Unsupervised Underwater Image Restoration: From a Homology Perspective

Zhenqi Fu[1][*], Huangxing Lin[1][*], Yan Yang[2], Shu Chai[1], Liyan Sun[1], Yue Huang[1], Xinghao Ding[1][†]

[1] School of Informatics, Xiamen University, China
[2] College of Computer Science & Technology, Hangzhou Dianzi University, China
{fuzhenqi, hxlin}@stu.xmu.edu.cn, yangyanyy@hdu.edu.cn, {shuchai, sunly}@stu.xmu.edu.cn,
{yhuang2010, dxh}@xmu.edu.cn

Abstract

Underwater images suffer from degradation due to light scattering and absorption. It remains challenging to restore such degraded images using deep neural networks since real-world paired data is scarcely available while synthetic paired data cannot approximate real-world data perfectly. In this paper, we propose an UnSupervised Underwater Image Restoration method (USUIR) by leveraging the homology property between a raw underwater image and a re-degraded image. Specifically, USUIR first estimates three latent components of the raw underwater image, i.e., the global background light, the transmission map, and the scene radiance (the clean image). Then, a re-degraded image is generated by randomly mixing up the estimated scene radiance and the raw underwater image. We demonstrate that imposing a homology constraint between the raw underwater image and the re-degraded image is equivalent to minimizing the restoration error and hence can be used for the unsupervised restoration. Extensive experiments show that USUIR achieves promising performance in both inference time and restoration quality.

Introduction

Deep learning has achieved great success in various low-level computer vision tasks. However, such success relies heavily on manually labeled datasets, which are time-consuming and expensive to obtain (Liu et al. 2020; Li et al. 2019a). For example, it is hard even impossible to collect a large-scale dataset with the clean ground truth image for underwater image restoration.

Underwater images suffer from color distortion and poor visibility (Akkaynak et al. 2017) because the light is absorbed and scattered when it propagates through turbid water mediums. Such degradations will lead to poor performance in underwater computer vision applications, e.g., underwater object tracking (Panetta et al. 2021), marine animal detection (Fan et al. 2020) and robotic navigation. The degradations in underwater images can be described via the modified Koschmieder's light scattering model (Jaffe 1990) where the transmission map is changed into a channel-wise component and the atmospheric light is replaced by the global background light.

Plenty of techniques have been proposed for underwater image restoration. Most of the approaches either provide handcrafted priors (Li et al. 2016) or deep learning based solutions for the restoration (Li et al. 2021b). Most of the deep learning based methods rely on synthetic training data due to the absence of large-scale real-world underwater image datasets with ground-truths. Data synthesis may not capture the complex features of the real-world degradation, and thus, suffer from domain-shift problems (Ben-David et al. 2007). In addition, a few deep learning based solutions belong to weakly supervised approaches that use generative adversarial networks (Li, Guo, and Guo 2018).

Recently, "zero-shot" approaches (Li et al. 2021a; Kar et al. 2021; Gandelsman, Shocher, and Irani 2019; Shocher, Cohen, and Irani 2018) have been developed which learn the restoration task using a single image and train a small image-specific network at test time. As such "zero-shot" methods do not use other supervision information beyond the input image itself, they require to leverage self/unsupervised losses or regularizers for training. For example, the unsupervised single image restoration (Kar et al. 2021) estimates the parameters of Koschmieder's model through a controlled perturbation. The unsupervised image decomposition method (Gandelsman, Shocher, and Irani 2019) applies multiple "Deep-image-Prior" (DIP) networks (Ulyanov, Vedaldi, and Lempitsky 2018) and the power of the internal patch recurrence to separate images into their basic components. Although "zero-shot" approaches need no extra training examples other than the input image, they are probably inefficient to be applied for real-world applications due to a large number of optimization iterations at test time.

In this paper, we propose an UnSupervised Underwater Image Restoration method (USUIR), a novel deep neural network trained for real-world and real-time underwater image restoration. Different from the "zero-shot" approaches, USUIR applies a set of training instances to train the deep neural network. To this end, we design a disentanglement network with the homology constraint. Concretely, in each iteration, we first disentangle the raw underwater image into three latent components, i.e., the global background light, the transmission map, and the scene radiance (the clean image). Then, we randomly mix up the estimated scene radiance and the raw underwater image to generate a re-

---

[*]These authors contributed equally.

[†]Corresponding author.

degraded image. We demonstrate that performing a homology constraint between the raw underwater image and the re-degraded image is equivalent to minimizing the restoration error. This finding motivates us to design an unsupervised underwater image restoration method from the perspective of homology. Note that the homology of underwater images defined in this paper should satisfy the two constraints: 1) the underwater images share the same underlying clean image; 2) the underwater images share the same formation model. USUIR is optimized in an unsupervised manner, and it is expected to learn the underwater degradation model using deep neural networks. Extensive experiments on two real-world underwater image datasets demonstrate the effectiveness and efficiency of the proposed method. To summarize, our work contributes in the following ways:

- We propose an unsupervised neural network (i.e., USUIR) for real-world and real-time underwater image restoration [1].

- USUIR performs layer disentanglement using three joint sub-modules. Among them, two of which are convolutional neural networks to estimate the clean image and the transmission map, another one is a Gaussian blur module to calculate the global background light.

- To achieve the unsupervised restoration, we design an underwater image re-degradation scheme and a homology constraint between the raw underwater image and the re-degraded image.

- Despite being an unsupervised approach, USUIR is comparable with the state-of-the-art methods in terms of the image quality and the inference time.

Related Work

In addition to traditional model-free methods, e.g., histogram equalization, automatic white balance, retinex-based method (Fu et al. 2014), and fusion-based approaches (Gao et al. 2019; Ancuti et al. 2012), the existing underwater image restoration techniques can be roughly classified into two groups: prior-based methods and learning-based methods.

Prior-Based Methods

Prior-based methods are widely used in the underwater image restoration community. They estimate the parameters of the underwater imaging model based on prior assumptions. For example, the work (Peng and Cosman 2017) proposed an underwater image depth estimation method based on light absorption and image blurriness. The clean image is restored via the underwater imaging model with the estimated depth map. The work (Peng, Cao, and Cosman 2018) proposed a generalization of the dark channel prior for image restoration under extreme imaging conditions. The work (Akkaynak and Treibitz 2018) proposed a new underwater formation model in which the coefficients of backscatter are different than those of direct transmission. Based on this model, the work (Akkaynak and Treibitz 2019) proposed a new underwater image restoration method using RGBD images.

Accurately estimating the underwater imaging parameters is challenging in current prior-based methods, especially for diverse and complex underwater scenes. As a result, prior-based methods are sensitive to different conditions of underwater images (Liu et al. 2020). Incorrect parameters estimation of the formation model tends to produce either over/under-restored outputs or introduce severe color distortions.

Learning-Based Methods

Motivated by the success of recent deep neural networks, there are several attempts made to improve the performance of underwater image restoration through a deep learning strategy. For example, the pioneering work (Li et al. 2018) employed the underwater image model and generative adversarial networks to generate synthetic image pairs for supervised training. The work (Li, Anwar, and Porikli 2020) trained ten underwater image restoration models for different types of underwater images. The work (Li et al. 2019b) collected a real-world underwater image dataset for training deep networks and proposed a gated fusion network for the restoration. The work (Uplavikar, Wu, and Wang 2019) applied adversarial learning to extract domain agnostic features of multiple water types and generated the clean image from those features. The work (Li et al. 2021b) proposed an underwater image restoration network via medium transmission guided multi-color space embedding. Currently, a few "zero-shot" approaches (Kar et al. 2021; Gandelsman, Shocher, and Irani 2019) have been proposed, which perform the restoration using a single image.

However, most of the learning-based methods rely on synthetic images to learn a deep neural network, which may suffer from domain shift issues. On the other hand, although "zero-shot" approaches leverage the input image itself to train a restoration model, they are probably time-consuming, prohibiting their applications in practice.

Method

Given a set of underwater images, we aim to learn the mapping from the degraded image to the clean image in an unsupervised manner. The basic idea of our approach is to disentangle each image into three latent components based on Koschmieder's model and leverage the homology constraint between the raw underwater image and a re-degraded image. As shown in Figure 1, our approach simultaneously feeds the raw underwater image into a scene radiance estimation network (J-Net), a transmission map estimation network (T-Net), and a global background light estimation module (GB). To supervise the disentangling process, we combine the three components to reconstruct the input image through Koschmieder's model. After that, a re-degraded image is generated by randomly mixing up the estimated scene radiance and the raw underwater image. We perform the homology constraint between the re-degraded image and the raw underwater image. In addition, we also leverage an unsupervised color loss to correct the potential color deviations of the restored image. Note that the whole model is learned in a self-supervised manner and all sub-modules are optimized in an end-to-end pipeline.

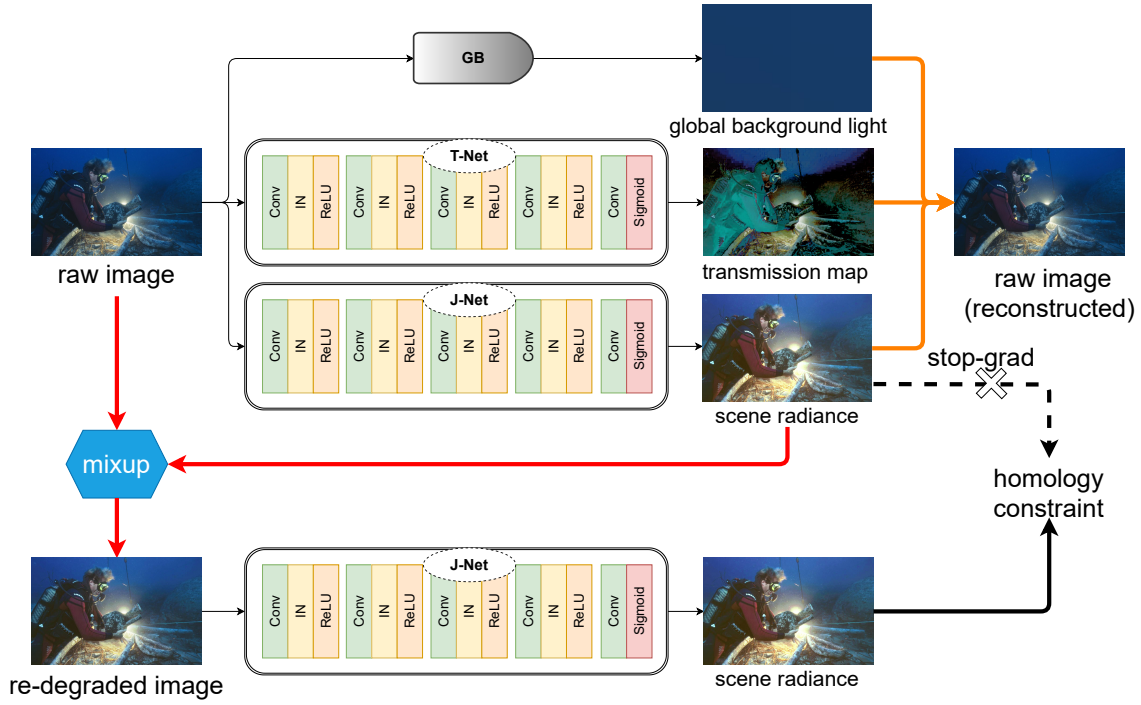[1]Our code is available at: https://github.com/zhenqifu/USUIR

Figure 1: The framework of USUIR, which consists of three joint sub-modules, i.e., the scene radiance estimation network (J-Net), the transmission map estimation network (T-Net), and the global background light estimation module (GB). Taking a raw underwater image as the input, these three sub-modules disentangle the input into three basic components, i.e., the global background light, the transmission map, and the scene radiance. USUIR generates a re-degraded image by randomly mixing up the raw underwater image and the estimated scene radiance. USUIR performs the homology constraint between the raw image and the re-degraded image. A reconstruction loss and a Gray-world assumption based color loss are applied at the top layer of USUIR to supervise the layer decomposition process and correct potential color deviations, respectively. Note that the two J-Nets share parameters, and a stop-gradient operation is applied on the raw image side when calculating the homology loss.

Underwater Image Degradation Model

USUIR is based on the modified Koschmieder's light scattering model that describes the degradations of underwater images, where the transmission map is modified into a channel-wise component and the atmospheric light is replaced by the global background light (Li et al. 2021a). As the Beer-Lambert law, light propagation is associated with an attenuation factor $e^{-\beta d}$ where d is the distance from the source and β is the channel-wise extinction coefficient depending on the water quality. Guided by the Beer-Lambert law, in the water medium, the effect of light scattering at a distance d from the source can be formulated as:

$$S(d) = \left(1 - e^{-\beta d(x)}\right) A \qquad (1)$$

where A represents the global background light and the attenuated quantity $S(d)$ is called the veiling light or space-light. Similarly, according to the existing notion (Narasimhan and Nayar 2003; Berman et al. 2021), the scene radiance J also gets attenuated by the factor $e^{-\beta d}$. Thus, we get:

$$D(x) = J(x) e^{-\beta d(x)} \qquad (2)$$

Based on the above effects of light scattering, the degradation of underwater images can be formulated as the additive combination of $S(d)$ and $D(x)$ as follows:

$$I(x) = J(x) e^{-\beta d(x)} + \left(1 - e^{-\beta d(x)}\right) A \qquad (3)$$

where I represents the observed underwater image, J is the scene radiance, x is an image pixel, and $d(x)$ is the scene depth at that pixel. Equation (3) is known as Koschmieder's light scattering model that explains underwater image formation. In (3), $t(x) = e^{-\beta d(x)}$ is known as the transmission value. For underwater light propagation, the transmission value $t(x)$ of the degraded image is different for each primary color channel due to the wavelength selective characteristics of the attenuation.

Exploiting Homology for Underwater Image Restoration

Supposing I_1 and I_2 are two degraded underwater images, as mentioned before, if I_1 and I_2 are homologous, they need to satisfy: 1) I_1 and I_2 share the same clean image; 2) I_1 and I_2 share the same degradation formulation.

Consider the underwater degradation model in (3), we rewrite here as follows for simplicity:

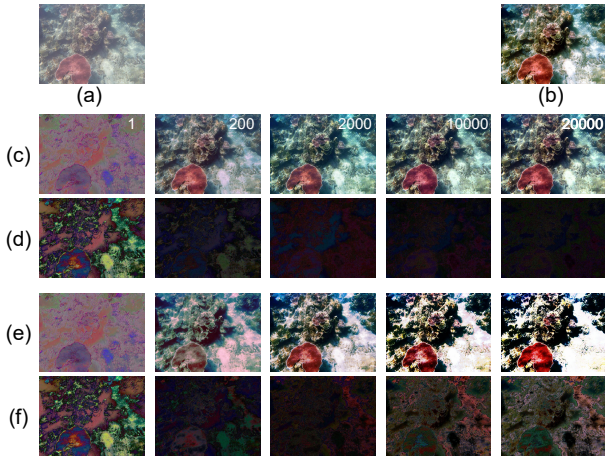$$I_1 = J t_1 + (1 - t_1) A_1 \qquad (4)$$

Figure 2: An example of the restoration error in different iterations. The numbers in the images indicate the training steps. (a) the raw image; (b) the reference image; (c) scene radiance with homology constraint; (d) the error map of (c); (e) scene radiance without homology constraint; (f) the error map of (e).

where I_1 represents the raw underwater image, J refers to the scene radiance. Our objective is to restore the scene radiance J from the degraded observation I_1.

Let $\hat{J} = J + \varepsilon$ be the estimation of i-th iteration of (4), ε denotes the error map. After obtaining $\hat{J}$, we generate a re-degraded image I_2 by simply mixing up the raw underwater image I_1 and the estimated scene radiance $\hat{J}$, which can be formulated as below:

$$I_2 = \alpha I_1 + (1 - \alpha)(J + \varepsilon) \tag{5}$$

where α is a random mixing ratio, $\alpha \sim \text{Beta}(1,1)$. Substituting I_1 from (4) in (5), we get:

$$I_2 = J(\alpha t_1 + 1 - \alpha) + \alpha(1 - t_1)A_1 + (1 - \alpha)\varepsilon \tag{6}$$

Let $t_2 = \alpha t_1 + 1 - \alpha$, equation (6) becomes:

$$I_2 = J t_2 + (1 - t_2)A_1 + (1 - \alpha)\varepsilon \tag{7}$$

Comparing (7) with (4), one can find that, when ε tends to zero, (7) becomes Koschmieder's model, which shows the formation of the re-degraded image I_2 from the scene radiance J, similar to (4) which explains the formation of I_1 from J. Specifically, when ε is a small value, I_1 and I_2 are approximately homologous. I_1 and I_2 are perfectly homologous when $\varepsilon = 0$. Therefore, it is reasonable to perform the homology constraint to regularize the estimated scene radiance of (4) and (7). In Figure 2, we exhibit the error map ε in different iterations using the proposed homology constraint. As can be seen, compared with (e), the error map of (c) gradually decreases as the iteration increases, which intuitively demonstrates the validity of our homology constraint.

The above analyses suggest that performing a homology constraint between the raw underwater image and the generated re-degraded image enables minimizing the restoration error. This property of re-degradation through Koschmieder's model can be leveraged for unsupervised training.

Network Architecture

Figure 1 illustrates the network architecture of the proposed USUIR which contains three sub-modules, i.e., J-Net, T-Net, and GB.

J-Net J-Net aims to estimate the scene radiance from the raw underwater image. As illustrated in Figure 1, J-Net is very simple, which only consists of five convolutional layers, four instance normalization layers, and four ReLU activation layers. In the output layer, J-Net employs the sigmoid function to normalize the output into [0, 1]. J-Net takes a non-degenerate network architecture, i.e., J-Net does not use down-sampling operations, thus can prevent the loss of details in the restored image. In J-Net, the kernel size of all convolutional layers is 3×3, and the channel size of all convolutional layers is 64. More details about the implementation of J-Net could refer to the experimental settings.

T-Net Since the scene radiance and the transmission map are dependent on the input image, USUIR adopts a similar network structure for J-Net and T-Net. The output layer of T-Net is of three channels. The reason is that light with different wavelengths is subjected to varying degrees of attenuation when it propagates through the water medium. Due to the wavelength selective characteristics of the attenuation, the transmission value is different for different primary color channels. This is a primary cause of color distortion. Note that there is no explicit loss for T-Net. The optimization of T-Net is based on a reconstruction loss at the top layer of USUIR, which will be described in the next section.

GB Gaussian Blur (GB) module aims to estimate the global background light from the input image. Different from the scene radiance and the transmission map, the global background light is independent of the image content and reflects the global property. Therefore, we perform the Gaussian blur to remove content details and obtain the global attribute. In our implementation, the filter size is $\left[\frac{1}{4}(w+h)\right] \times 2 + 1$, where $[*]$ denotes the nearest integer less than or equal to $*$, w and h denote the width and height of the input image, respectively. Note that GB enforces the global background light to be smooth, and therefore, the high-frequency details will be enforced on the output of J-Net and T-Net.

Loss Functions

Since USUIR is based on the image decomposition and the homology constraint, a self-supervised reconstruction loss and a homology loss are employed in this paper. The former enforces the latent components could reconstruct the input image, while the latter performs the homology constraint on input images. In addition, USUIR adopts an unsupervised color loss to correct the potential color distortion of the restored image. The adopted loss functions are detailed below.

Homology Loss The homology loss $\mathcal{L}_H$ is one of the primary losses developed in this paper. To supervise J-Net, we design a re-degradation scheme that randomly mixes up the raw underwater image and the estimated scene radiance. We demonstrate that performing a homology constraint between the raw underwater image and the re-degraded image

is equivalent to minimizing the restoration error. Considering the definition of homology is from two aspects, the constraint of homology should also have two folds.

Since we have demonstrated that the re-degraded image has a similar formation with the raw underwater image in (7), it is easy to implement the constraint of sharing the same formation model. Instead of explicitly imposing a loss, we adopt the natural capability of the Siamese network. As illustrated in Figure 1, the two J-Nets are required to share parameters, which enforces the two degradation models (we focus on J-Net) are thoroughly the same.

Then, the constraint can be converted into that we only need to ensure the two outputs of J-Nets are the same. Therefore, we compute the homology loss as follows:

$$\mathcal{L}_H = \|J_2 - stopgrad(J_1)\|_2^2 \tag{8}$$

where J_1 and J_2 refer to the clean images estimated from the raw underwater image and the re-degraded image, respectively. Note that a stop-gradient operation ($stopgrad$) is performed on J_1, which makes the training stable.

Reconstruction Loss   Reconstruction loss is another basic loss function used in USUIR to supervise the layer decomposition process. USUIR simultaneously feeds the raw underwater image into a scene radiance estimation network (J-Net), a transmission map estimation network (T-Net), and a global background light estimation module (GB). Naturally, the outputs of them can be further combined to reconstruct the input image at the top layer of USUIR through the formation model. Formally, we calculate the reconstruction loss $\mathcal{L}_R$ as follows:

$$\mathcal{L}_R = \|I(x) - x\|_2^2 \tag{9}$$

where x denotes the raw underwater image, and $I(x)$ denotes the reconstructed image according to Koschmieder's model in (3).

Color Loss   Following the Gray-world assumption of natural image statistics, we design a color loss to correct the potential color deviations in the restored image. The color loss $\mathcal{L}_C$ used in this paper is expressed as:

$$\mathcal{L}_C = \sum_{c \in \Omega} \|\mu(J^c) - 0.5\|_2^2 \tag{10}$$

where $\Omega = \{R, G, B\}$ is a set of color channels with Red (R), Green (G), and Blue (B) colors. J refers to the restored image. $\mu(J^c)$ represents the mean of the color channel c. $\mathcal{L}_C$ penalizes deviation from the Gray-world assumption.

Total Loss   The total loss of our USUIR is expressed as:

$$\mathcal{L} = \omega_1 \mathcal{L}_H + \omega_2 \mathcal{L}_R + \omega_3 \mathcal{L}_C \tag{11}$$

where ω_1, ω_2, and ω_3 are the weights. We set $\omega_1 = 1$, $\omega_2 = 1$, and $\omega_3 = 0.01$ empirically.

Experiments

In this section, we evaluate our method on two real-world underwater image restoration datasets. We compare our method with six baseline methods in terms of full-reference and no-reference quality assessment metrics. In the following contents, we will first introduce the experimental settings, and then show the qualitative and quantitative results on the datasets. Furthermore, we conduct experiments to investigate the inference time of USUIR with baseline methods. The ablation study is also presented to verify the effectiveness of our method. Finally, we extend the USUIR for image dehazing.

Experimental Settings

In this part, we introduce the used datasets, the baselines, the evaluation metrics, and the implementation details.

Datasets   We conduct experiments on two underwater image restoration datasets, i.e., UIEBD (Li et al. 2019b) and RUIE (Liu et al. 2020). UIEBD dataset consists of 890 real-world underwater images with corresponding reference images. Note that the reference image in UIEBD is obtained from 12 existing underwater image restoration methods. Specifically, UIEBD conducts subjective comparisons to select the best result among 12 restored images as the potential reference image. RUIE is a large-scale underwater benchmark that contains three subsets, i.e., underwater image quality subset (3630 images), underwater color cast subset (300 images), and underwater higher-level task-driven subset (300 images). Different from UIEBD that contains reference images, RUIE only contains raw underwater images. In this paper, we choose the underwater image quality subset for training and testing since it contains various underwater conditions and different levels of image quality. As the training and testing images are not clearly assigned in UIEBD and RUIE, we divide UIEBD and RUIE as follows: 1) we apply the first 700 images for training and the rest 190 images for testing on the UIEBD dataset; 2) the underwater image quality subset of RUIE is originally divided into five parts evenly (each part has 726 images) according to the image quality. Therefore, we use the first 600 images of each part for training and the rest for testing.

Baselines   We compare the proposed USUIR with six methods which are divided into three groups, namely, two prior-based methods (IBLA (Peng and Cosman 2017) and Histogram-Prior (Li et al. 2016)), two fully-supervised methods (Water-Net (Li et al. 2019b) and Ucolor (Li et al. 2021b)), and two "zero-shot" methods (DIP (Ulyanov, Vedaldi, and Lempitsky 2018) and DDIP (Gandelsman, Shocher, and Irani 2019)). It should be noted that both the prior-based and "zero-shot" methods restore underwater images without using the ground-truth clean image, and their major difference is that the former is hand-crafted while the latter is based on deep neural networks.

Evaluation Metrics   Since UIEBD consists of ground truth images, the evaluation can be conducted in a standard way. Similar to (Guo et al. 2020) and (Wang et al. 2021), two popular full-reference image quality assessment metrics (i.e., PSNR and SSIM) are used in the quantitative comparisons. Since RUIE only contains raw underwater images, we use two no-reference underwater image quality assessment metrics (i.e., UIQM (Panetta, Gao, and Agaian 2015) and

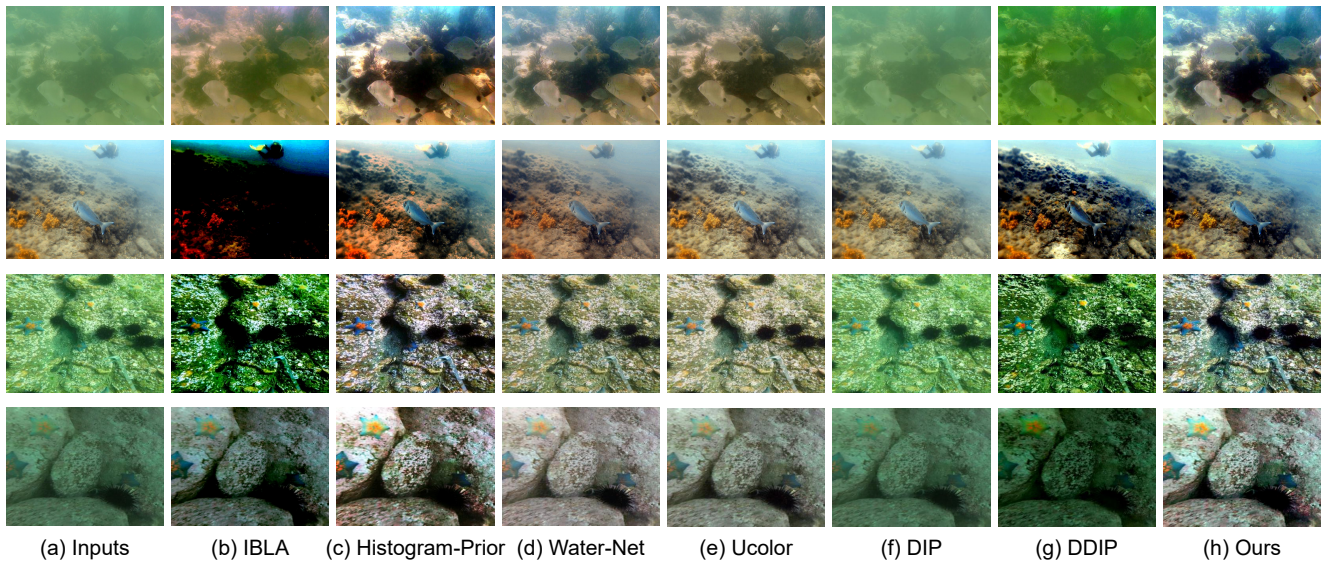| (a) Inputs | (b) IBLA | (c) Histogram-Prior | (d) Water-Net | (e) Ucolor | (f) DIP | (g) DDIP | (h) Ours |

Figure 3: Visual comparisons of different restoration approaches on real-world underwater images, from which we can observe that our method obviously improves the contrast and restores the colors.

UCIQE (Yang and Sowmya 2015)) to objectively measure the quality of restored images.

Experimental Configurations We conduct experiments on an NVIDIA RTX 2080Ti GPU in PyTorch. We employ the ADAM optimizer to optimize USUIR, the default learning rate is $1e^{-4}$, the batch size is 1, and the maximal epoch is 50. We augment the training data with rotation, flipping horizontally and vertically. The training images are resized to 128×128. We record the results of all baselines by using the source codes provided by the authors and adopting the original parameter settings for fair comparisons.

Visual and Perceptual Comparisons

Figure 3 presents the visual results on UIEBD and RUIE datasets. As can be seen, our approach outperforms the competitors in terms of naturalness preservation, contrast and brightness improvement, and color cast correction. IBLA and Histogram-Prior produce over/under-restoration artifacts because they employ handcrafted priors to estimate parameters of the underwater formation model, their predictions may be dominated by the characteristics of the priors. "zero-shot" approaches tend to produce color deviation and incorrect contrast. USUIR yields natural colors and clear details and achieves comparable even better visual quality compared with supervised methods.

Quantitative Comparisons

Table 1 presents the quantitative results on UIEBD and RUIE datasets according to full-reference and no-reference quality assessment metrics. Higher values of these metrics indicate better performance. As can be seen, prior-based methods depend on the characteristics of the priors, their performance is not stable. USUIR performs better or as well as the supervised approaches, despite that it is an unsuper-
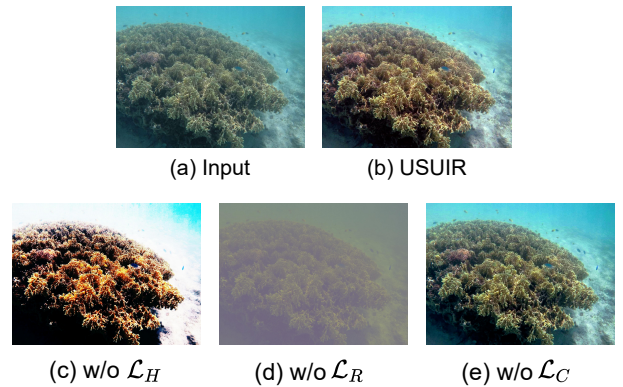

| (a) Input | (b) USUIR |

| (c) w/o $\mathcal{L}_H$ | (d) w/o $\mathcal{L}_R$ | (e) w/o $\mathcal{L}_C$ |

Figure 4: Ablation study of the contribution of each loss (homology loss $\mathcal{L}_H$, reconstruction loss $\mathcal{L}_R$, color loss $\mathcal{L}_C$).

vised method. Compared with the "zero-shot" approaches, USUIR enables leveraging more information from massive training data with the well-designed unsupervised pipeline. As a result, USUIR obtains better performance.

Comparisons on Inference Time

In this subsection, we conduct experiments to show the computational efficiency of USUIR. We measure the inference time of different methods averaged on 100 images of size 512×512 on an NVIDIA RTX 2080Ti GPU and Intel(R) Xeon(R) E5-2678 v3 @ 2.50GHz CPU. Table 2 shows the experimental results. For IBLA and Histogram-Prior, only the codes of the CPU version are available. In Table 2, "zero-shot" approaches are time-consuming because they need to train the network at test time. USUIR is computationally efficient, benefited from the lightweight network structure with the homology constraint.

Datasets	Metrics	Techniques						
		Prior-based		Supervised		"Zero-shot"		Unsupervised
		IBLA	Histogram-Prior	Water-Net	Ucolor	DIP	DDIP	Ours
UIEBD	PSNR	14.39	18.51	19.31	**21.65**	17.08	12.42	<u>20.31</u>
	SSIM	0.573	0.762	0.830	<u>0.840</u>	0.619	0.381	**0.841**
RUIE	UIQM	1.249	**4.210**	3.562	3.622	1.840	2.172	<u>3.788</u>
	UCIQE	24.66	**33.74**	28.42	28.95	19.68	23.44	<u>32.01</u>

Table 1: Quantitative comparisons on UIEBD and RUIE datasets in terms of objective quality assessment metrics. The best result is marked in bold while the second best one is underlined.

Methods		IBLA	Histogram-Prior	Water-Net	Ucolor	DIP	DDIP	Ours
Time(s)	CPU	23.33	**0.87**	2.34	21.19	>5k	>5k	<u>1.07</u>
	GPU	–	–	<u>0.49</u>	3.36	1005	827	**0.004**

Table 2: Inference time comparisons (in second). The best result is marked in bold while the second best one is underlined.

$\mathcal{L}_H$	$\mathcal{L}_R$	$\mathcal{L}_C$	PSNR/SSIM
	✓	✓	12.46/0.521
✓		✓	12.57/0.457
✓	✓		18.40/0.810
✓	✓	✓	20.31/0.841

Table 3: Ablation study on UIEBD dataset.

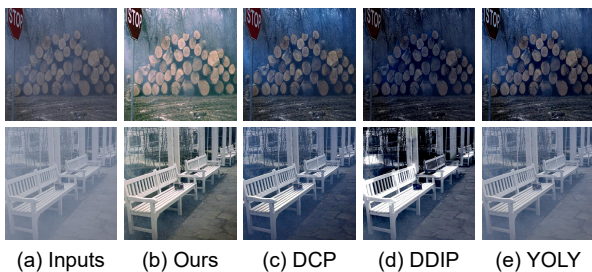(a) Inputs   (b) Ours   (c) DCP   (d) DDIP   (e) YOLY

Figure 5: Visual results of real-world hazing images.

Ablation Study

To verify the effectiveness of our loss functions, we conduct an ablation study on the UIEBD dataset by using different combinations of loss functions. The quantitative results are shown in Table 3 and the visual results are presented in Figure 4, one could observe that: 1) the homology loss $\mathcal{L}_H$ and the reconstruction loss $\mathcal{L}_R$ mainly drive our unsupervised framework. Removing either $\mathcal{L}_H$ or $\mathcal{L}_R$, the restoration performance drops significantly; 2) the color loss $\mathcal{L}_C$ is not necessary for stably training the network, although it could improve the performance to a certain extent; 3) each loss has its respective role in restoring the underwater image, and the combination of all losses achieves the best performance.

Image Dehazing

Similar to underwater images, haze in an image can be described by Koschmieder's model, where the pixel-wise transmission map is the same for all primary color channels. In this subsection, we further perform image dehazing using the proposed unsupervised framework. In the experiment, we simply replace the three-channel transmission map component of the model in (3) with a one-channel map for image dehazing. Other parameter settings and the network structure are the same. The network is trained on real-world dehazing datasets (i.e., I-Haze (Ancuti et al. 2018a) and O-Haze (Ancuti et al. 2018b)). We randomly select 80% of images for training and the rest for testing. We show the visual comparisons with three typical image dehazing methods (i.e., DCP (He, Sun, and Tang 2010), DDIP (Gandelsman, Shocher, and Irani 2019), and YOLY (Li et al. 2021a)) in Figure 5. As can be seen, our method presents a favorite recovery with natural colors and contrasts, which demonstrates the effectiveness of our unsupervised framework. Note that USUIR is not optimized for image dehazing currently, since that we leverage a simple GB module to calculate the atmospheric light. We assume that a more accurate estimation of atmospheric light will promote the dehazing performance.

Conclusion

We proposed USUIR, an unsupervised deep neural network for underwater image restoration from a homology perspective. USUIR is simple yet effective. The basic idea is to disentangle the underwater image into three latent components and leverage the homology constraint between the raw underwater image and the re-degraded image. Experiments demonstrate the superiority of USUIR against existing methods. The potential usage of our framework for image dehazing is also demonstrated. The success of our unsupervised framework shows that the homology property can be leveraged for unsupervised training through a suitable re-degradation operation. In future works, we will try to investigate and design similar frameworks for other applications, where such a homology constraint can be employed.

Acknowledgements

References

Akkaynak, D.; and Treibitz, T. 2018. A Revised Underwater Image Formation Model. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6723–6732.

Akkaynak, D.; and Treibitz, T. 2019. Sea-Thru: A Method for Removing Water From Underwater Images. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1682–1691.

Akkaynak, D.; Treibitz, T.; Shlesinger, T.; Loya, Y.; Tamir, R.; and Iluz, D. 2017. What is the Space of Attenuation Coefficients in Underwater Computer Vision? In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 568–577.

Ancuti, C.; Ancuti, C. O.; Haber, T.; and Bekaert, P. 2012. Enhancing underwater images and videos by fusion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 81–88.

Ancuti, C.; Ancuti, C. O.; Timofte, R.; and De Vleeschouwer, C. 2018a. I-HAZE: a dehazing benchmark with real hazy and haze-free indoor images. In *International Conference on Advanced Concepts for Intelligent Vision Systems*, 620–631. Springer.

Ancuti, C. O.; Ancuti, C.; Timofte, R.; and De Vleeschouwer, C. 2018b. O-haze: a dehazing benchmark with real hazy and haze-free outdoor images. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 754–762.

Ben-David, S.; Blitzer, J.; Crammer, K.; Pereira, F.; et al. 2007. Analysis of representations for domain adaptation. *Advances in neural information processing systems*, 19: 137.

Berman, D.; Levy, D.; Avidan, S.; and Treibitz, T. 2021. Underwater Single Image Color Restoration Using Haze-Lines and a New Quantitative Dataset. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(8): 2822–2837.

Fan, B.; Chen, W.; Cong, Y.; and Tian, J. 2020. Dual refinement underwater object detection network. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XX 16*, 275–291. Springer.

Fu, X.; Zhuang, P.; Huang, Y.; Liao, Y.; Zhang, X.; and Ding, X. 2014. A retinex-based enhancing approach for single underwater image. In *2014 IEEE International Conference on Image Processing (ICIP)*, 4572–4576.

Gandelsman, Y.; Shocher, A.; and Irani, M. 2019. "Double-DIP": Unsupervised Image Decomposition via Coupled Deep-Image-Priors. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 11018–11027.

Gao, S.-B.; Zhang, M.; Zhao, Q.; Zhang, X.-S.; and Li, Y.-J. 2019. Underwater Image Enhancement Using Adaptive Retinal Mechanisms. *IEEE Transactions on Image Processing*, 28(11): 5580–5595.

Guo, C.; Li, C.; Guo, J.; Loy, C. C.; Hou, J.; Kwong, S.; and Cong, R. 2020. Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1777–1786.

He, K.; Sun, J.; and Tang, X. 2010. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12): 2341–2353.

Jaffe, J. 1990. Computer modeling and the design of optimal underwater imaging systems. *IEEE Journal of Oceanic Engineering*, 15(2): 101–111.

Kar, A.; Dhara, S. K.; Sen, D.; and Biswas, P. K. 2021. Zero-shot Single Image Restoration through Controlled Perturbation of Koschmieder's Model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 16205–16215.

Li, B.; Gou, Y.; Gu, S.; Liu, J. Z.; Zhou, J. T.; and Peng, X. 2021a. You only look yourself: Unsupervised and untrained single image dehazing neural network. *International Journal of Computer Vision*, 129(5): 1754–1767.

Li, B.; Ren, W.; Fu, D.; Tao, D.; Feng, D.; Zeng, W.; and Wang, Z. 2019a. Benchmarking Single-Image Dehazing and Beyond. *IEEE Transactions on Image Processing*, 28(1): 492–505.

Li, C.; Anwar, S.; Hou, J.; Cong, R.; Guo, C.; and Ren, W. 2021b. Underwater Image Enhancement via Medium Transmission-Guided Multi-Color Space Embedding. *IEEE Transactions on Image Processing*, 30: 4985–5000.

Li, C.; Anwar, S.; and Porikli, F. 2020. Underwater scene prior inspired deep underwater image and video enhancement. *Pattern Recognition*, 98: 107038.

Li, C.; Guo, C.; Ren, W.; Cong, R.; Hou, J.; Kwong, S.; and Tao, D. 2019b. An underwater image enhancement benchmark dataset and beyond. *IEEE Transactions on Image Processing*, 29: 4376–4389.

Li, C.; Guo, J.; Cong, R.; Pang, Y.; and Wang, B. 2016. Underwater Image Enhancement by Dehazing With Minimum Information Loss and Histogram Distribution Prior. *IEEE Transactions on Image Processing*, 25(12): 5664–5677.

Li, C.; Guo, J.; and Guo, C. 2018. Emerging From Water: Underwater Image Color Correction Based on Weakly Supervised Color Transfer. *IEEE Signal Processing Letters*, 25(3): 323–327.

Li, J.; Skinner, K. A.; Eustice, R. M.; and Johnson-Roberson, M. 2018. WaterGAN: Unsupervised Generative Network to Enable Real-Time Color Correction of Monocular Underwater Images. *IEEE Robotics and Automation Letters*, 3(1): 387–394.

Liu, R.; Fan, X.; Zhu, M.; Hou, M.; and Luo, Z. 2020. Real-world Underwater Enhancement: Challenges, Benchmarks, and Solutions under Natural Light. *IEEE Transactions on Circuits and Systems for Video Technology*, 1–1.

Narasimhan, S.; and Nayar, S. 2003. Contrast restoration of weather degraded images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(6): 713–724.

Panetta, K.; Gao, C.; and Agaian, S. 2015. Human-visual-system-inspired underwater image quality measures. *IEEE Journal of Oceanic Engineering*, 41(3): 541–551.

Panetta, K.; Kezebou, L.; Oludare, V.; and Agaian, S. 2021. Comprehensive Underwater Object Tracking Benchmark Dataset and Underwater Image Enhancement With GAN. *IEEE Journal of Oceanic Engineering*, 1–17.

Peng, Y.-T.; Cao, K.; and Cosman, P. C. 2018. Generalization of the Dark Channel Prior for Single Image Restoration. *IEEE Transactions on Image Processing*, 27(6): 2856–2868.

Peng, Y.-T.; and Cosman, P. C. 2017. Underwater Image Restoration Based on Image Blurriness and Light Absorption. *IEEE Transactions on Image Processing*, 26(4): 1579–1594.

Shocher, A.; Cohen, N.; and Irani, M. 2018. Zero-Shot Super-Resolution Using Deep Internal Learning. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3118–3126.

Ulyanov, D.; Vedaldi, A.; and Lempitsky, V. 2018. Deep Image Prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Uplavikar, P. M.; Wu, Z.; and Wang, Z. 2019. All-in-One Underwater Image Enhancement Using Domain-Adversarial Learning. In *CVPR Workshops*, 1–8.

Wang, L.; Wang, Y.; Dong, X.; Xu, Q.; Yang, J.; An, W.; and Guo, Y. 2021. Unsupervised Degradation Representation Learning for Blind Super-Resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10581–10590.

Yang, M.; and Sowmya, A. 2015. An underwater color image quality evaluation metric. *IEEE Transactions on Image Processing*, 24(12): 6062–6071.