Physics-guided Paired Underwater Image Synthesis and Underwater Image Degradation Removal Network

Anonymous Authors

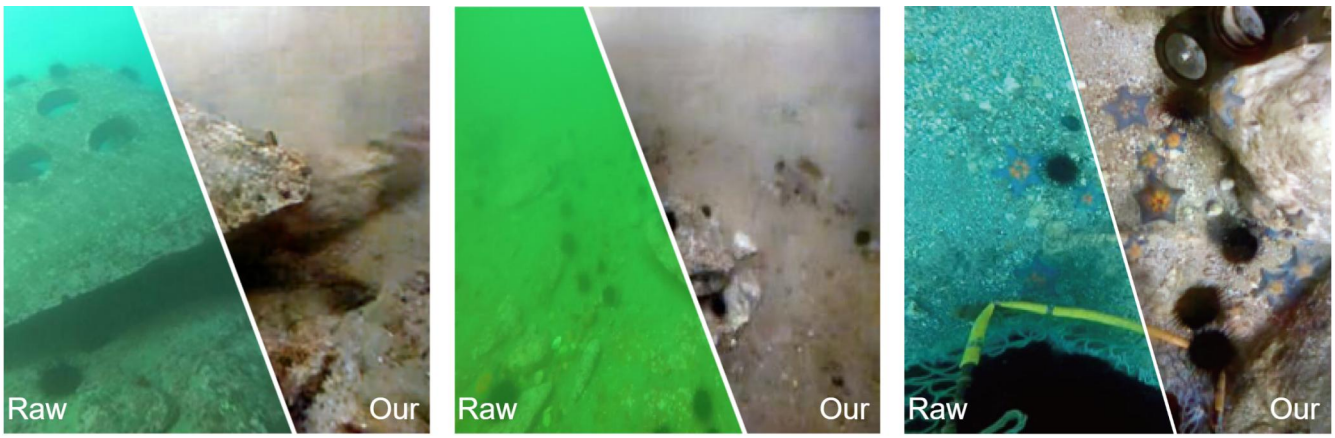

Figure 1: Visual comparisons of raw underwater images and processed by our method

ABSTRACT

Underwater imaging faces inherent challenges such as scattering and differential attenuation, leading to issues like blurring and color cast, which affect vision-based underwater task performance. The effectiveness of deep neural network (DNN) methods in mitigating or removing the influence of underwater environments on images largely depends on the quality of the water-removed images in training datasets. However, the complexity of underwater environments makes obtaining perfectly paired underwater and water-removed images difficult. This work presents Physics-guided Paired Underwater Image Synthesis (PPUIS) to construct paired realistic underwater and water-removed images. PPUIS synthesizes underwater images by randomizing imaging physical model's parameters and combining this with land RGBD images, where the parameter range is matched with the depth information range of the image. We refine the selection of realistic underwater images from the synthesized set by leveraging the color distribution of real underwater images and a discriminator trained to distinguish between real and synthesized underwater images. These land images processed by color constancy method are considered water-removed images. To showcase PPUIS's effectiveness, we propose an Underwater Image Degradation Removal Network (UIDRN). UIDRN's preprocessing layer provides a learnable image of the same size as the input image, aimed at bridging the gap between realistic and real underwater images, while also aiding the adaptation of real underwater images to the network structure. The encoder, based on the Swin Transformer V2, ensures that UIDRN can capture the complex degradation features of underwater images. The decoder incorporates spatial and channel attention gates to achieve higher-quality water-removed image output. We also incorporate multi-scale and multi-category deep supervision losses during training to broaden the scales and dimensions of the training. Experimental results demonstrate that our method has obvious improvement over SOTA methods quantitatively and qualitatively.

CCS CONCEPTS

• **Computing methodologies** → **Reconstruction**.

KEYWORDS

Paired underwater image dataset construction, Underwater Image Enhancement, Underwater Image Degradation Removal

1 INTRODUCTION

Unlike land imaging environments, underwater environments present unique challenges due to wavelength-dependent light absorption and scattering, which degrade the quality of captured underwater images. Specifically, aquatic plankton and suspended particles scatter light, leading to blurring in the captured images. Furthermore, shorter-wavelength blue light attenuates more slowly in water, travelling greater distances, whereas longer-wavelength red light attenuates quickly, travelling shorter distances. This difference results in a prevalent, severe blue or green color cast in underwater images [20][10]. These issues pose significant challenges for vision-based underwater tasks.

ACM MM, 2024, Melbourne, Australia

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Common underwater image enhancement methods aim to improve visibility and quality through techniques such as color correction, contrast enhancement, and noise reduction, making underwater images more visually appealing. Among these methods, removing degradation in underwater images is more complex and challenging, as it often involves accurately simulating and reversing the physical processes in underwater environments, focusing on alleviating or eliminating image degradation effects caused by the unique optical properties of water. The degradations removed from underwater images only eliminate the effects of the underwater environment on imaging, without altering other information of the target scene. This high-fidelity image restoration is particularly crucial for downstream tasks that require detailed image analysis and decision-making, such as 3D reconstruction and biometric identification.

Methods for underwater image degradation removal that rely on DNN necessitate paired datasets. These datasets should include degraded images affected by underwater environments (underwater images), and their corresponding images unaffected by underwater environments (water-removed images). By training on these datasets, the network learns to map from underwater images to water-removed images [28][29]. The success of this learning largely depends on the quality of the datasets, particularly the consistency of factors like target scenes, background, object positions, lighting, and viewing angles between underwater and water-removed images [45] [7]. While capturing real underwater images is relatively straightforward, obtaining perfectly paired water-removed images can be challenging due to the complexity of underwater environments and issues like equipment jitters, water flows, and scene changes. Many methods have been proposed in recent years to build paired underwater image datasets to address this issue. However, these methods still have room for improvement, and we will delve into the reasons for this in detail in Section 2.

In the process of learning to map from underwater images to water-removed images, the network should adequately consider the complexity of light propagation in the underwater environment. Due to the multiple scatterings and reflections of light by substances like plankton and particles, light at different distances interferes with each other. Moreover, different parts of the same large object or complex structures may exhibit significantly different degradation features due to factors like occlusion, geometric shape, material characteristics, or their relative positions to light sources.

Therefore, the network's ability to account for the dependencies among different regions of underwater image will enhance its learning of the correct mapping. Convolution has been proven to effectively capture local features of images, while Transformers, due to their self-attention mechanism, can effectively grasp a broader global context and the long-term dependencies among different regions in underwater images. To this end, we used the Swin Transformer v2 [26] as the encoder of UIDRN to capture the complexity of underwater image degradation. Furthermore, the spatial and channel attention gates [40][14][30] introduced in the UIDRN's decoder can reconstruct the complex color and luminance information of water-removed images. Visual comparisons of underwater images and our method's results are shown in Figure 1. Our contributions can be summarized as follows:

1. We introduce PPUIS, a method that refines land images using color constancy for true color representation and synthesizes underwater images with depth and parameter alignment, further enhanced by leveraging real underwater image color distributions and discrimination to ensure realism.

2. We develop UIDRN, a deep neural network designed for eliminating underwater image degradation, comprising a unique preprocessing layer that bridges the gap between real and realistic underwater images, a Swin Transformer V2-based encoder to capture complex degradation patterns, and an attention-enhanced decoder, all fine-tuned with multi-scale and multi-category deep supervision to precisely remove degradation across various scales and dimensions.

3. Our method surpasses existing methods in correcting underwater image color cast and restoring scene details with state-of-the-art precision, uniquely enhancing details in shadowed regions, correcting overexposure, and preserving real colors under various lighting conditions.

2 RELATED WORK

Traditional underwater image enhancement methods fall into two categories: classical image processing methods and physical imaging model-based methods. Classical image processing methods typically involve direct stretching or redistribution of image pixels in the color space or frequency domain. Techniques include histogram equalization, contrast stretching, color correction, and wavelet transform [11][15][47][43][44], all of which aim to improve the image's contrast, brightness, and saturation. On the other hand, physical imaging model-based methods seek to improve image quality by understanding and modelling the underwater imaging process. The physical underwater imaging process is commonly represented as shown in Equation 1 and Equation 2.

$$I_c = J_c e^{-\beta_c d} + B_c^{\infty}\left(1 - e^{-\beta_c d}\right) \tag{1}$$

Where c $\in$ {R, G, B}. I_c is the underwater image captured by the sensor, B_c^{∞} is the veiling light. J_c is the ideal image of the target scene. $t_c = e^{-\beta_c d}$ is the transmission map, which represents the transmission of light from the target scene to the camera, d is the depth map and represents the distance between the sensor and the target scene, and β_c is the backscatter and wideband attenuation coefficient. When considering the differential attenuation of light in water for each wavelength band, i.e. B_c^{∞} and β_c of RGB channels are different, this imaging model would be used for underwater imaging. This equation is also used for land image dehazing when considering B_c^{∞} and β_c of RGB channels are equal. Equation 1 assumes wideband attenuation and backscatter coefficients are same, but [3][1][2] consider them to be distinct and optimized Equation 1, resulting in the optimized model shown in Equation 2:

$$I_c = J_c e^{-\beta_c^D d} + B_c^{\infty}\left(1 - e^{-\beta_c^B \cdot d}\right) \tag{2}$$

Where $c \in R, G, B$. β_c^{D} is wideband attenuation coefficients, β_c^{B} is backscatter coefficient. I_c, J_c, B_c^{∞} and d have the same definitions as those in Equation 1. [2] also proposed a method for estimating β_c^D and the signal of backscatter $B_c^{\infty}\left(1 - e^{-\beta_c^B \cdot d}\right)$, and archive high-quality underwater image degradation removal for underwater

images. we synthesize underwater images based on Equation 2 in Section 3.1.

Physical imaging model-based methods generally rely on estimating the parameters of the imaging model, i.e. β_c^B, β_c^D, or t_c, to reduce or eliminate the degradation of underwater image [9][5]. Traditional underwater image processing methods often struggle to adapt to various water types and lighting conditions simultaneously, leaving room for improvement in their robustness and generalization.

Compared to traditional underwater image enhancement, DNN-based methods for underwater image processing can learn features from a vast amount of images, effectively handle complex tasks, and exhibit strong robustness and generalization. Convolution-based DNN methods have been proven to effectively capture and process local features of underwater images, and have shown commendable performance in related tasks. Moreover, the Transformer [37] has been gradually introduced into vision tasks due to its superior handling of long-range dependencies and contextual information in images [8][27][26] [32], achieving high-performance metrics in several vision tasks [23][6]. Transformer provides crucial support for learning the complex features involved in interactions between different regions of underwater images.

To achieve effective network training, a high-quality dataset is crucial. We will delve into the various methods for constructing paired datasets for underwater image enhancement. Generally, the construction of paired datasets for underwater image enhancement can be divided into two categories: methods that synthesize ground truth images and methods that synthesize underwater images. Both approaches have their advantages and disadvantages, and each affects subsequent network training and image enhancement outcomes differently. For ground truth image synthesis methods, Li et al. [22] organized 50 participants to evaluate various image enhancement and dehazing techniques. Participants then selected the images that best aligned with human visual preferences to construct the UIEBD dataset. Peng et al. [32] employed a more objective and detailed process than [22] to ensure the highest quality of ground truth images in their LSUI dataset. Both UIEBD and LSUI offer high-quality ground truth images that cater to human visual preferences. However, these methods may neglect certain features of real underwater images, leading to an overemphasis on human visual preferences. Islam et al. [17] introduced a technique using the FUnIE-GAN network to generate ground truth images for the EUVP dataset. Nonetheless, the capabilities of FUnIE-GAN might restrict the quality of the ground truth images. Huang et al. [16] produced ground truth images via a teacher model. However, since the teacher model is trained on datasets provided by [22] and [21], limitations inherent in these datasets could impact the authenticity of the ground truth images generated by the teacher model.

In terms of methods for synthesizing underwater images, Li et al. [21][4] synthesized a large number of paired underwater images by summarizing the t_c across multiple types of water and employing the NYU v2 dataset [34] along with Equation 1, resulting in the construction of the UWCNN dataset. Although this method offers an effective approach, it might compromise the authenticity of the synthesized underwater images by neglecting the depth information between the target scenes and sensors. Hou et al. [13] predicted the parameters of Equation 1 and then synthesized underwater images using outdoor images, forming the SUID dataset. However, this approach is constrained by the precision of the parameter prediction method itself and does not incorporate depth information. Wen et al. [39] predicted underwater image parameters related to Equation 2 using the method provided by [2] and synthesized underwater images employing RGBD images alongside these predicted parameters. Nonetheless, they overlooked the inconsistency between the depth information of the underwater image used for parameter prediction and the depth information of the RGBD image utilized for underwater image synthesis. We refer to the dataset synthesized based on this method as SyreaNet. These methods consider the land images involved in underwater image synthesis as ground truth images, but neglect the fact that these land images might not accurately reflect the true colors of the target scene due to illumination effects. Furthermore, due to the complexity of underwater environments, Equations 1 and 2 might not accurately represent the physical process of underwater imaging. Hence, both the land images used for underwater image synthesis and the synthesized underwater images may require further processing to ensure that the constructed paired datasets are suitable for underwater image degradation removal tasks.

3 METHOD

3.1 PPUIS: Physics-guided Paired Underwater Image Synthesis

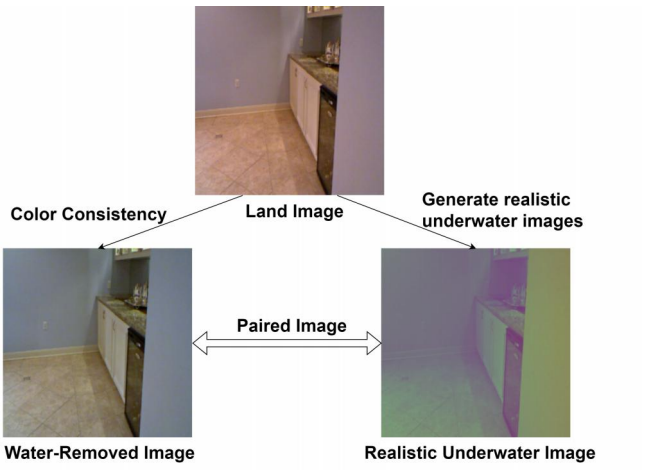

Figure 2: The principle of PPUIS. Land images from RGBD datasets are processed using the color constancy method and considered water-removed images. Other methods of constructing paired images through synthesizing underwater images seem to overlook the issue of color cast in land images caused by illumination. Realistic underwater images are synthesized based on land images according to the steps introduced in Section 3.1.

"Finding Regularities in Constrained Stochastic Properties" inspired our development of PPUIS, a comprehensive method for

constructing paired underwater and water-removed images, as illustrated in Figure 2. To obtain water-removed images that reflect the true colors of the target scene, we processed the land images from the RGBD datasets [34] and [36] using a color constancy method provided by [41], as shown in Figure 2.

To synthesize realistic underwater images, we initially synthesize underwater images based on Equation 2. Specifically, as the depth information of land images in the NYUv2 and DIODE [36] datasets is concentrated within the range of [0,10] meters, we were inspired by [2] to uniformly and randomly select values within the [0,5] range. These values are considered as the wideband attenuation coefficient β_c^{D} and the backscatter coefficient β_c^{B}. We also uniformly and randomly select a value within the [0,1] range, which is regarded as the veiling light in Equation 2. Based on Equation 2, we can synthesize a large number of non-repetitive underwater images. Ideally, if Equation 2 is considered to fully characterize the physical features of underwater imaging, the synthesized underwater images should include real underwater images due to the uniform random value setting of the parameters. This is because the combination of underwater imaging model parameters β_c^{D}, β_c^{B}, B_c^{∞}, and d obtained through random selection should encompass the parameter combinations of real underwater images. Furthermore, most of the synthesized underwater images that meet our criteria are derived from the synthesis of indoor images. This is because outdoor images contain a large number of depth values set to 0, either due to the performance limitations of the depth sensor or specific processing.

Then, we select realistic underwater images from the synthesized underwater images. We train a discriminator based on EfficientNet [35] that can distinguish between real and fake underwater images. This discriminator can score the realism of underwater images within the range of [0, 1], with a score of 1 assigned for real underwater images and a score of 0 for fake underwater images. The real underwater images in the training and validation sets are from [2] and RUIE (Real-world underwater enhancement) dataset provided by [24], while the fake underwater images are synthesized underwater images from UWCNN and SyreaNet. Images in both training and validation sets are selected randomly. The average scores of the discriminator scoring on the validation set are shown in Table 1.

Table 1: The average Scores Of the discriminator for real and fake underwater images of the validation set.

Underwater images	Average Scores
Real ↑	0.7070
Fake ↓	0.4325

Additionally, we conduct a statistical analysis on the real underwater images provided by [24] and [2]. We found that the mean value of the G channel in these images is higher than that of the R and B channels, as depicted in Figure 3. Consequently, we also consider "The mean value of the G channel should be larger than the mean value of the R channel and the B channel" as an essential criterion. Based on the above, the conditions for choosing realistic

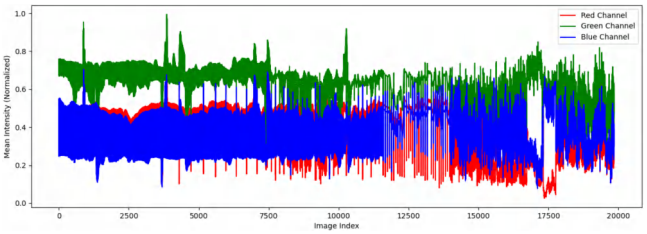

Figure 3: The comparison of the mean value of RGB channel of 20000 images provided by RUIE and [2].

underwater images from synthesized underwater images can be summarized as follows:

1. Discriminator score greater than 0.7. We compared the impact of different discriminator score conditions on the selection of realistic images, as shown in Figure 4.

Figure 4: The impact of the discriminator score on selecting realistic images shows that as the discriminator score value increases, the selected realistic underwater images are more affected by the water environment, and the features of the indoor scenes in the images gradually become blurred.

2. The mean value of the image G channel should be larger than the mean value of the image R channel and the mean value of the image B channel.

$$F(I) = \begin{cases} \text{True Water Image,} & \text{if } D(I) > 0.7 \text{ and } \{\text{mean}(I_G) > \text{mean}(I_R)\} \text{ and } \{\text{mean}(I_G) > \text{mean}(I_B)\} \\ \text{Fake Water Image,} & \text{Otherwise} \end{cases} \quad (3)$$

Where I denotes the synthesized underwater image based on Equation 2. $D(I)$ stands for the discriminator score of the synthesized underwater image. The synthesized underwater images provided by UWCNN and SyreaNet, along with their respective discriminator scores, are displayed. Similarly, the realistic underwater images produced by our PPUIS and their corresponding discriminator scores are also presented. Additionally, the real underwater images sourced from RUIE and their discriminator scores are also exhibited. These images and scores can be found in Figure 5.

Such selection criteria help us filter realistic underwater images that exhibit features similar to real underwater images and meet the scoring requirements of the discriminator. In our method, enhancing the discriminator's performance assists us in accurately

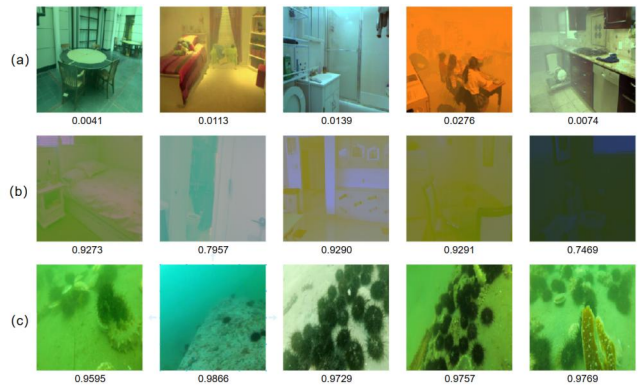

Figure 5: Images and the discriminator's scores for them. In (a) are the synthesized underwater images provided by UWCNN and SyreaNet, and their scores. Their scores are very low, i.e. the probability that the discriminator thinks they are real underwater images is very low. In (c) are the real underwater images provided by RUIE and their scores. Their scores are high, i.e. the probability that the discriminator thinks they are real underwater images is high. In (b) are the realistic underwater images provided PPUIS, and they all meet our choosing conditions.

identifying the differences between synthesized underwater images and real underwater images with more complex features. By improving the discriminator's performance, we can select realistic underwater images from the pool of synthesized images, which have a broader range of features similar to those of real underwater images.

3.2 UIDRN: Underwater Image Degradation Removal Network

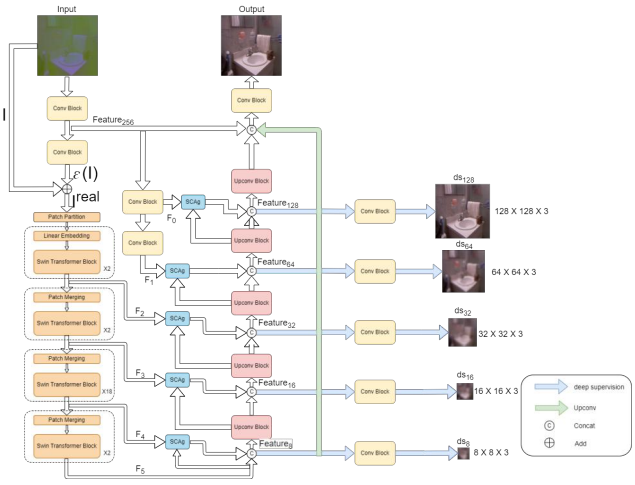

Figure 6: UIDRN: Underwater Images Degradation Removal Network.

In Section 3.1, we get the realistic underwater images, which have characteristics similar to real underwater images. To further bridge the gap between realistic and real underwater images, we designed the image preprocessing layer in UIDRN. UIDRN's architecture is shown in figure 6. The real underwater images I^{real} fed to the encoder are as shown in Equation 4:

$$I^{real} = I + \varepsilon(I) \tag{4}$$

Where I represents realistic underwater images, ε denotes the difference between realistic and real underwater images, and is a learnable image of the same size as the input. $\varepsilon(I) = Conv(Conv(I))$, where $\varepsilon(I)$ is obtained by performing two convolutional processes on I. When the network processes real underwater images, the image preprocessing layer adjusts the feature representation of the real underwater image to more closely match the distribution of the training data, enabling the UIDRN network to more effectively apply the learned features and weights. Simultaneously, we obtain features $F_0 = Conv(Conv(I))$, $F_1 = Conv(Conv(Conv(I)))$, and $Feature_{256} = Conv(I)$, which are used to reconstruct the water-removed image.

Then we feed I^{real} to Swin Transformer v2 to achieve feature extraction as Equation 5.

$$F_2, F_3, F_4, F_5 = \text{SwinTransformer}(I^{real}) \tag{5}$$

And then we have a more fine-grained fusion of image region and channel features based on spatial and channel attention gates(SCAg), as shown in Figure 7:

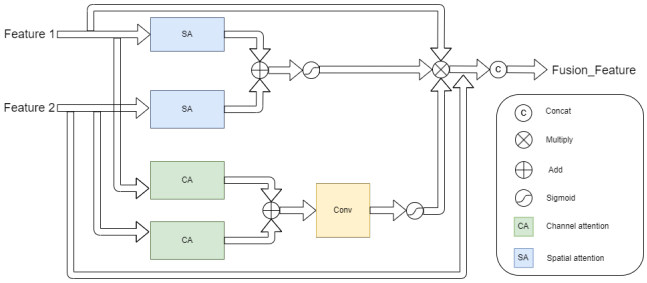

Figure 7: Features fusion based on SCAg.

$$Feature_{2^i} = \begin{cases} \text{Cat}[F_5, SCAg(F_4, F_5)] & \text{if } i = 3 \\ \text{Cat}[Upconv(F_{2^{i-1}}), & \\ SCAg(F_{7-i}, Upconv(F_{2^{i-1}}))] & \text{if } i = 4, 5, 6, 7 \end{cases} \tag{6}$$

Where $Cat()$ represents concatenating features along the channel dimension, $SCAg()$ represents fusing features through SCAg, $Feature_8, Feature_{16}, Feature_{32}, Feature_{64}, Feature_{128}$ respectively represent features at scales of 8×8, 16×16, 32×32, 64×64, 128×128. $Upconv()$ represents upsampling and convolving features. $F_i, i = 0, 1, 2, 3, 4, 5$ are shown in Figure 6.

Finally, we upsample and convolve these features to 256×256 to fuse them, and reconstruct water-removed image $\bar{I}$ based on convolution:

$$\bar{I} = Conv(Cat(Upconv(Feature_{2^i}), Feature_{256})), i = 3, 4, 5, 6, 7 \tag{7}$$

In the process of reconstructing the water-removed image, we also compress the features into images at each scale based on convolution and use these images for deep supervision to accelerate network training:

$$ds_{2^i} = Conv(Feature_{2^i}), i = 3, 4, 5, 6, 7, 8 \tag{8}$$

We use L1 loss, L2 loss, SSIM loss, and Perceptual loss to form the reconstruction loss.

$$loss_{reconstruct} = loss_{ssim}(\bar{I}, \text{gt}) + loss_{perceptual}(\bar{I}, \text{gt}) + |\bar{I} - \text{gt}|_1 + \|\bar{I} - \text{gt}\|^2 \tag{9}$$

Where gt is the ground truth water-removed image. We also introduce deep supervision loss in training. To ensure that UIDRN can effectively reconstruct water-removed images at different scales and accelerate network training, the depth supervision loss consists of L1 loss, L2 loss, and SSIM loss at multiple scales. Specifically, $ds_{reconstruct} = \{ds_8, ds_{16}, ds_{32}, ds_{64}, ds_{128}\}$ represents that output images at different scales 8×8, 16×16, 32×32, 64×64, 128×128 in the process of reconstructing the water-removed image by UIDRN. $ds_{gt} = \{gt_8, gt_{16}, gt_{32}, gt_{64}, gt_{128}\}$ represents the downsampled images of the ground truth water-removed image at the corresponding scale. Calculate the loss of the corresponding scale for the image of the corresponding scale, and the weights of different scales are $weight = \{w_8, w_{16}, w_{32}, w_{64}, w_{128}\} = \{0.1, 0.15, 0.20, 0.25, 0.3\}$.

$$\begin{aligned} loss_{deepsupervision} = & loss_{ssim}\left(ds_{reconstruct}, dswr_{gt}\right) \\ & + \left|ds_{reconstruct} - dswr_{gt}\right|_1 \\ & + \left\|ds_{reconstruct} - dswr_{gt}\right\|^2 \end{aligned} \tag{10}$$

Finally, we optimize UIDRN through the following Equation 11:

$$loss_{total} = loss_{reconstruct} + loss_{deepsupervision} \tag{11}$$

4 EXPERIMENTS

4.1 Implementation Details

Training Details: The discriminator, as mentioned in Section 3.1, is trained using NVIDIA RTX 4090 GPUs with an initial learning rate of 0.0001, batch size set as 10, and the learning rate is halved every 200 epochs. The training set comprises real underwater images, including 19,400 images from RUIE [24] and 600 images from [2], as well as fake underwater images: 10,000 images each from UWCNN and SyreaNet. The validation set consists of real underwater images (5,711 from RUIE and 289 from [2]) and fake underwater images (3,000 each from UWCNN and SyreaNet). Images for both the training and validation sets are selected randomly.

UIDRN is trained on NVIDIA A40 GPU with an initial learning rate of 0.0001, batch size set as 4, and the learning rate is multiplied by 0.5 every 400 epochs. The training set contains 5000 pairs of paired water-removed images and realistic underwater images provided by PPUIS.

Competitors & Metrics: Methods used for experimental comparison include traditional underwater image enhancement methods such as MLLE [44] and WWPF [43], as well as DNN-based methods: Ucolor [20], TOPAL [18], U-shape [32], and Semi-UIR [16]. As discussed in Section 3.1, the realistic underwater images provided by PPUIS have similar features to real underwater images provided by RUIE and [2]. Therefore, UIDRN trained on PPUIS can better mitigate the degradation in real underwater images provided by RUIE and [2]. Therefore, we randomly select 500 real underwater images from RUIE and [2], for **Comparison on Real Underwater Images with Similar Features to Realistic Underwater Images**. To further verify the capabilities of our method for general real underwater images, we also randomly selected 90 images from the UIEB dataset, 329 images from the EUVP dataset, and 400 images from the LSUI dataset for **Comparison on General Real Underwater Images**.

In terms of quantitative experimental comparisons, considering that the PPUIS dataset, like SyreaNet and UWCNN, contains ground truth images are land images rather than high-quality underwater enhanced images, it is inappropriate to use metrics such as UIQM [31] and UCIQE [42] for evaluating the quality of our method's output images. This is because UIQM and UCIQE are specifically designed and optimized for underwater image characteristics, with UCIQE's weighting coefficients derived from training on underwater image datasets. Applying UIQM and UCIQE to evaluate water-removed images may not yield accurate results. Therefore, to quantitatively compare our method with others in terms of removing underwater image degradation, we conducted experimental comparisons on synthesized underwater images and selected PSNR [19] and SSIM [12] as metrics to quantitatively compare the closeness between the processed images by different methods and land images. Among the methods involved in the experimental comparison, training sets of Ucolor and Semi-UIR include images from UWCNN, while UIDRN is trained on images provided by PPUIS. Considering that DNN models tend to perform exceptionally well when tested on their training sets, to mitigate this effect, we randomly selected 400 images each from SyreaNet, UWCNN, and PPUIS, and mixed them for use in the **Comparison on Synthesized Underwater Images**.

4.2 Experiment Details

Comparison on Real Underwater Images with Similar Features to Realistic Underwater Images: It is noteworthy that, as shown by the areas within the red boxes in the second rows of images in Figure 8, image processed by our method does not exhibit blue or green color casts in regions that are far from the camera. Furthermore, in areas closer to the camera, images processed by our method have restored more details on the stone surfaces, indicating that our method can address the issue of details being overly dark due to shadow occlusion. The comparison of images in the first and second rows also demonstrates that our method can effectively remove the blue or green color cast and blur in underwater images.

Comparison on General Real Underwater Images: To demonstrate the generalization and robustness of our method, we also compared it with other methods on general real underwater images, as shown in Figure 9. In all comparisons, our method effectively

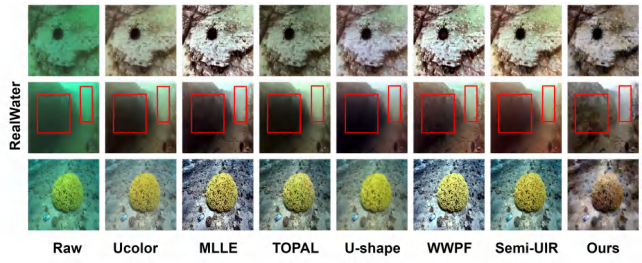

Figure 8: Comparison on real underwater images with similar features to realistic underwater images. The blue or green color cast and blur in the images have been significantly corrected.

corrected the blue or green color cast in the images. In the second and third rows, images processed by our method not only removed the blue or green color cast but also preserved the true colors of the clownfish within the red-boxed area. In the image comparison in the seventh row, only the image processed by our method resolved the overexposure problem in the target scene caused by lighting. In the comparison in the ninth row, the original image exhibited red areas on the fish's body due to lighting issues. Only the image processed by our method showed colors on the fish's body that appeared more normal and closer to natural colors. This is attributed to the fact that the water-removed images in the PPUIS dataset are processed using the color constancy method, and UIDRN has learned the ability to restore the true colors of underwater environments and objects.

Comparison on Synthesized Underwater Images: To quantitatively compare the closeness between the reconstructed images by different methods and land images, we also conducted comparisons on synthesized underwater images. The qualitative comparison results are shown in Figure 10, and the quantitative comparison results are presented in Table 2. The most noticeable observation from the comparison in Figure 10 is that the images processed by our method do not exhibit any color cast, and from the comparison within the red box area in the images of the third row, only our method restored the reflection of a person in the mirror. This further demonstrates the generalization ability and robustness of our method in correcting color cast and blur in underwater images. The comparison of PSNR and SSIM in Table 2 indicates that, compared to other methods, the images processed by our method have significantly higher metric values, suggesting that the synthesized underwater images processed by our method are the closest to their corresponding land images.

We condact **ablation studies** to separately verify the advancements of PPUIS and UIDRN. We trained Unet [33], CycleGAN [46], and U-shape on images provided by PPUIS. We then quantitatively compared the closeness of the images processed by each method to land images. The metircs results are shown in Table 3.

Based on the experimental comparisons in Table 3, among the methods trained on the same PPUIS dataset, the UIDRN achieves the highest PSNR and SSIM values. This indicates that among the models trained on the same dataset, the images processed by UIDRN are closest to the land images.

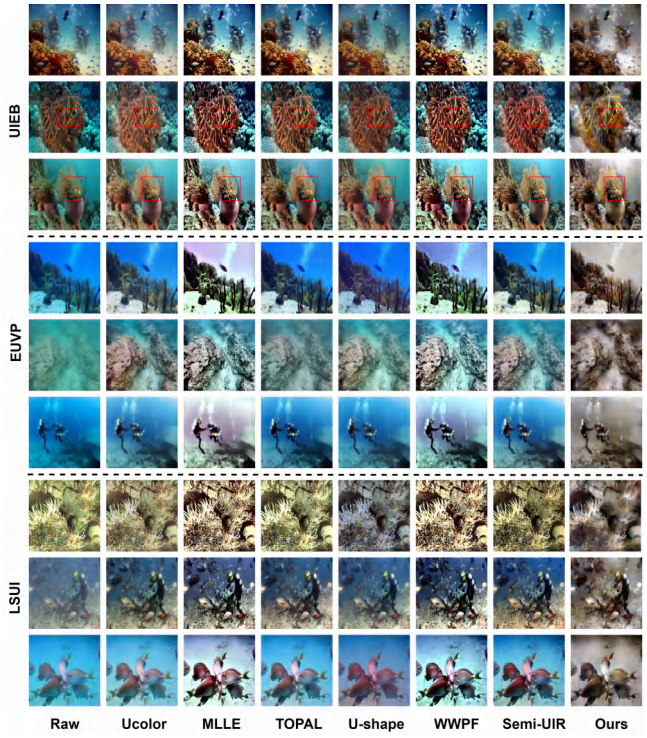

Figure 9: Comparison on general real underwater images.

In the experiments shown in Table 2, the U-shape was trained on images provided by UIEB, LSUI, and UWCNN. In the experiments displayed in Table 3, the U-shape was trained on images provided by PPUIS. A clear comparison shows that the U-shape trained on the PPUIS has **its PSNR increased from the original 15.7217 to 18.7085, an improvement of 19.00%. Its SSIM increased from the original 0.3611 to 0.6248, an improvement of 73.03%.** Such experimental comparisons prove that for the task of removing underwater image degradation, the data provided by PPUIS is superior to that provided by other methods. The comparison in Table 3 also demonstrates that among the networks trained on the same training set, UIDRN outperforms Unet, U-shape, and Cycle-GAN in performing the task of removing underwater image degradation. We are actively reaching out to other researchers, hoping to have more DNN models retrained on images provided by PPUIS to evaluate UIDRN and PPUIS more objectively and comprehensively.

5 CONCLUSION

In this work, we introduced PPUIS and UIDRN. Based on PPUIS, we synthesized 5000 pairs of images that can be used to train DNN models to remove degradation from underwater images. Our proposed UIDRN model, due to its special structural design, is more suited for the task of removing underwater image degradation, and experiments have qualitatively and quantitatively proven that our method has obvious improvement over SOTA methods. It's worth noting that the purpose of our work is solely to remove degradation from underwater images, aiming to preserve other information, including noise, as much as possible. This design is

Table 2: PSNR and SSIM of comparison experiments in synthesized underwater images.

	Ucolor (2021)	MLLE (2022)	TOPAL (2022)	U-shape (2023)	WWPF (2023)	Semi-UIR (2023)	Ours
PSNR ↑	17.6527	14.7977	15.8665	15.7217	14.6133	18.3730	**26.8264**
SSIM ↑	0.5190	0.4623	0.4215	0.3611	0.4660	0.5594	**0.7105**

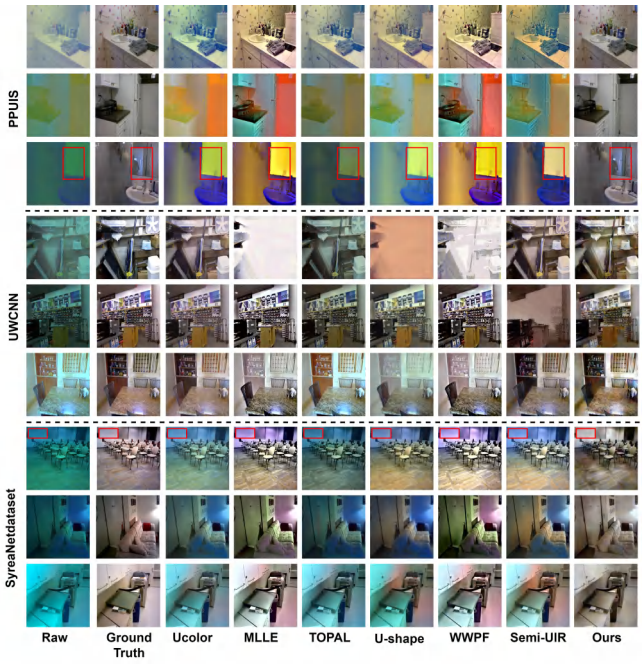

Figure 10: Comparison on synthesized underwater images. The comparison of images indicates that the images processed by our method do not exhibit color cast and blur.

Table 3: PSNR and SSIM of Ablation Studies.

	Unet	U-shape	Cycle-GAN	Ours
PSNR ↑	16.7796	18.7085	18.0037	**26.8264**
SSIM ↑	0.5367	0.6248	0.5034	**0.7105**

motivated by two reasons: 1. We hope that this design can provide relatively complete and scientific image and video information for extending land-based visual tasks to the underwater environment. 2. We aim to explore an underwater image enhancement method that can effectively assist advanced tasks such as underwater target detection and semantic segmentation. Through a large number of experiments, [38] has been proven that existing underwater image enhancement methods might suppress the performance of underwater target detection due to the potential increase in background interference [25]. We believe that ensuring the information of underwater target scenes is not altered by enhancement is crucial for high-level vision tasks, hence we undertook this work. We will also publish our findings on "How can underwater image enhancement effectively assist high-level underwater vision tasks?" in the future.

REFERENCES

[1] Derya Akkaynak and Tali Treibitz. 2018. A revised underwater image formation model. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 6723–6732.

[2] Derya Akkaynak and Tali Treibitz. 2019. Sea-thru: A method for removing water from underwater images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 1682–1691.

[3] Derya Akkaynak, Tali Treibitz, Tom Shlesinger, Yossi Loya, Raz Tamir, and David Iluz. 2017. What is the space of attenuation coefficients in underwater computer vision?. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4931–4940.

[4] Saeed Anwar and Chongyi Li. 2020. Diving deeper into underwater image enhancement: A survey. *Signal Processing: Image Communication* 89 (2020), 115978.

[5] Dana Berman, Deborah Levy, Shai Avidan, and Tali Treibitz. 2020. Underwater single image color restoration using haze-lines and a new quantitative dataset. *IEEE transactions on pattern analysis and machine intelligence* 43, 8 (2020), 2822–2837.

[6] Hanting Chen, Yunhe Wang, Tianyu Guo, Chang Xu, Yiping Deng, Zhenhua Liu, Siwei Ma, Chunjing Xu, Chao Xu, and Wen Gao. 2021. Pre-trained image processing transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 12299–12310.

[7] Dandan Ding, Shiwei Gan, Long Chen, and Ben Wang. 2023. Learning-based underwater image enhancement: An efficient two-stream approach. *Displays* 76 (2023), 102337.

[8] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020).

[9] Paul Drews, Erickson Nascimento, Filipe Moraes, Silvia Botelho, and Mario Campos. 2013. Transmission estimation in underwater single images. In *Proceedings of the IEEE international conference on computer vision workshops*. 825–830.

[10] Xianping Fu, Zheng Liang, Xueyan Ding, Xinyue Yu, and Yafei Wang. 2020. Image descattering and absorption compensation in underwater polarimetric imaging. *Optics and Lasers in Engineering* 132 (2020), 106115.

[11] Ahmad Shahrizan Abdul Ghani and Nor Ashidi Mat Isa. 2015. Underwater image quality enhancement through integrated color model with Rayleigh distribution. *Applied soft computing* 27 (2015), 219–230.

[12] Alain Hore and Djemel Ziou. 2010. Image quality metrics: PSNR vs. SSIM. In *2010 20th international conference on pattern recognition*. IEEE, 2366–2369.

[13] Guojia Hou, Xin Zhao, Zhenkuan Pan, Huan Yang, Lu Tan, and Jingming Li. 2020. Benchmarking underwater image enhancement and restoration, and beyond. *IEEE Access* 8 (2020), 122078–122091.

[14] Jie Hu, Li Shen, and Gang Sun. 2018. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 7132–7141.

[15] Dongmei Huang, Yan Wang, Wei Song, Jean Sequeira, and Sébastien Mavromatis. 2018. Shallow-water image enhancement using relative global histogram stretching based on adaptive parameter acquisition. In *MultiMedia Modeling: 24th International Conference, MMM 2018, Bangkok, Thailand, February 5-7, 2018, Proceedings, Part I 24*. Springer, 453–465.

[16] Shirui Huang, Keyan Wang, Huan Liu, Jun Chen, and Yunsong Li. 2023. Contrastive semi-supervised learning for underwater image restoration via reliable bank. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 18145–18155.

[17] Md Jahidul Islam, Youya Xia, and Junaed Sattar. 2020. Fast underwater image enhancement for improved visual perception. *IEEE Robotics and Automation Letters* 5, 2 (2020), 3227–3234.

[18] Zhiying Jiang, Zhuoxiao Li, Shuzhou Yang, Xin Fan, and Risheng Liu. 2022. Target oriented perceptual adversarial fusion network for underwater image enhancement. *IEEE Transactions on Circuits and Systems for Video Technology* 32,

10 (2022), 6584–6598.
[19] Jari Korhonen and Junyong You. 2012. Peak signal-to-noise ratio revisited: Is simple beautiful?. In *2012 Fourth International Workshop on Quality of Multimedia Experience*. IEEE, 37–38.
[20] Chongyi Li, Saeed Anwar, Junhui Hou, Runmin Cong, Chunle Guo, and Wenqi Ren. 2021. Underwater image enhancement via medium transmission-guided multi-color space embedding. *IEEE Transactions on Image Processing* 30 (2021), 4985–5000.
[21] Chongyi Li, Saeed Anwar, and Fatih Porikli. 2020. Underwater scene prior inspired deep underwater image and video enhancement. *Pattern Recognition* 98 (2020), 107038.
[22] Chongyi Li, Chunle Guo, Wenqi Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao. 2019. An underwater image enhancement benchmark dataset and beyond. *IEEE Transactions on Image Processing* 29 (2019), 4376–4389.
[23] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. 2021. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*. 1833–1844.
[24] Risheng Liu, Xin Fan, Ming Zhu, Minjun Hou, and Zhongxuan Luo. 2020. Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light. *IEEE transactions on circuits and systems for video technology* 30, 12 (2020), 4861–4875.
[25] Risheng Liu, Zhiying Jiang, Shuzhou Yang, and Xin Fan. 2022. Twin adversarial contrastive learning for underwater image enhancement and beyond. *IEEE Transactions on Image Processing* 31 (2022), 4922–4936.
[26] Ze Liu, Han Hu, Yutong Lin, Zhuliang Yao, Zhenda Xie, Yixuan Wei, Jia Ning, Yue Cao, Zheng Zhang, Li Dong, et al. 2022. Swin transformer v2: Scaling up capacity and resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 12009–12019.
[27] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*. 10012–10022.
[28] Pan Mu, Haotian Qian, and Cong Bai. 2022. Structure-inferred bi-level model for underwater image enhancement. In *Proceedings of the 30th ACM International Conference on Multimedia*. 2286–2295.
[29] Pan Mu, Hanning Xu, Zheyuan Liu, Zheng Wang, Sixian Chan, and Cong Bai. 2023. A generalized physical-knowledge-guided dynamic model for underwater image enhancement. In *Proceedings of the 31st ACM International Conference on Multimedia*. 7111–7120.
[30] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. 2018. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999* (2018).
[31] Karen Panetta, Chen Gao, and Sos Agaian. 2015. Human-visual-system-inspired underwater image quality measures. *IEEE Journal of Oceanic Engineering* 41, 3 (2015), 541–551.
[32] Lintao Peng, Chunli Zhu, and Liheng Bian. 2023. U-shape transformer for underwater image enhancement. *IEEE Transactions on Image Processing* (2023).
[33] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer, 234–241.
[34] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. 2012. Indoor segmentation and support inference from rgbd images. In *Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part V 12*. Springer, 746–760.
[35] Mingxing Tan and Quoc Le. 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*. PMLR, 6105–6114.
[36] Igor Vasiljevic, Nick Kolkin, Shanyi Zhang, Ruotian Luo, Haochen Wang, Falcon Z. Dai, Andrea F. Daniele, Mohammadreza Mostajabi, Steven Basart, Matthew R. Walter, and Gregory Shakhnarovich. 2019. DIODE: A Dense Indoor and Outdoor DEpth Dataset. *CoRR* abs/1908.00463 (2019). http://arxiv.org/abs/1908.00463
[37] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).
[38] Yudong Wang, Jichang Guo, Wanru He, Huan Gao, Huihui Yue, Zenan Zhang, and Chongyi Li. 2023. Is Underwater Image Enhancement All Object Detectors Need? *IEEE Journal of Oceanic Engineering* (2023).
[39] Junjie Wen, Jinqiang Cui, Zhenjun Zhao, Ruixin Yan, Zhi Gao, Lihua Dou, and Ben M Chen. 2023. SyreaNet: A Physically Guided Underwater Image Enhancement Framework Integrating Synthetic and Real Images. *arXiv preprint arXiv:2302.08269* (2023).
[40] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. 2018. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*. 3–19.
[41] Kai-Fu Yang, Shao-Bing Gao, and Yong-Jie Li. 2015. Efficient illuminant estimation for color constancy using grey pixels. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2254–2263.
[42] Miao Yang and Arcot Sowmya. 2015. An underwater color image quality evaluation metric. *IEEE Transactions on Image Processing* 24, 12 (2015), 6062–6071.
[43] Weidong Zhang, Ling Zhou, Peixian Zhuang, Guohou Li, Xipeng Pan, Wenyi Zhao, and Chongyi Li. 2023. Underwater image enhancement via weighted wavelet visual perception fusion. *IEEE Transactions on Circuits and Systems for Video Technology* (2023).
[44] Weidong Zhang, Peixian Zhuang, Hai-Han Sun, Guohou Li, Sam Kwong, and Chongyi Li. 2022. Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement. *IEEE Transactions on Image Processing* 31 (2022), 3997–4010.
[45] Zengxi Zhang, Zhiying Jiang, Jinyuan Liu, Xin Fan, and Risheng Liu. 2023. Waterflow: heuristic normalizing flow for underwater image enhancement and beyond. In *Proceedings of the 31st ACM International Conference on Multimedia*. 7314–7323.
[46] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*. 2223–2232.
[47] Peixian Zhuang, Jiamin Wu, Fatih Porikli, and Chongyi Li. 2022. Underwater image enhancement with hyper-laplacian reflectance priors. *IEEE Transactions on Image Processing* 31 (2022), 5442–5455.