow$	KROCC $\uparrow$	RMSE $\downarrow$
UIQM	0.435	0.390	0.267	12.350
UCIQE	0.350	0.278	0.189	11.950
NIQE	0.125	0.106	0.072	34.200
CCF	0.398	0.622	0.438	15.560
NUIQ	0.672	0.649	0.477	13.960
B-FEN	0.852	0.839	0.657	6.633
ResNet-50	0.826	0.805	0.598	11.230
DenseNet-121	0.840	0.822	0.612	9.512
Proposed RUIQA	<b>0.904</b>	<b>0.900</b>	<b>0.728</b>	<b>5.470</b>

methods (ResNet-50, DenseNet-121, B-FEN [44]). All models are retrained using the same data as our method. Four metrics are adopted to measure the performance, including Pearson linear correlation coefficient (PLCC), Spearman rank order correlation coefficient (SROCC), Kendall rank order

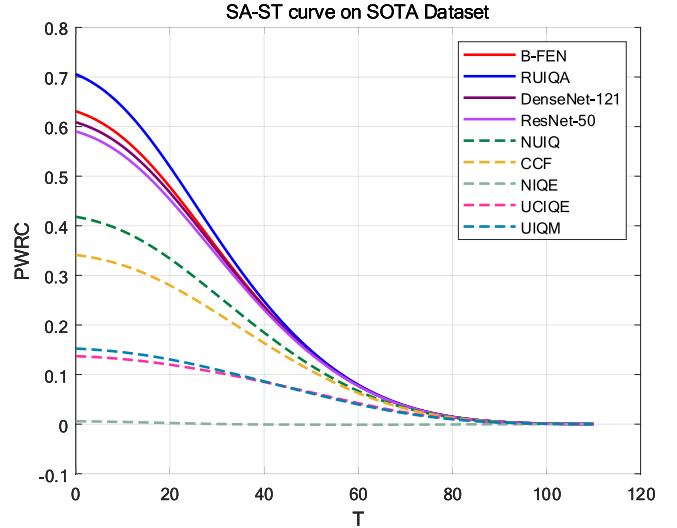


Fig. 17. SA-ST curve of the PWRC indicator on ranking testing dataset. The horizontal axis and vertical axis represent the sensory threshold (ST) and the sorting accuracy (SA), respectively. Solid lines refer to learning-based methods, and dashed lines refer to non-learning methods.

correlation coefficient (KROCC) and root mean square error (RMSE). A higher PLCC, SROCC and KROCC or a lower RMSE score denotes that the image quality scores obtained by the algorithm are more consistent with the subjective perceived qualities.

Table IV reports the quantitative comparison results, where the performance of deep learning-based methods are all better than non-learning methods. Among them, our method achieves the best performance, and even has good correlation with MOS on the order of 0.900 and achieves the gain of 0.5 to 0.65 in comparison to commonly used UCIQE and UIQM, showing the superiority of our metric. NIQE shows a very low correlation since the characteristic of this indicator is not suitable for evaluating underwater images. We also show a visual comparison of the proposed RUIQA and other underwater image quality assessment methods in Fig.16. A larger value means that a better perceptual quality. It is evident that the proposed RUIQA can more accurately reflect the perceptual quality of underwater images.

To more accurately compare different IQA methods, we use the perceptually weighted rank correlation (PWRC) with SA-ST curves [56] to evaluate the correlation of algorithms and subjectively perceived preferences. Different from previous methods that adopting a single rank correlation coefficient,

TABLE V  
COMPARISONS OF DIFFERENT METHODS ON THE UIEB DATASET

Models	PSNR ↑	SSIM ↑	RUIQA ↑
Inputs	17.555	0.844	45.968
UIEM	16.444	0.866	43.736
UNTV	15.772	0.603	48.956
UGAN	18.981	0.779	48.614
UWCNN	14.198	0.691	38.255
FUIE-GAN	18.865	0.881	52.363
LCNet	18.899	0.883	51.631
TUDA	<b>19.361</b>	<b>0.919</b>	<b>64.021</b>

TABLE VI  
SROCC AND PLCC RESULTS OF DIFFERENT METHODS AND ABLATION STUDIES ON THE RANK TEST SET

Method	UCIQE	UIQM	UIQA	PUIQA	RUIQA
SROCC ↑	0.245	0.400	0.796	0.835	<b>0.900</b>
PLCC ↑	0.326	0.422	0.826	0.862	<b>0.904</b>

TABLE VII  
THE ABLATION STUDY OF THE PROPOSED INTER-DOMAIN ADAPTATION MODULE ON THE SYNTHETIC TEST SET AND THE REAL TEST TEST-R1000

Methods	Test-S1800↑		Test-R1000↑	
	PSNR↑	SSIM↑	RUIQA↑	Perceptual Score↑
BL	<b>27.300</b>	<b>0.964</b>	55.534	2.914
BL+ITE	26.768	0.962	<b>61.791</b>	<b>3.501</b>

TABLE VIII  
THE ABLATION STUDY OF THE PROPOSED INTRA-DOMAIN ADAPTATION MODULE ON THE REAL TEST SET TEST-R1000

Method	RUIQA ↑	Perceptual Score ↑
BL + ITE	61.791	3.501
BL + ITE + ITA	<b>63.674</b>	<b>4.033</b>

TABLE IX  
THE ABLATION STUDY ON HYPER-PARAMETER  $\lambda$  FOR DIVIDING THE REAL UNDERWATER DATA INTO THE EASY AND HARD SAMPLES

$\lambda$	0.40	0.45	0.50	0.55	0.60	0.65
RUIQA↑	63.99	<b>64.01</b>	63.86	63.55	63.71	63.29

PWRC interprets the sorting accuracy SA with a complicated SA-ST curve by adjusting the perception threshold ST. A higher SA value denotes that the IQA algorithm is more consistent with human visual perception. Fig. 17 presents the SA-ST curve of different methods on ranking testing dataset. As presented, learning-based IQA methods have better performance than non-learning methods, in which the proposed RUIQA obtains the highest value at any threshold ST. Such results demonstrate that our RUIQA is consistent with human subjective visual preference results.

We further conduct a quantitative comparison on the UIEB dataset in terms of PSNR/SSIM/RUIQA to verify the consistency between the proposed indicator and paired evaluation indicators. The average scores of PSNR and SSIM and RUIQA of different methods are reported in Table V,

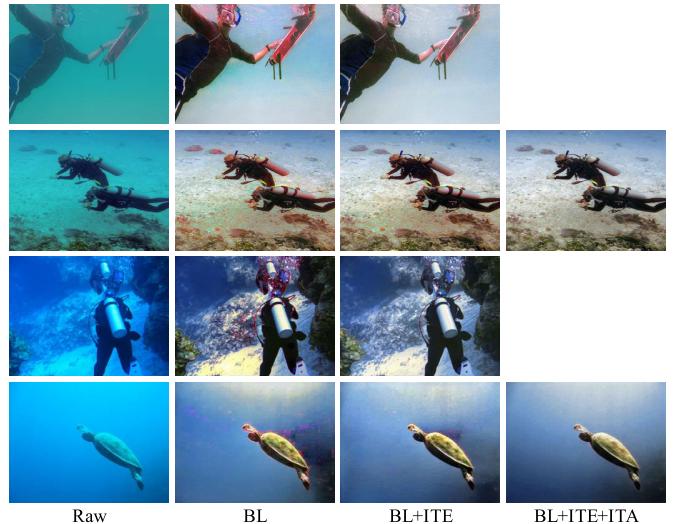


Fig. 18. Visual comparisons on some easy and hard samples. We can clearly see that our method can effectively handle easy and hard samples, especially on hard samples, our full model generates the most visually pleasing results.

TABLE X  
THE PARAMETERS, FLOPS AND RUNNING TIME OF DIFFERENT LEARNING-BASED METHODS

	Our TUDA	UGAN	FUIE-GAN	Water-Net
FLOPs (G)	174.4	3887	0.008	1937
Parameters (M)	31.36	57.17	4.216	1.091
Running time (s)	0.051	0.009	0.083	0.582

where we exclude Water-Net and Ucolor since they are trained based on the UIEB dataset. As shown, we can see that the rank order of the same algorithms on the three metrics is almost the same. For example, UWCNN achieves the lowest PSNR and SSIM scores. For the RUIQA scores, UWCNN still obtains the lowest value. Observing the scores of three all metrics, our method outperforms all competing methods. Such results demonstrate the effectiveness of our TUDA and prove that our RUIQA is consistent with the paired evaluation indicators.

In addition, an ablation study is conducted to analyze the contribution of each component: 1) UIQA: using the IQA network to directly predict image quality score; 2) PUIQA: using the ResNet50 network pre-trained on ImageNet data as our initialization backbone model; 3) RUIQA: using the ResNet50 network pre-trained on our rank data as our initialization backbone model. As presented in Table VI, we can see that our RUIQA achieves the best evaluation performance and is significantly better than UIQA and PUIQA. It's worth mentioning that the ImageNet has more than 1.28 million images and our rank training dataset only contains 720 image pairs. This indicates that the pre-trained ResNet50 network on the rank dataset can capture sufficient perceptual quality information of underwater image, and then quickly help the IQA task better predict the image quality score.

2) *Effectiveness of the Inter-Domain Adaptation Phase:* We perform an ablation study of 60 images randomly chosen from enhanced images in Test-R1000 to evaluate

the effectiveness of the inter-domain adaptation, as follows: 1) BL: baseline network (trained on synthetic data); 2) BL+ITE: baseline network with the inter-domain adaptation, i.e., triple-alignment network. Results are listed in Table VII. It is evident that baseline network has only slightly higher PSNR values in comparison with our triple-alignment network, but the perceptual quality is far worse than the triple-alignment network (6.257 and 0.587 lower on average in RUIQA and perceptual score of user study, respectively). Such results show that the inter-domain adaptation part generates the enhanced results with well reconstructed details (high fidelity) and good perceptual visual quality. Some examples are presented in Fig.18. It is observed that our inter-domain adaptation better corrects color casts and avoids over-enhancement than baseline network.

3) *Effectiveness of the Intra-Domain Adaptation Phase:* In the intra-domain adaptation part, we conduct an ablation study of 60 images randomly selected from enhanced results in **Test-R1000** with the following settings: 1) BL+ITE: baseline network with the inter-domain adaptation; 2) BL+ITE+ITA: baseline network with the inter-domain and the intra-domain adaption. The average RUIQA value and perceptual score are presented in Table VIII. It can be seen that BL+ITE+ITA achieves better performance, even the average performance gains up to 1.883 and 0.532 in two metrics, respectively. This indicates that intra-domain adaptation can effectively process hard samples and significantly improve the perceptual quality of the image, making enhanced results more subject to human preferences. In addition, a few samples are illustrated in Fig.18. It can be noted that if only the inter-domain adaptation phase is conducted, the results of some hard samples still contain some noises and over-enhancement artifacts in some region. In other words, the proposed intra-domain adaptation part is robust for real-world extremely hard underwater image enhancement, producing visually more pleasing results.

4) *Analysis of Hyperparameter  $\lambda$ :*  $x_n$  denotes the real underwater images for inter-domain training,  $n \in (1, 2900)$ . The inter-domain enhancement part  $G_{inter}^E$  receives the input  $x_n$  and outputs the enhanced image  $\hat{x}_n$ . The proposed RUIQA is used to evaluate their perceptual quality score, i.e.,  $MOS_n = RUIQA(\hat{x}_n)$ . We rank the  $MOS_n$  value (i.e.,  $MOS_n^{Rank}$ ) and select the corresponding  $MOS_n^{Rank}$  value of hyperparameter  $\lambda$  as the threshold (i.e.,  $MOS_n^{\lambda*Rank}$ ) to separate the real underwater data  $x_n$  into easy and hard samples (i.e.,  $x_e$  and  $x_h$ ) for intra-domain training and the whole framework testing. Thus, different values of hyperparameter  $\lambda$  will have a significant impact on subsequent operations. Some experiments are conducted to decide the optimal  $\lambda$  in our framework. For a selected hyperparameter  $\lambda$ , we first conduct intra-domain adaptation training. Then, the 2900 real training data (**Train-R2900**) is utilized as validate data in the test pipeline (see Fig.8). Finally, we predict the average perceptual quality score of 2900 enhanced images in terms of the RUIQA metric, and set it as the metric for selecting  $\lambda$ . Results are reported in Table IX. It can be observed that when  $\lambda = 0.45$ , the proposed TUDA can achieve better performance.

### E. Model Complexity Analysis

We compare the FLOPs, parameters and time cost of some representative learning-based methods on a PC with an Intel(R) i5-10500 CPU, 16.0GB RAM, and a NVIDIA GeForce RTX 2080 Super. The test dataset is UIEB benchmark, which includes 890 images and its size is  $256 \times 256 \times 3$ . The source codes and test parameters of all methods are provided by their authors, and the results are presented in Table X.

As presented, the computational aspect and time cost of our method are ideal. UGAN has the shortest running time, but its FLOPs and parameters are the most, far exceeding our proposed method. The size, computation and time cost of FUIE-GAN are less than our method. However, the generalization performance on real underwater benchmarks is limited, not as good as our method. The parameters of Water-Net is the least, but its FLOPs and time cost are large since it does not use down-sampling operation in the whole network structure. These comparison results demonstrate that our TUDA can achieve good performance and efficiency.

## V. CONCLUSION

In this paper, a novel two-phase underwater domain adaptation method is proposed for enhancing underwater images, which contains an inter-domain adaptation phase and an intra-domain adaptation phase to jointly optimize the inter-domain gap and the intra-domain gap. Firstly, a triple-alignment network is proposed to jointly perform image-level, feature-level and output-level alignment using adversarial learning for better closing the inter-domain gap. Secondly, a simple yet efficient rank-based underwater IQA method is developed, which can evaluate the perceptual quality of underwater images with the aid of rank information, named RUIQA. Finally, coupled with the proposed RUIQA, an easy/hard adaptation technique is conducted to effectively reduce the intra-domain gap between easy and hard samples. Extensive experiments on four real underwater benchmarks demonstrate that the proposed method can significantly perform favorably against other state-of-the-art algorithms, particularly on eliminating color deviation, increasing contrast and avoiding over-enhancement.

## REFERENCES

- [1] R. Hummel, “Image enhancement by histogram transformation,” *Comput. Graph. Image Process.*, vol. 6, pp. 184–195, Apr. 1977.
- [2] J. Hu, Q. Jiang, R. Cong, W. Gao, and F. Shao, “Two-branch deep neural network for underwater image enhancement in HSV color space,” *IEEE Signal Process. Lett.*, vol. 28, pp. 2152–2156, 2021.
- [3] 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.
- [4] K. Zuiderveld, “Contrast limited adaptive histogram equalization,” in *Proc. Graph. Gems*, 1994, pp. 474–485.
- [5] M. S. Hitam, W. N. J. H. W. Yussof, E. A. Awalludin, and Z. Bachok, “Mixture contrast limited adaptive histogram equalization for underwater image enhancement,” in *Proc. Int. Conf. Comput. Appl. Technol. (ICCAT)*, Jan. 2013, pp. 1–5.
- [6] 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.

- [7] X. Fu, P. Zhuang, Y. Huang, Y. Liao, X.-P. Zhang, and X. Ding, "A retinex-based enhancing approach for single underwater image," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 4572–4576.
- [8] S. Zhang, T. Wang, J. Dong, and H. Yu, "Underwater image enhancement via extended multi-scale retinex," *Neurocomputing*, vol. 245, pp. 1–9, Jul. 2017.
- [9] 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. 2018.
- [10] C. O. Ancuti, C. Ancuti, C. D. Vleeschouwer, and M. Sbert, "Color channel compensation (3C): A fundamental pre-processing step for image enhancement," *IEEE Trans. Image Process.*, vol. 29, pp. 2653–2665, 2020.
- [11] 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. 2012.
- [12] P. L. J. Drews, 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. 2016.
- [13] 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.
- [14] Y.-T. Peng and P. C. Cosman, "Underwater image restoration based on image blurriness and light absorption," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1579–1594, Apr. 2017.
- [15] 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. 48, no. 8, pp. 2822–2837, Aug. 2021.
- [16] D. Akkaynak and T. Treibitz, "Sea-thru: A method for removing water from underwater images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1682–1691.
- [17] J. Xie, G. Hou, G. Wang, and Z. Pan, "A variational framework for underwater image dehazing and deblurring," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 6, pp. 3514–3526, Jun. 2022.
- [18] W. Song, Y. Wang, D. Huang, A. Liotta, and C. Perra, "Enhancement of underwater images with statistical model of background light and optimization of transmission map," *IEEE Trans. Broadcast.*, vol. 66, no. 1, pp. 153–169, Mar. 2020.
- [19] C. Li, J. Guo, B. Wang, R. Cong, Y. Zhang, and J. Wang, "Single underwater image enhancement based on color cast removal and visibility restoration," *J. Electron. Imag.*, vol. 25, no. 3, Jun. 2016, Art. no. 033012.
- [20] 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.
- [21] J. Li, K. A. Skinner, R. M. Eustice, and M. Johnson-Roberson, "WaterGAN: Unsupervised generative network to enable real-time color correction of monocular underwater images," *IEEE Robot. Autom. Lett.*, vol. 3, no. 1, pp. 387–394, Jan. 2018.
- [22] C. Fabbri, M. J. Islam, and J. Sattar, "Enhancing underwater imagery using generative adversarial networks," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 7159–7165.
- [23] C. Li, J. Guo, and C. Guo, "Emerging from water: Underwater image color correction based on weakly supervised color transfer," *IEEE Signal Process. Lett.*, vol. 25, no. 3, pp. 323–327, Mar. 2018.
- [24] C. Li et al., "An underwater image enhancement benchmark dataset and beyond," *IEEE Trans. Image Process.*, vol. 29, pp. 4376–4389, 2020.
- [25] X. Chen, J. Yu, S. Kong, Z. Wu, X. Fang, and L. Wen, "Towards real-time advancement of underwater visual quality with GAN," *IEEE Trans. Ind. Electron.*, vol. 66, no. 12, pp. 9350–9359, Dec. 2019.
- [26] 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.
- [27] N. Jiang, W. Chen, Y. Lin, T. Zhao, and C.-W. Lin, "Underwater image enhancement with lightweight cascaded network," *IEEE Trans. Multimedia*, vol. 24, pp. 4301–4313, 2021.
- [28] 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.
- [29] J. S. Jaffe, "Computer modeling and the design of optimal underwater imaging systems," *IEEE J. Ocean. Eng.*, vol. 15, no. 2, pp. 101–111, Apr. 1990.
- [30] A. Dudhane, P. Hambarde, P. Patil, and S. Murala, "Deep underwater image restoration and beyond," *IEEE Signal Process. Lett.*, vol. 27, pp. 675–679, 2020.
- [31] Y. Shao, L. Li, W. Ren, C. Gao, and N. Sang, "Domain adaptation for image dehazing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2808–2817.
- [32] F. Pan, I. Shin, F. Rameau, S. Lee, and I. S. Kweon, "Unsupervised intra-domain adaptation for semantic segmentation through self-supervision," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3763–3772.
- [33] I. Shin, S. Woo, F. Pan, and I. Kweon, "Two-phase pseudo label densification for self-training based domain adaptation," 2020, *arXiv:2012.04828*.
- [34] C. Zheng, T.-J. Cham, and J. Cai, "T2Net: Synthetic-to-realistic translation for solving single-image depth estimation tasks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 767–783.
- [35] S. Zhao, H. Fu, M. Gong, and D. Tao, "Geometry-aware symmetric domain adaptation for monocular depth estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 9780–9790.
- [36] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of Wasserstein GANs," in *Proc. Neural Inf. Process. Syst.*, Mar. 2017, pp. 5767–5777.
- [37] 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.
- [38] 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.
- [39] D. Li, T. Jiang, W. Lin, and M. Jiang, "Which has better visual quality: The clear blue sky or a blurry animal?" *IEEE Trans. Multimedia*, vol. 21, no. 5, pp. 1221–1234, May 2019.
- [40] S. Su et al., "Blindly assess image quality in the wild guided by a self-adaptive hyper network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3664–3673.
- [41] W. Zhang, Y. Liu, C. Dong, and Y. Qiao, "RankSrgan: Generative adversarial networks with ranker for image super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 3096–3105.
- [42] 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.
- [43] 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.
- [44] Q. Wu, L. Wang, K. N. Ngan, H. Li, F. Meng, and L. Xu, "Subjective and objective de-raining quality assessment towards authentic rain image," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 11, pp. 3883–3897, Nov. 2020.
- [45] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, Int. Telecommun. Union, Geneva, Switzerland, 2002.
- [46] *Subjective Video Quality Assessment Methods for Multimedia Applications*, Int. Telecommun. Union, Geneva, Switzerland, 1999.
- [47] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.
- [48] N. Ponomarenko et al., "Color image database TID2013: Peculiarities and preliminary results," in *Proc. Eur. Workshop Vis. Inf. Process. (EUVIP)*, 2013, pp. 106–111.
- [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] 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. 5967–5976.
- [51] B. Li et al., "Benchmarking single-image dehazing and beyond," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 492–505, Jan. 2019.
- [52] M. J. Islam, M. Fulton, and J. Sattar, "Toward a generic diver-following algorithm: Balancing robustness and efficiency in deep visual detection," *IEEE Robot. Autom. Lett.*, vol. 4, no. 1, pp. 113–120, Jan. 2019.
- [53] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.

- [54] Y. Wang et al., "An imaging-inspired no-reference underwater color image quality assessment metric," *Comput. Electr. Eng.*, vol. 70, pp. 904–913, Aug. 2018.
- [55] Q. Jiang, Y. Gu, C. Li, R. Cong, and F. Shao, "Underwater image enhancement quality evaluation: Benchmark dataset and objective metric," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 9, pp. 5959–5974, Sep. 2022.
- [56] Q. Wu, H. Li, F. Meng, and K. N. Ngan, "A perceptually weighted rank correlation indicator for objective image quality assessment," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2499–2513, May 2018.



**Zhengyong Wang** received the B.S. degree from the School of Communication and Information Engineering, Yangzhou University, Yangzhou, China, in 2014. He is currently pursuing the Ph.D. degree with the Communication and Information Systems Laboratory, Shanghai University, Shanghai, China. His research interests include some low-level vision tasks such as underwater image enhancement, dehazing, deraining, deblurring, super-resolution, and some high-level vision tasks such as underwater object detection and segmentation.



**Liquan Shen** received the B.S. degree in automation control from Henan Polytechnic University, Henan, China, in 2001, and the M.E. and Ph.D. degrees in communication and information systems from Shanghai University, Shanghai, China, in 2005 and 2008, respectively. Since 2008, he has been with the Faculty of the School of Communication and Information Engineering, Shanghai University, where he is currently a Professor. From 2013 to 2014, he was at the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA, as a Visiting Professor. He has authored or coauthored more than 100 refereed technical papers in international journals and conferences in video coding and image processing. He also holds ten patents in the areas of image or video coding and communications. His research interests include versatile video coding (VVC), perceptual coding, video codec optimization, 3DTV, and video quality assessment.



**Mai Xu** (Senior Member, IEEE) received the B.S. degree from Beihang University, Beijing, China, in 2003, the M.S. degree from Tsinghua University, Beijing, in 2006, and the Ph.D. degree from Imperial College London, London, U.K., in 2010. From 2010 to 2012, he was a Research Fellow at the Department of Electrical Engineering, Tsinghua University. Since January 2013, he has been with Beihang University, where he was an Associate Professor and promoted as a Full Professor in 2019. From 2014 to 2015, he was a Visiting Researcher at MSRA. He has authored or coauthored more than 100 technical papers

in international journals and conference proceedings, such as *International Journal of Computer Vision (IJCV)*, *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (TPAMI)*, *IEEE TRANSACTIONS ON IMAGE PROCESSING (TIP)*, *IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING (JSTSP)*, *CVPR*, *ICCV*, *ECCV*, and *AAAI*. His research interests include image processing and computer vision. He is an Elected Member of the Multimedia Signal Processing Technical Committee and IEEE Signal Processing Society. He was a recipient of best paper awards of three IEEE conferences. He served as the Area Chair or a TPC Member for many conferences, such as *ICME* and *AAAI*. He served as an Associate Editor for *IEEE TRANSACTIONS ON IMAGE PROCESSING (TIP)* and *IEEE TRANSACTIONS ON MULTIMEDIA (TMM)* and also a Lead Guest Editor for *IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING (JSTSP)*.



**Mei Yu** received the B.S. and M.S. degrees from the Hangzhou Institute of Electronics Engineering, Hangzhou, China, in 1990 and 1993, respectively, and the Ph.D. degree from Ajou University, Suwon, South Korea, in 2000. She is currently a Professor with the Faculty of Information Science and Engineering, Ningbo University, Ningbo, China. Her research interests mainly include image or video coding and visual perception.



**Kun Wang** received the B.S. and M.E. degrees in communication engineering from Shanghai University, Shanghai, China, in 2019 and 2021, respectively. His research interests include underwater image/video enhancement, dehazing, deraining, and machine learning.



**Yufei Lin** received the B.S. degree in communication engineering from the Nanjing University of Science and Technology, Nanjing, China, in 2019, and the M.E. degree in communication engineering from Shanghai University, Shanghai, China, in 2021. His research interests include underwater image/video enhancement, dehazing, deraining, and machine learning.