

Domain Adaptation for Underwater Image Enhancement

Zhengyong Wang, Liquan Shen, Mei Yu, Kun Wang, Yufei Lin and Mai Xu

Abstract—Recently, learning-based algorithms have shown impressive performance in underwater image enhancement. Most of them resort to training on synthetic data and achieve outstanding performance. However, these 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 underwater scenarios. Furthermore, 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, a new dual-alignment network is designed in the first phase, including a translation part for enhancing realism of input images, followed by an enhancement part. With performing image-level and feature-level adaptation in two parts by jointly adversarial learning, the network can better build invariance across domains and thus bridge the inter-domain gap. In the second phase, we perform an easy-hard classification of real data according to the assessed quality of enhanced images, where a 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 our TUDA is 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 images. Excellent underwater image enhancement 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 high-level computer vision tasks, such as object detection and recognition.

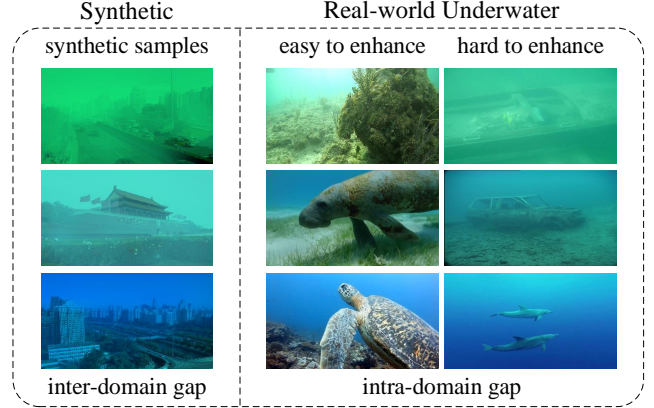


Fig. 1. Illustration of two challenges for underwater image enhancement. (1) Inter-domain gap challenge: the domain shift between the synthetic images and real images is often ignored, and thus the deep models trained on synthetic data often suffer a great performance drop in some real underwater images 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.

In the past decades, many algorithms have been proposed to enhance underwater images, ranging from traditional methods (image-based [1]–[8] and physical-based [9]–[14]) to learning-based methods [15]–[21]. Compared to traditional methods, learning-based methods tend to design end-to-end modules or integrate networks with physical priors to solve problems, which have better feature representation that benefits from the large data and powerful computational ability. Unfortunately, it is impractical in real-world to collect a number of real underwater images with distortion-free counterparts. Compared to real underwater data, synthetic data is much easier to be obtained. Thus, most deep methods exploit synthetic data to train the proposed models, achieving relatively promising performance. However, most of them ignore the domain shift problem from synthetic to real data, i.e., inter-domain gap, as shown in Fig. 1. These models learned from synthetic data often suffer a severe performance drop when facing some real underwater images with different distortion distributions.

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. As shown 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

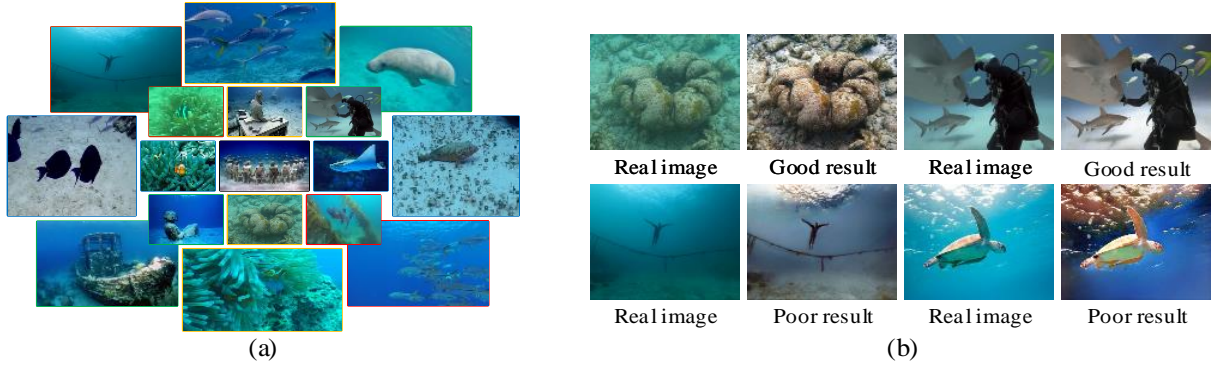


Fig. 2. (a) Examples of real-world underwater images, which have obvious different characteristics of underwater image quality degradation, e.g., color casts and blurred details. (b) Some results come from our inter-domain adaptation phase. Obviously, the results of some samples have higher perceptual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

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 presented 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)** in 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. To be specific, a new dual-alignment network is designed in the first phase, including a translation part and an enhancement part, one for the synthetic-to-real translation and another for image enhancement. Coupled with both image-level and feature-level adaptations in an end-to-end manner, two parts can cooperative with each other for learning more domain-invariant representations to better reduce the inter-domain gap.

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 presented in Fig.3. To the best of our knowledge, this is the first work that successfully explores the inter-domain and intra-domain adaptation jointly in the underwater image enhancement community. The main contributions of this paper are summarized as follows:

- 1) We propose a novel two-phase underwater domain adaptation network, called 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 dual-alignment architecture is designed in the inter-domain adaptation phase, which can effectively perform image-level and feature-level adaptations using jointly adversarial learning. Two alignment parts can improve each other, and the combination of them can better build invariance across domains and thus bridge 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

In this section, we briefly review previous related works in two aspects, i.e. underwater image enhancement and domain adaptation.

A. Underwater Image Enhancement

Underwater image enhancement approaches can be roughly categorized into three branches, i.e., image-based methods, physical-based methods and learning-based methods.

Image-based methods [1]–[8] mainly modify pixel values of underwater images to improve visual quality, including pixel values adjustment [1]–[3], [8], retinex decomposition [5], [6] and image fusion [4], [7], etc. For example, Zhang et al. [6] 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. [7] 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. [8] introduce a new color-channel-compensation pre-processing step in the opponent color channel to better

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

Most **physical-based methods** [9]–[14] are based on the underwater image formation model [22], in which the background light and transmission map are estimated by some priors. The priors include underwater dark channel prior [10], minimum information prior [11], blurriness prior [12] and color-line prior [13], etc. For example, built on underwater image blurriness and light absorption, Peng *et al.* [12] 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.* [11] estimate an optimal transmission map to restore underwater images, and exploit a histogram distribution prior to effectively improve the contrast and brightness. Recently, Berman *et al.* [13] 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 can restore underwater images well in some cases. However, when the priors are invalid, undesired artifacts and color casts are still inevitable appear in some regions.

Recently, with the development of deep learning, **learning-based methods** [15]–[21] have made significant progresses in underwater image enhancement. There are many methods improve performance by training their models on real underwater images. For example, to relax the need of paired training data, Li *et al.* [18] develop a weakly supervised underwater color transfer model based on cycle-consistent generative adversarial network (CycleGAN) and real data to correct color. As a pioneering work, Li *et al.* [19] build a real underwater image enhancement dataset, including totally 950 underwater raw images and reference images. The reference images are produced by 12 enhancement algorithms, and scored by 50 volunteers to choose the final results. With these images, Li *et al.* [19] design a gate fusion network, in which three confidence maps are learned to fuse three pre-processing versions into a decent result. Recently, Li *et al.* [21] develop an underwater image enhancement network in medium transmission-guided multi-color space for more robust enhancement. The methods trained on real data can produce visually pleasing results. However, they cannot restore the color and structure of objects well and tend to produce inauthentic results since the reference images are not the actual ground truths.

There are also many algorithms to train their networks using data synthesized from Generative Adversarial Networks [16], [17] or physical models [22], [23]. For example, combined with the knowledge of underwater imaging, Li *et al.* [16] 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.* [17] directly employ a CycleGAN to generate paired training data, and then a fully convolutional encoder-decoder is trained to improve underwater image quality. In addition, Li *et al.* [15] propose to synthesize ten types of underwater images based on an un-

derwater image formation model and some scene parameters. With the synthetic data, Li *et al.* [15] develop an end-to-end model to directly recover the clear underwater latent image first, and then conduct a post-processing to improve subjective visual effects. Dudhane *et al.* [23] improve the work of [15] 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 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 systematic studied in underwater image enhancement field. However, it has a wide range of applications in other fields such as image hazing [24], semantic segmentation [25], [26] and depth prediction [27], [28], etc. For example, Shao *et al.* [24] propose a domain adaptation for single image dehazing based on CycleGAN, in which a new bidirectional translation network is design to reduce the gap between synthetic and real images by jointly synthetic-to-real and real-to-synthetic image-level adaptations. Zhao *et al.* [28] 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 bridge the gap between synthetic and real domains, and thus generate high-quality depth maps.

More recently, Pan *et al.* [25] 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 are 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 dual-alignment network used for inter-domain adaptation and a novel underwater image quality assessment algorithm used for intra-domain adaptation are proposed. The detail 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 network is proposed, which consists of two parts: inter-domain adaptation and intra-domain adaptation. As shown in Fig. 3, in the inter-domain

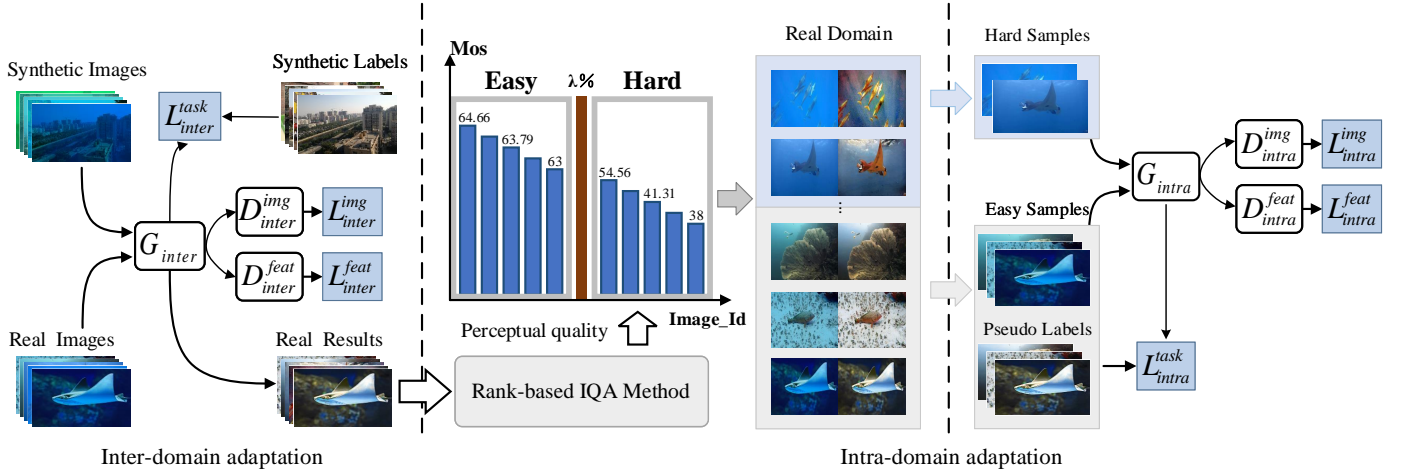


Fig. 3. Illustration of our proposed TUDA, which consists of two phases, inter-domain adaptation and intra-domain adaptation. In the inter-domain adaptation phase, G_{inter} can effectively reduce the distribution discrepancy between synthetic and real images using the image-level and feature-level discriminator D_{inter}^{img} and D_{inter}^{feat} . 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 λ 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} and D_{intra}^{feat} .

phase, a new dual-alignment network G_{inter} is developed to jointly perform image-level and feature-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 of synthetic to real 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., RUIQA) 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. 1st phase: Inter-domain Adaptation

Our proposed dual-alignment network aims to reduce the inter-domain adaptation gap between the synthetic and real data domain in both image level and feature level, as shown in Fig.4. The proposed network composes 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 [29] is adopted to perform image-level alignment, set as:

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

where \hat{I} represents the sampling distribution which is sampled uniformly from x_r and x_{st} , and λ_{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 [30]. 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 the same semantic content as x_s , but with a different appearance. Therefore, 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

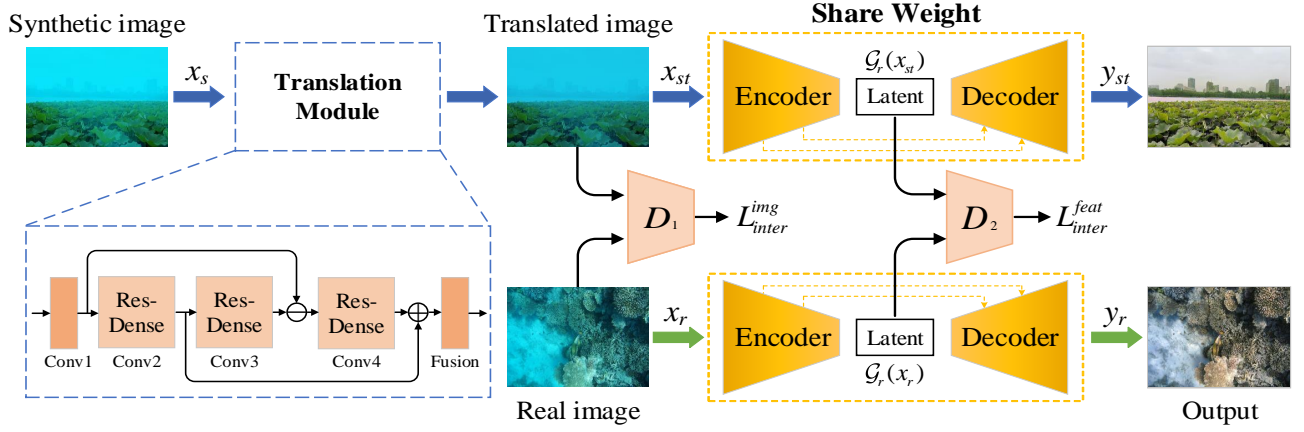


Fig. 4. Illustration of our dual-alignment network proposed in the inter-domain adaptation phase, trained on synthetic underwater image pairs and unpaired real images, which consists of an image translation part for enhancing realism of input images, followed by an image enhancement part. They are cooperatively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

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:

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

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 $\mathcal{G}_r(x_r)$. λ_{feat} is the penalty parameter, set as 10 in our experiments.

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

$$L_{inter} = \lambda_1 L_{inter}^{img} + \lambda_2 L_{inter}^{cont} + \lambda_3 L_{inter}^E + \lambda_4 L_{inter}^{feat} \quad (5)$$

where λ_1 , λ_2 , λ_3 and λ_4 are trade-off weights. In our work, they are set as 1, 100, 10 and 0.0005, respectively.

B. 2nd phase: Intra-domain Adaptation

As mentioned above, the intra-domain gap exists among real underwater images itself, and thus a straightforward method is to divide and conquer. 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 [31] and UIQM [32]. However, the experimental results in [19] 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 [33], [34]. Although these metrics achieve good results on ground images to some extent, the performance is 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 [35] 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. And then we fine-tune it to more accurately predict the perceptual quality of enhanced images. Differently, in [35], 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, our RUIQA consists of three stages: generating rank images, training ranker and fine-tuning network.

Generating rank images: A large number of underwater images are first collected from online and some public datasets [19], [30], [36], and then carefully selected and refined. Most of the collected images are weeded out, and about 3900 candidate images are remained. We randomly

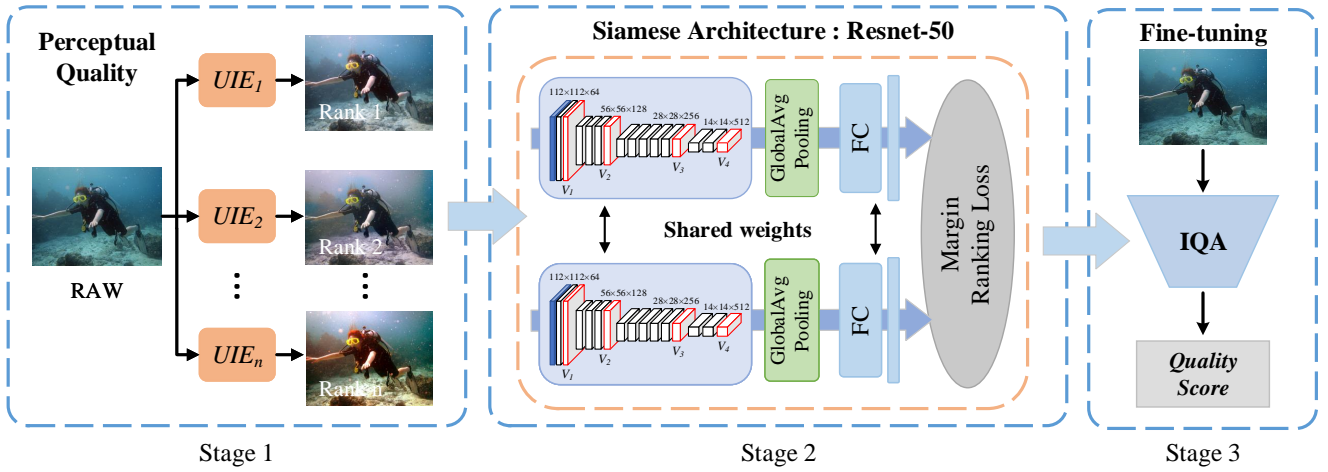


Fig. 5. The proposed RUIQA consists of three stages, namely stage 1: build a real-world underwater ranking dataset based on their perceptual quality; stage 2: train the Siamese architecture ResNet-50 using the ranking dataset; stage 3: perform a fine-tuning technique to predict the image quality score.

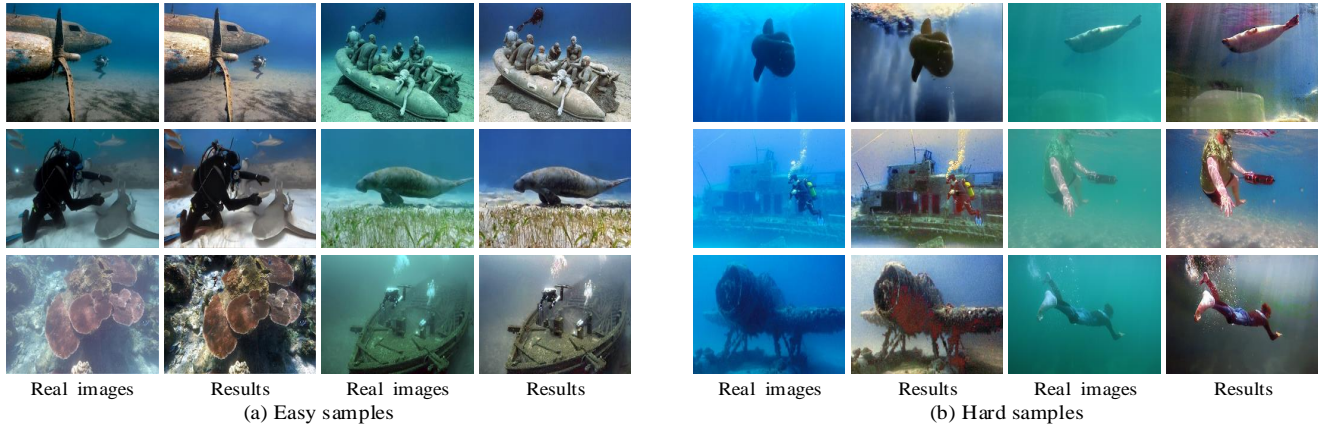


Fig. 6. Illustration of easy and hard samples. Their results come from the same inter-domain adaptation network. Obviously, the results of easy samples have higher perceptual quality, whereas the results of hard samples suffer from local artifacts, noise, color casts and over-enhancement, etc.

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 Fusion-12 [4], Fusion-18 [7], Two-step-based [37], Histogram prior [11], Blurriness-based [12], Water-Net [19], FUIE-GAN [36] 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 map 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 [38], the raw scores are refined by means of some standardized settings [39], [40] and the Mean Opinion Score (MOS) are calculated [41], [42], obtaining reliable subjective rating results. Our dataset and code will be publicly released on <https://github.com/Underwater-Lab/TUDA>.

Training ranker: With the obtained MOS values, the pairwise images and the corresponding ranking order labels can be obtained. Meanwhile, ResNet-50 [43] is employed as the Siamese network architecture to extract ranking information. The Siamese network is trained by a margin-ranking loss proposed in [35], which is beneficial for 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 Ranking images, is extracted to initialize our backbone network.

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 1×1 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 this 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 percep-

tual quality and are near to the human perception. In practice, a ratio λ is introduced to help the separation, which means the ratio of easy samples to total samples. The corresponding MOS value of the specified ratio λ 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-D4, how to obtain the best ratio λ is explored. It is very important for the intra-domain training and finally TUDA testing pipeline.

2) *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

$$L_{intra}^{img} = \mathbb{E}_{x_{et}} [D_{intra}^{img}(x_{et})] - \mathbb{E}_{x_h} [D_{intra}^{img}(x_h)] + \lambda_{img} \mathbb{E}_{\hat{I}} \left(\left\| \nabla_{\hat{I}} D_{intra}^{img}(\hat{I}) \right\|_2 - 1 \right)^2 \quad (6)$$

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:

$$L_{intra}^{cont} = w_k \sum_{k \in L_c} \|\phi_k(x_e) - \phi_k(x_{et})\|_1 \quad (7)$$

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:

$$L_{intra}^{task} = c * \|y_e - y_{et}\|_1 + d * \sum_{k \in L_c} \|\phi_k(y_e) - \phi_k(y_{et})\|_1 \quad (8)$$

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-world domain, we also perform a feature-level adaptation, 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:

$$L_{intra}^{feat} = \mathbb{E}_{x_{et}} [D_{intra}^{feat}(\mathcal{G}_h(x_{et}))] - \mathbb{E}_{x_h} [D_{intra}^{feat}(\mathcal{G}_h(x_h))] + \lambda_{feat} \mathbb{E}_{\hat{I}} \left(\left\| \nabla_{\hat{I}} D_{intra}^{feat}(\hat{I}) \right\|_2 - 1 \right)^2 \quad (9)$$

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 $\mathcal{G}_h(x_h)$. λ_{feat} is the penalty factor, set as 10 in this work. G_{intra}^T and G_{intra}^E are trained in an end-to-end manner, and thus the full loss function is as follow:

$$L_{intra} = \lambda_a L_{intra}^{img} + \lambda_b L_{intra}^{img} + \lambda_c L_{intra}^{task} + \lambda_d L_{intra}^{feat} \quad (10)$$

where $\lambda_a, \lambda_b, \lambda_c$ and λ_d are trade-off weights. In our work, we set them as 1, 100, 10 and 0.0005, respectively.

C. Architecture Details

The detail architecture of two transform modules (G_{inter}^T, G_{intra}^T) is shown in Fig.7. The down-sampling layer is not employed in the translator for avoiding valuable information loss. For image discriminators ($D_{inter}^{img}, D_{intra}^{img}$) and feature discriminator networks ($D_{inter}^{feat}, D_{intra}^{feat}$), PatchGANs [44] 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)¹ is used as our enhancement parts (G_{inter}^E, G_{intra}^E). It's worth mentioning that our test pipeline only need 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, we first describe the implementation details and experiment settings of our TUDA. Then, we compare it with existing representative methods on four publicly available real underwater benchmarks. 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 follow the physical model proposed in the project page of ANA-SYN. The synthetic dataset contains 9 water types² defined in [22], and each type has 1000 images which are randomly chosen from RTTS dataset [45]. The constructed dataset is divided into two parts, 7200 (800×9) images for training, denoted as **Train-S7200** and 1800 (200×9) 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 two parts, 2900 images for training, denoted as **Train-R2900** and 1000 images for testing **Test-R1000**. All images are resized to 256×256 and the pixel values are normalized to $[-1, 1]$. Furthermore, several data augmentation techniques are performed in the training phase, such as random rotating $90^\circ, 180^\circ, 270^\circ$ and horizontal flipping.

Our TUDA and RUIQA are implemented in Pytorch framework and all experiments are carried out on two NVIDIA Titan

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

²Type I, II, III, IA and IB for open ocean water and type 1C, 3C, 5C and 7C for coastal water

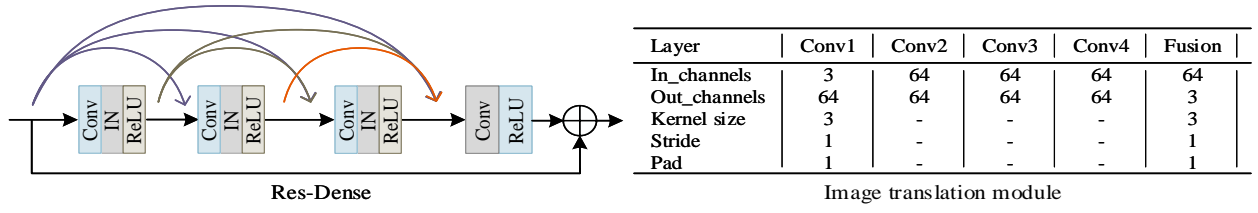


Fig. 7. Configurations of image translation module. "Conv1" is combined by a convolutional layer, a IN layer and a ReLU activation function. "Conv2", "Conv3", "Conv4" denote Res-Dense block. "Fusion" is combined by a convolution layer and a Tanh activation function.

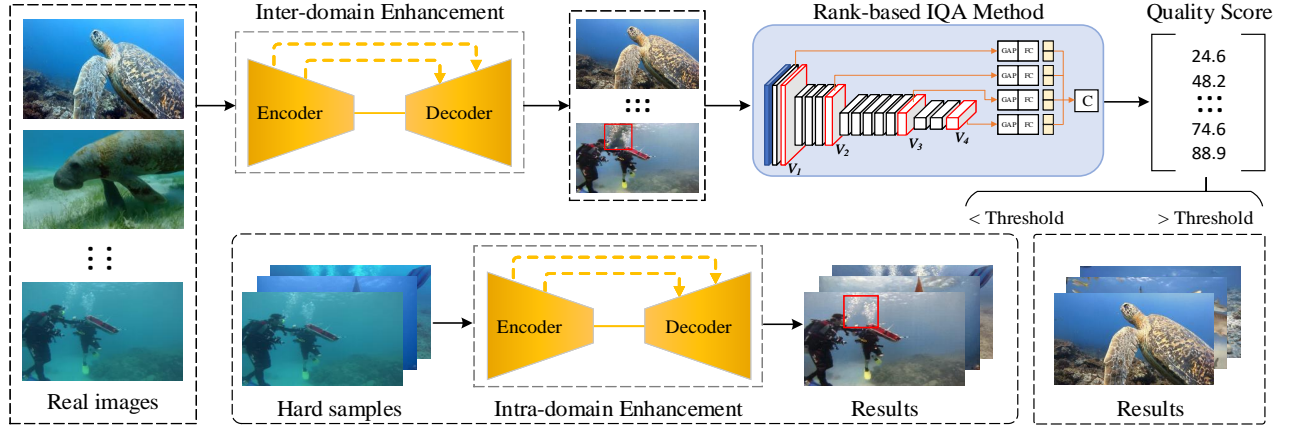


Fig. 8. An overview of our testing pipeline. The inter-domain enhancement part first takes real underwater images as input and outputs the corresponding inter-domain enhancement results. Then, our proposed rank-based IQA method evaluates the perceived quality of the enhancement result. When the score 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.

V GPUs. Adam optimizer with a learning rate of 1×10^{-4} is utilized to train G_{inter}^T , G_{inter}^E , G_{intra}^T and G_{intra}^E . For D_{inter}^{img} , D_{inter}^{feat} , D_{intra}^{img} and D_{intra}^{feat} , we adopt an Adam optimizer with learning rate of 2×10^{-4} as the optimization method. Default values of β_1 and β_2 are set as 0.5 and 0.999. 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 four publicly available real-world underwater benchmarks, i.e., **SQUID**³ [13], **UIEB**⁴ [19], **EUVP**⁵ [36] and **UFO-120**⁶ [46]. The compared algorithms include Fusion-12 [4], Fusion-18 [7], HE-Prior [11], UIBLA [12], UGAN [17], FUIE-GAN [36] and Water-Net [19]. The first four algorithms are traditional methods, while the remaining are deep-learning methods. For all the above-mentioned methods, we use the released test models and parameters to produce their results.

For results on real images, performances are measured by three no-reference underwater quality assessment metrics: UCIQE, UIQM and our 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 [19], [21], and thus the scores of UCIQE and UIQM are only for reference.

In addition, a user study is conducted to more accurately evaluate the visual quality of the results, in which 30 images are randomly selected from each testing dataset to scored. 15 volunteers are invited in this evaluation, and the scoring range is 0 to 5 levels, referring Bad, Poor, Fair, Good and Excellent, respectively. To evaluate the color restoration accuracy of different methods, we also calculate the color restoration accuracy on the average angular reproduction error [13] on the 16 representative examples presented in the project page⁷ of SQUID. The smaller color error, the better performance.

C. Comparisons with State-of-the-Art Methods

In this section, we conduct quantitative and visual comparisons on diverse challenging testing dataset. The results of different methods are reported in the following subsections.

Quantitative Comparisons: The quantitative comparison results of different methods on real challenging set are listed in Table I and Table II. As presented, HE-Prior achieves the highest scores in term of UCIQE, while our TUDA ranks the fourth best on all challenging set. For the UIQM scores, our method almost achieves the best performance across all datasets, and UGAN ranks the second best. The average values of the color restoration accuracy of different methods on 16 representative examples of SQUID are reported in the second

³The SQUID dataset contains 57 real underwater images

⁴The UIEB dataset contains 950 real underwater images

⁵The EUVP dataset contains about 1910 real underwater images

⁶The UFO-120 dataset contains about 3255 synthetic and real images

⁷http://csms.haifa.ac.il/profiles/tTreibitz/datasets/ambient_forwardlooking/index.html

TABLE I
QUANTITATIVE RESULTS (AVERAGE UIQM/UCIQE) OF DIFFERENT METHODS ON REAL BENCHMARKS (TEST-R1000, SQUID, UIEB, EUVP, AND UFO-120). THE TOP THREE RESULTS ARE MARKED IN RED, BLUE AND GREEN. '-' REPRESENTS THE RESULTS ARE NOT AVAILABLE.

Methods	UCIQE↑						UIQM↑					
	Test-R1000	SQUID	UIEB	EUVP	UFO-120	Avg	Test-R1000	SQUID	UIEB	EUVP	UFO-120	Avg
Fusion-12 [4]	0.607	0.561	0.612	0.610	0.619	0.613	2.729	1.507	2.870	2.779	2.769	2.769
Fusion-18 [7]	0.583	0.582	0.589	0.588	0.596	0.591	2.879	2.268	3.030	3.153	2.987	3.016
HE-Prior [11]	0.673	0.582	0.678	0.673	0.678	0.675	–	–	2.637	2.656	2.565	2.604
UIBLA [12]	0.581	0.500	0.595	0.594	0.624	0.605	2.030	–	2.262	2.053	2.061	2.183
Water-Net [19]	0.576	0.537	0.584	0.578	0.598	0.587	3.066	2.124	2.756	2.851	2.748	2.815
FUIE-GAN [36]	0.548	0.488	0.560	0.556	0.582	0.566	2.777	1.791	2.949	3.079	2.954	2.952
UGAN [17]	0.611	0.584	0.619	0.604	0.624	0.615	3.130	2.780	3.180	3.237	3.152	3.172
Our TUDA	0.596	0.589	0.605	0.596	0.608	0.602	3.125	2.643	3.203	3.313	3.166	3.200

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

Methods	Color Error↓	RUIQA↑					Perceptual Scores↑				
	SQUID	Test-R1000	SQUID	UIEB	EUVP	UFO-120	Test-R1000	SQUID	UIEB	EUVP	UFO-120
Fusion-12 [4]	31.653	49.204	44.958	51.185	47.375	50.379	2.442	2.049	2.404	2.464	2.462
Fusion-18 [7]	6.008	59.850	60.430	60.631	57.208	59.499	3.209	3.208	3.320	3.153	3.216
HE-Prior [11]	21.586	32.287	23.239	35.420	36.549	34.403	1.731	1.072	1.558	1.973	1.580
UIBLA [12]	34.214	42.727	41.395	43.170	40.227	42.213	2.020	1.997	2.084	1.891	2.073
Water-Net [19]	23.352	45.040	46.793	53.004	50.611	52.976	2.093	2.740	2.593	2.575	2.618
FUIE-GAN [36]	26.847	50.784	46.061	51.928	49.150	52.871	2.522	2.042	2.782	2.608	2.802
UGAN [17]	10.158	46.996	46.461	48.391	47.412	48.218	1.842	1.874	2.031	2.156	2.007
Our TUDA	11.718	61.436	61.127	60.952	59.262	60.067	3.595	3.993	3.675	3.531	3.537

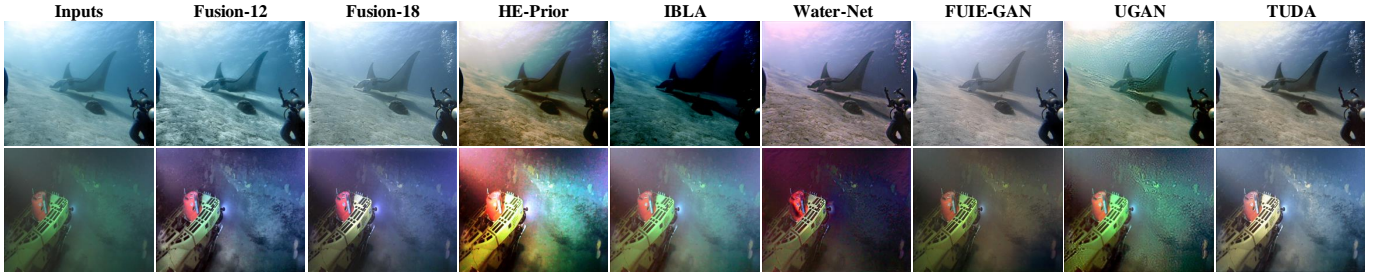


Fig. 9. Visual comparisons on challenging underwater images sampled from **Test-R1000**. From left to right are raw underwater images, and the results of Fusion-12 [4], Fusion-18 [7], HE-Prior [11], UIBLA [12], Water-Net [19], FUIE-GAN [36], UGAN [17] and our proposed TUDA.

column of Table II. It can be observed that our TUDA achieves relatively low average error, making a more effectively recovery of a scene's colors. Fusion-18 obtains the lowest color error than other methods. However, its performance is not as good as our TUDA in terms of RUIQA and Perceptual Scores. Among them, our TUDA achieves the best performance. Such results demonstrate that our TUDA produces visually more convincing results and has more robust performance in handling images taken in a diversity of underwater scenes.

The deep methods trained based on real data including Water-Net and FUIE-GAN perform relatively well, but they do not restore green or some excessively distorted images due to ignoring the intra-domain gap among real underwater images itself, and thus its performance is limited. UGAN trains the model using the synthetic data generated by GAN methods. Since the inter-domain gap is not effectively reduced, the results often contain various artifacts, and thus the subjective effect is not good, and the score is relatively low.

There is an interesting finding from the quantitative results. HE-Prior almost achieves the highest UCIQE scores on all real datasets. However, its perceptual score is the lowest, which means it has the worst subjective quality. In our opinion, this is mainly due to the fact that UCIQE pays too much attention to local features (color) and ignores the entire image, and thus it is not consistent with human visual perception in some cases, especially when the enhanced image is over-enhanced (please refer to Fig. 14) [19].

Visual Comparisons: Some visual comparisons on **Test-R1000** and **SQUID** are shown in Fig. 9 and Fig. 10, respectively. For these images, except for Fusion-18 [7] and our TUDA, other competing methods cannot achieve satisfactory results. Some of them even introduce undesirable color artifacts in their enhanced results to some extent, such as Fusion-12 [4], HE-Prior [11], FUIE-GAN [36] and UIBLA [12]. Meanwhile, most methods fail to restore the structural details of underwater scene, in which UGAN [17] even introduces

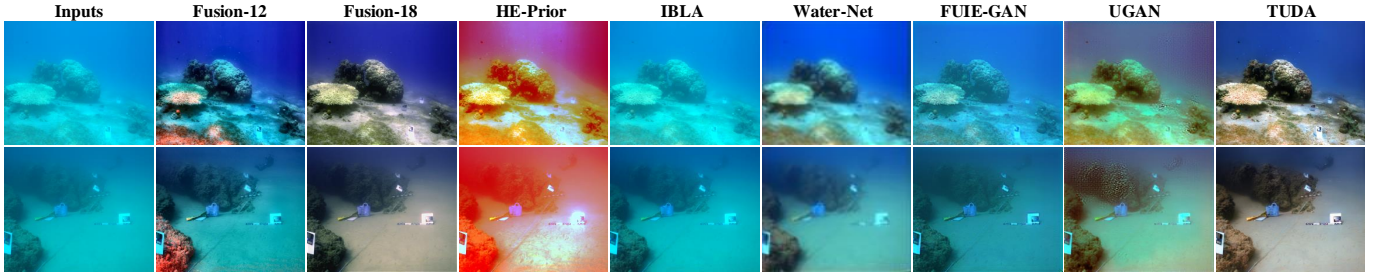


Fig. 10. Visual comparisons on challenging underwater images sampled from **SQUID**. From left to right are raw underwater images, and the results of Fusion-12 [4], Fusion-18 [7], HE-Prior [11], UBLA [12], Water-Net [19], FUIE-GAN [36], UGAN [17] and our proposed TUDA.

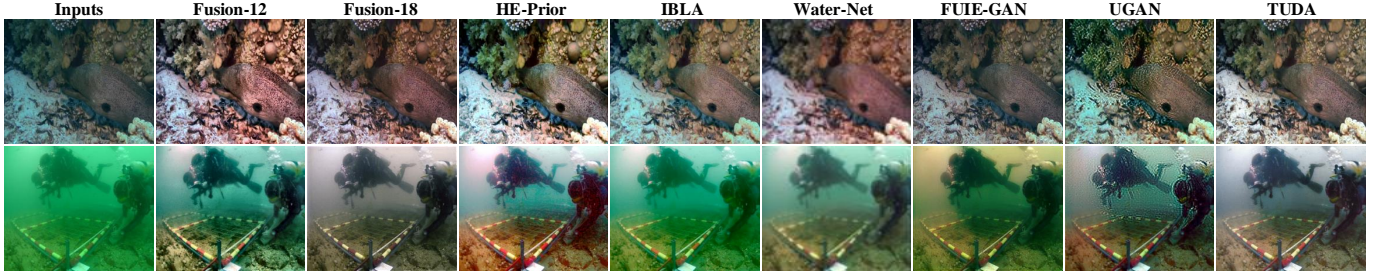


Fig. 11. Visual comparisons on challenging underwater images sampled from **UIEB**. From left to right are raw underwater images, and the results of Fusion-12 [4], Fusion-18 [7], HE-Prior [11], UBLA [12], Water-Net [19], FUIE-GAN [36], UGAN [17] and our proposed TUDA.

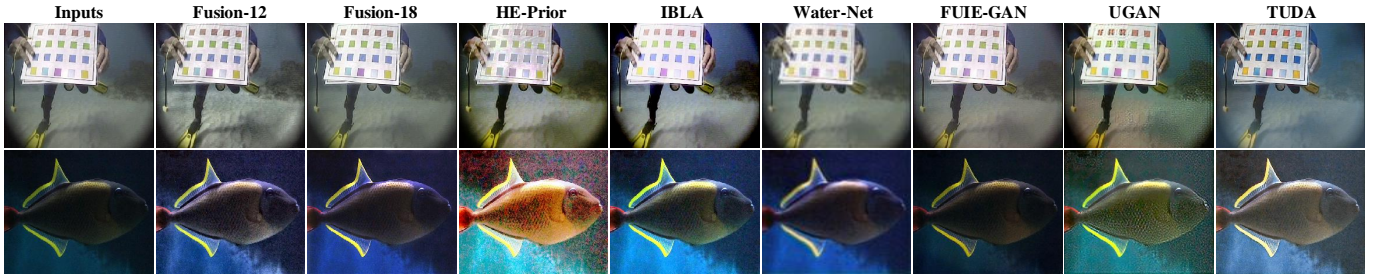


Fig. 12. Visual comparisons on challenging underwater images sampled from **EUVP**. From left to right are raw underwater images, and the results of Fusion-12 [4], Fusion-18 [7], HE-Prior [11], UBLA [12], Water-Net [19], FUIE-GAN [36], UGAN [17] and our proposed TUDA.

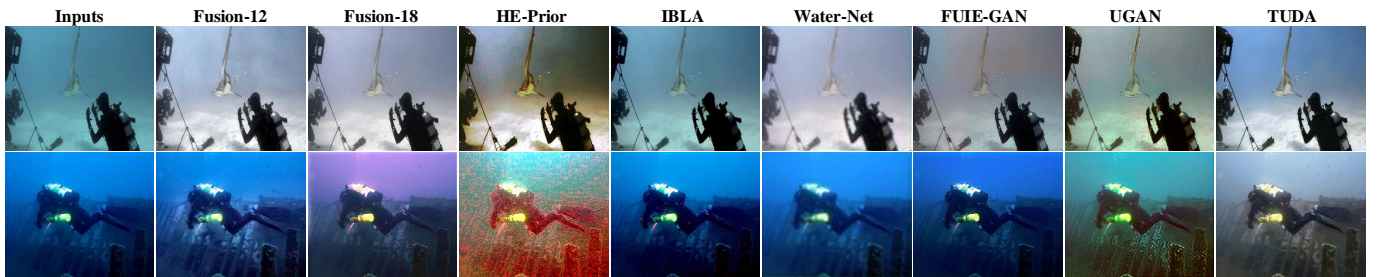


Fig. 13. Visual comparisons on challenging underwater images sampled from **UF0-120**. From left to right are raw underwater images, and the results of Fusion-12 [4], Fusion-18 [7], HE-Prior [11], UBLA [12], Water-Net [19], FUIE-GAN [36], UGAN [17] and our proposed TUDA.

serious artifacts at the boundary of objects. Fusion-18 [7] can restore better color than other methods, but the performance in recovering object details is not as good as our TUDA.

The results of different methods on challenging underwater images sampled from **UIEB** and **EUVP** are presented in Fig. 11 and Fig. 12. As presented, for the image with the greenish tone, our TUDA significant removes the haze and color casts, and effectively recovers details, producing visually pleasing results. In comparison, all the comparison methods

cannot produce the realistic color. Most of them suffer obvious over-enhancement and under-enhancement, such as Fusion-12 [4], Fusion-18 [7], HE-Prior [11], UBLA [12] and Water-Net [19]. Fusion-18 [7] even lost the original color of the object in the second image. 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 TUDA not only can effectively increase the brightness

of images but also refine the object edges, produce realistic results with correct color from extremely noisy.

Visual comparisons on challenging underwater images sampled from **UFO-120** are shown in Fig. 13. Compared to most existing methods, our TUDA significantly reduces color distortion and satisfactorily removes blurriness. It can be seen that the images enhanced by HE-Prior [11] have obvious reddish color shift and artifacts in some regions. Besides, UGAN [17] often introduces undesirable artifacts at the boundary of objects. Most methods cannot correct the colors well, even amplify color deviation (e.g., the color of background). Fusion-18 [7] can produce relatively good results. However, they still contain numerous noises and color distortion.

All the above quantitative and visual comparison results demonstrate that considering both reduce the inter-domain gap and the intra-domain gap in our TUDA can produce visually pleasing results and have more robust performance. Due to the limited space, more experimental results are given in the supplementary material.

D. Ablation Studies and Analysis

In this section, 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. In addition, we study the influence of different ratio of λ 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 generalization ability of our RUIQA, we compare it with two state-of-the-art methods: UCIQE and UIQM in terms of two metrics: Spearman Rank Order Correlation Coefficient (SROCC) and Pearson's linear correlation coefficient (PLCC). Table III lists the corresponding comparison results. As shown, our method achieves the best performance, 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 UCIQE and UIQM, showing the superiority of our RUIQA metric. A visual comparison is also shown in Fig. 14. The larger values indicate a better perceptual quality. Obviously, our RUIQA can more accurately reflect the perceptual quality of the image.

In addition, an ablation study is conducted to analyze the contribution of each individual component using the following settings: 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 III, 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,

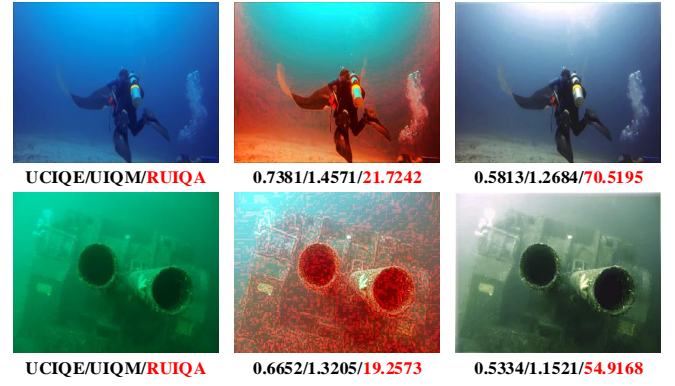


Fig. 14. Visual comparisons in terms of UCIQE, UIQM and our proposed RUIQA metrics. It is obvious that our quantitative scores can better represent subjectively quality.

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

Method	UCIQE	UIQM	UIQA	PUIQA	RUIQA
SROCC \uparrow	0.245	0.400	0.796	0.835	0.900
PLCC \uparrow	0.326	0.422	0.826	0.862	0.904

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

Methods	Test-S1800 \uparrow		Test-R1000 \uparrow	
	PSNR \uparrow	SSIM \uparrow	RUIQA \uparrow	Perceptual Score \uparrow
BL	27.284	0.949	55.534	3.146
BL+ITE	27.065	0.947	59.250	3.390

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

Method	RUIQA \uparrow	Perceptual Score \uparrow
BL + ITE	59.250	3.390
BL + ITE + ITA	61.436	3.595

and then quickly help the IQA task better predict the image quality score.

2) *Effectiveness of the inter-domain adaptation phase*: In this part, 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., dual-alignment network. Results are listed in Table IV. It can be seen that baseline network has only slightly higher PSNR values in comparison to our dual-alignment network (0.219dB higher on average), but the perceptual quality is far worse than the dual-alignment network (3.716 and 0.244 lower on average in RUIQA and perceptual score of user study, respectively). Such results show that our inter-domain adaptation part generates the enhanced results with well reconstructed details (high fidelity) and good perceptual visual quality. Some examples are shown in Fig. 15, in which our inter-domain adaptation phase can better correct color casts and avoid over-

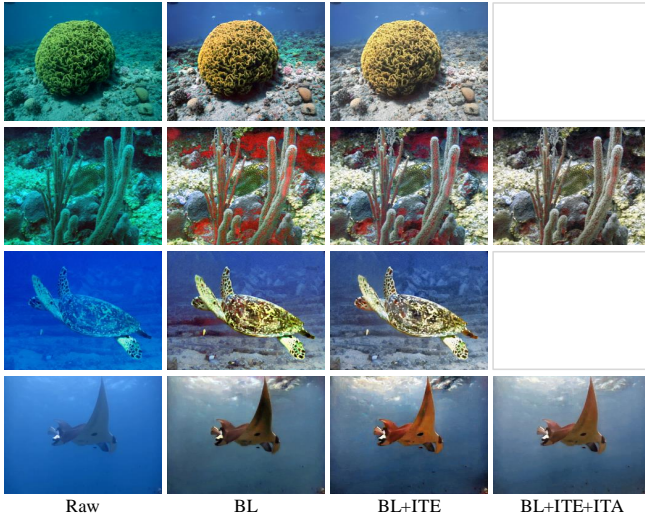


Fig. 15. Visual comparisons on some easy and hard samples. We can clearly see that our method can effectively handle easy and hard samples, especially on hard samples, our full model generates the most visually pleasing results.

TABLE VI
[t]

THE ABLATION STUDY ON HYPER-PARAMETER λ FOR DIVIDING THE REAL UNDERWATER DATA INTO THE EASY AND HARD SAMPLES.

λ	0.40	0.45	0.50	0.55	0.60	0.65
RUIQA \uparrow	62.03	61.39	61.78	61.31	61.20	61.39

TABLE VII
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

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 averaged RUIQA value and the averaged perceptual score are reported in Table V. It can be seen BL+ITE+ITA achieves better performance, even the average performance gains up to 2.186 and 0.205 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 are more subject to human preferences. In addition, a few samples are illustrated in Fig. 15. It can be noted that if only conduct inter-domain adaptation phase, the results of some hard samples still contain some noises and over-enhancement artifacts in some region. In other words, our intra-domain adaptation part is robust for real-world extremely hard underwater image enhancement, producing visually more pleasing results.

4) *Analysis of Hyperparameter λ* : x_n denotes the real underwater images for inter-domain training, in our work,

$n \in (1, 2900)$. 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 the ratio λ 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 the ratio λ will have a significant impact on subsequent operations. Here, some experiments are conducted to decide the optimal λ in our framework. For a selected ratio λ , we first conduct intra-domain adaptation training. Then, the 2900 real training data (**Test-S2900**) is used as validate data in the test pipeline (see Fig. 8). Finally, we predict the average perceptual quality score of 2900 enhanced images in term of the RUIQA metric, and set it as the metric for selecting λ . Results are shown in Table VI, where it can be observed that $\lambda = 0.4$, the proposed method can achieve better performance.

E. Model Complexity Analysis

We compare the flops, parameters and time cost of different 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 950 images and its size is 256x256x3. The source codes and test parameters of all the compared methods are provided by their authors, and the results are presented in Table VII.

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 method. The size, computation and time cost of FUIE-GAN are less than our TUDA. However, the generalization performance on four 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. Such 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 and an intra-domain adaptation phase to jointly optimize the inter-domain gap and the inter-domain gap. Firstly, a dual-alignment network is introduced to jointly perform image-level and feature-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 our TUDA 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," *Computer Graphics and Image Processing*, vol. 6, pp. 184–195, 1975.
- [2] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics Gems*, 1994.
- [3] M. S. Hitam, E. A. Awalludin, W. N. Jawahir Hj Wan Yussof, and Z. Bachok, "Mixture contrast limited adaptive histogram equalization for underwater image enhancement," in *Proc. Int. Conf. Computer Applications Technology (ICCAT)*, pp. 1–5, 2013.
- [4] C. Ancuti, C. O. Ancuti, T. Haber, and P. Bekaert, "Enhancing underwater images and videos by fusion," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 81–88, June 2012.
- [5] 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 Processing (ICIP)*, pp. 4572–4576, June 2014.
- [6] S. Zhang, T. Wang, J. Dong, and H. Yu, "Underwater image enhancement via extended multi-scale retinex," *Neurocomputing*, vol. 245, pp. 1–9, 2017.
- [7] 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, 2018.
- [8] C. Ancuti, C. O. 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.
- [9] J. Chiang and Y. Chen, "Underwater image enhancement by wavelength compensation and dehazing," *IEEE Trans. Image Process.*, vol. 21, pp. 1756–1769, 2012.
- [10] P. Drews, E. R. Nascimento, S. Botelho, and M. Campos, "Underwater depth estimation and image restoration based on single images," *IEEE Comput. Graph. Appl.*, vol. 36, pp. 24–35, Mar. 2016.
- [11] C. Li, J. Guo, R. Cong, Y. Pang, and B. Wang, "Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior," *IEEE Trans. Image Process.*, vol. 25, pp. 5664–5677, 2016.
- [12] Y.-T. Peng and P. Cosman, "Underwater image restoration based on image blurriness and light absorption," *IEEE Trans. Image Process.*, vol. 26, pp. 1579–1594, Apr. 2017.
- [13] 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.*, 2020.
- [14] D. Akkaynak and T. Treibitz, "Sea-thru: A method for removing water from underwater images," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 1682–1691, 2019.
- [15] C. Li, S. Anwar, and F. Porikli, "Underwater scene prior inspired deep underwater image and video enhancement," *Pattern Recognit.*, vol. 98, 2020.
- [16] J. Li, K. Skinner, R. 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, pp. 387–394, 2018.
- [17] C. Fabbri, M. J. Islam, and J. Sattar, "Enhancing underwater imagery using generative adversarial networks," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, pp. 7159–7165, 2018.
- [18] 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, pp. 323–327, 2018.
- [19] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao, "An underwater image enhancement benchmark dataset and beyond," *IEEE Trans. Image Process.*, vol. 29, pp. 4376–4389, 2020.
- [20] X. Chen, J. Yu, S. Kong, Z. Wu, X. Fang, and L. Wen, "Towards real-time advancement of underwater visual quality with gan," *IEEE Trans. Ind. Electron.*, vol. 66, pp. 9350–9359, 2019.
- [21] 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.
- [22] J. Jaffe, "Computer modeling and the design of optimal underwater imaging systems," *IEEE J. Ocean. Eng.*, vol. 15, pp. 101–111, 1990.
- [23] A. Dudhane, P. Hambarde, P. Patil, and S. Murala, "Deep underwater image restoration and beyond," *IEEE Signal Process. Lett.*, vol. 27, pp. 675–679, Apr. 2020.
- [24] Y. Shao, L. Li, W. Ren, C. Gao, and N. Sang, "Domain adaptation for image dehazing," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 2805–2814, 2020.
- [25] 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. Computer Vision and Pattern Recognition (CVPR)*, pp. 3763–3772, 2020.
- [26] I. Shin, S. Woo, F. Pan, and I.-S. Kweon, "Two-phase pseudo label densification for self-training based domain adaptation," *ArXiv*, vol. abs/2012.04828, 2020.
- [27] C. Zheng, T.-J. Cham, and J. Cai, "T2net: Synthetic-to-realistic translation for solving single-image depth estimation tasks," in *Proc. European Conference on Computer Vision (ECCV)*, pp. 798–814, Springer International Publishing, 2018.
- [28] S. Zhao, H. Fu, M. Gong, and D. Tao, "Geometry-aware symmetric domain adaptation for monocular depth estimation," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 9780–9790, 2019.
- [29] M. A. V. D. I. Gulrajani, F. Ahmed and A. Courville, "Improved training of wasserstein gans," in *Neural Information Processing Systems*, p. 5767–5777, Mar. 2017.
- [30] 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, pp. 4861–4875, 2020.
- [31] M. Yang and A. Sowmya, "An underwater color image quality evaluation metric," *IEEE Trans. Image Process.*, vol. 24, pp. 6062–6071, 2015.
- [32] K. Panetta, C. Gao, and S. Agaian, "Human-visual-system-inspired underwater image quality measures," *IEEE J. Ocean. Eng.*, vol. 41, pp. 541–551, July 2016.
- [33] 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, pp. 1221–1234, May 2019.
- [34] S. Su, Q. Yan, Y. Zhu, C. Zhang, X. Ge, J. Sun, and Y. Zhang, "Blindly assess image quality in the wild guided by a self-adaptive hyper network," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 3664–3673, June 2020.
- [35] W. Zhang, Y. Liu, C. Dong, and Y. Qiao, "Ranksrgan: Generative adversarial networks with ranker for image super-resolution," in *Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV)*, pp. 3096–3105, Oct. 2019.
- [36] 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, 2020.
- [37] X. Fu, Z. Fan, M. Ling, Y. Huang, and X. Ding, "Two-step approach for single underwater image enhancement," in *Proc. Int. Symp. Intelligent Signal Processing and Communication Systems (ISPACS)*, pp. 789–794, 2017.
- [38] 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, 2020.
- [39] R. I.-R. BT, "Methodology for the subjective assessment of the quality of television pictures," *International Telecommunication Union*, 2002.
- [40] P. ITU-T RECOMMENDATION, "Subjective video quality assessment methods for multimedia applications," *International telecommunication union*, 1999.
- [41] 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, 2006.
- [42] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, L. Jin, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, et al., "Color image database tid2013: Peculiarities and preliminary results," in *European workshop on visual information processing (EUVIP)*, pp. 106–111, IEEE, 2013.
- [43] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
- [44] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 5967–5976, 2017.
- [45] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE Trans. Image Process.*, vol. 28, pp. 492–505, Jan. 2019.
- [46] 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, pp. 113–120, Jan. 2019.