Re-Degradation and Contrastive Learning for Zero-shot Underwater Image Restoration

Nisha Varghese
nishavarghese15@gmail.com
A. N. Rajagopalan
raju@ee.iitm.ac.in

Indian Institute of Technology, Madras
Chennai
India

Abstract

Restoring underwater (UW) images is an important task in ocean exploration applications and is quite challenging due to its fundamental ill-posedness. Traditional methods for UW image restoration struggle when there is a mismatch between the adopted prior and actual scene conditions. Deep models require large-scale paired or unpaired real-world data for training which are scarce in the UW scenario; synthetic datasets typically suffer from domain-shift issues. To alleviate the limitations of prior-based and data-driven UW restoration methods, "zero-shot" approach is an attractive solution. In this paper, we propose an UnderWater Zero-shot image Restoration method (UWZR) by harnessing the physical model for UW image formation. A re-degradation strategy is introduced to generate another UW image that respects the same image formation model. The network is optimized to disentangle the input UW image in such a manner that the relationships between the components of the input UW image and the re-degraded image are satisfied. A contrastive learning strategy is added that ensures that the restored image is pulled closer to a clean image and pushed far away from the UW image in the representation space. Extensive experiments on four real UW datasets establish the superiority of our proposed UWZR over prior art for UW image restoration.

1 Introduction

Underwater (UW) images often suffer from low contrast and contain inaccurate colors due to wavelength-dependent light scattering and absorption in water. In ocean exploration applications [34, 35, 47], UW image restoration is an important task. Reconstruction of colors in UW images is challenging and has attracted a lot of attention. UW image restoration methods can be roughly divided into traditional and deep-learning (DL) based approaches.

Traditional methods can be classified as model-free color correction methods and model-based restoration methods. Model-free methods [11, 15, 22, 48] modify each image pixel to improve the visual quality of UW images. Since they ignore the UW image formation model, they introduce over-enhancement and over-saturation. Physical model-based enhancement methods [6, 37, 38, 42, 49] estimate the parameters of the UW imaging model using image priors. Due to the mismatch between the adopted prior and actual image conditions, the results are not always satisfactory [33].

DL-based approaches have emerged to mitigate the dependence on predetermined priors. Because of the scarcity of large-scale real-paired or unpaired UW datasets, designing supervised [13, 30, 31] or unsupervised Generative Adversarial Network (GAN)-based

[17, 18, 28, 32] networks is not reliable. In order to alleviate the need for large-scale datasets for training data-driven methods, "zero-shot" models can be a solution that utilize only the information from the observed single UW image. A number of zero-shot approaches have been developed for single-image restoration [5, 14, 23, 26, 39, 44]. Zero-shot methods for UW image restoration [5, 23, 39] estimate the parameters of the Koschmieder's model [10] by disentangling the input UW image. But, their performance is not up to the mark.

In this paper, we propose an UnderWater Zero-shot image Restoration method (UWZR) based on the revised UW image formation model [1] that does not employ any hand-crafted priors for optimization unlike the zero-shot methods [26, 39, 44]. We introduce a new effective re-degradation strategy to generate a secondary UW image from the input UW image and we theoretically derive the relation between the latent components of the two UW images. Using a DL network, we disentangle both the UW images into their components and utilize their relationships to optimize the network parameters. In order to improve the performance of our zero-shot network, inspired by the contrastive learning strategy for image dehazing [45] and UW image restoration [20], we introduce a contrastive learning scheme that aims to pull the restored output towards a clean image and push it away from the distribution of the UW image. We perform extensive experiments on several real-world UW image datasets to establish the effectiveness of UWZR for UW image restoration.

Our main contributions are as follows.

1. We propose an UnderWater Zero-shot image Restoration method (UWZR) which does not use any hand-crafted priors. An effective and novel re-degradation strategy is proposed to generate a UW image from the input UW image, and we theoretically show the relationships between their latent components which paves the way for our zero-shot approach.
2. A contrastive learning scheme is also introduced to improve the restoration quality. To the best of our knowledge, UWZR is the first work that introduces a contrastive learning scheme along with zero-shot learning for UW image restoration.
3. Despite being a zero-shot approach, our approach outperforms the state-of-the-art (SoTA) in real-world underwater image restoration.

2 Related Works

2.1 Traditional UW image restoration methods

Traditional model-free UW image restoration methods [3, 11, 15, 22, 48] do not consider the underlying image formation model for enhancement. They simply modify each pixel of the UW image to obtain enhanced images with improved visual quality and hence can end up generating over-saturated or over-enhanced images. Physical model-based enhancement methods utilize the UW image formation model for estimating the parameters with the help of hand-crafted priors in their optimization. Different modifications of dark channel prior (DCP) [21] have been applied for UW image restoration [6, 9, 38]. An adaptive attenuation-curve prior is proposed by Chau *et al*. [42] while Li *et al*. [27] propose a histogram distribution prior. Berman *et al*. [4] consider different types of water and distinct spectral profiles for each type to refine the restored images. Akkayanak *et al*. [2] rely on the depth map for restoration.

2.2 Data-driven UW image restoration methods

Due to the scarcity of real paired UW datasets, UW image restoration using supervised DL-based methods is challenging. [30] created a paired real underwater dataset (UIEB) where

pseudo-ground-truth is subjectively selected based on human perception of the outputs of different enhancement techniques. Supervised networks [13, 31, 43] have used the UIEB dataset [30] for supervision. To partially overcome the challenge of paired datasets, GAN-based methods [17, 18, 28, 32] have emerged. UWCNN [29] trains 10 image enhancement models corresponding to each water type. Utilizing the attenuation coefficient as a cue, [7] proposes a generative network to restore UW images. The work [8] utilizes depth as a cue for UW image restoration where depth is estimated from a pre-trained SoTA model [16]. The unsupervised methods [12, 41] perform physics-based disentanglement of underwater images where [12] uses a homology constraint on the enhanced image and [41] utilizes the view-synthesis constraint on neighboring frames in monocular videos for self-supervision.

2.3 Zero-shot image restoration methods

Zero-shot approach for image restoration outputs the restored image by learning the information from a single degraded image. Double-DIP [14] is a unified framework for layer decomposition of a single degraded image, based on coupled "Deep-image-Prior" (DIP) [25] networks. The work [26] is a zero-shot method for single-image dehazing where the input hazy image is disentangled by utilizing DCP [21] as prior. The zero-shot image restoration method [23] shows that a suitable degradation of the input image provides a controlled perturbation of the Koschmeider's model [24] parameters. Along with dehazing and low-light enhancement, [23] restores UW images based on the classic UW image formation model [10]. A recent zero-shot image dehazing method [44] generates a re-degraded hazy image for their zero-shot strategy. In the zero-shot UW image restoration method [39], the original image is fed into a zero-shot network like [26] and further enhanced by an improved level-adjustment methodology. [5] is based on layer disentanglement, and the disentangled components are further combined to reconstruct the UW image.

2.4 Contrastive learning for UW image restoration

Contrastive learning for UW image restoration is a less-explored area. [20] and [19] leverage contrastive learning and GAN to maximize the mutual information between raw and restored UW images where the corresponding patches of the raw UW image and the restored image are treated as positive samples, and the other patches from the raw UW image are treated as negative samples. [45] is a dehazing network utilizing contrastive learning where hazy images and the ground truth clean images are treated as negative and positive samples, respectively. This ensures that the dehazed image is pulled closer to the clean image and pushed away from the hazy image in the representation space.

We also make use of a contrastive learning strategy as in [45] but without utilizing ground truth images for positive samples. We leverage the haze information from the image patches to move the restored image from UWZR closer to the clean image representation and away from the UW image representation.

3 Proposed Method

UW image formation model is based on Koschmieder's light scattering model as given in [10]. Akkayanak *et al*. [1] observe that the transmission maps corresponding to backscatter and direct signal are different, although the classic model [10] treats them to be the same. The revised UW image formation model as proposed in [1] is given by

$$I(x) = J(x)T_D(x) + (1 - T_B(x))A \tag{1}$$

where x is pixel location, I is the UW image, J is the scene radiance which is the underlying clean image, A is global background light, while T_D and T_B are the transmission maps corresponding to direct signal and backscatter, respectively. Generally, $T_*(x) = e^{-\beta_c^* D(x)} : \{* = (D,B)\}$ where β_c^* is the channel-wise extinction coefficient, and $D(x)$ is the scene depth.

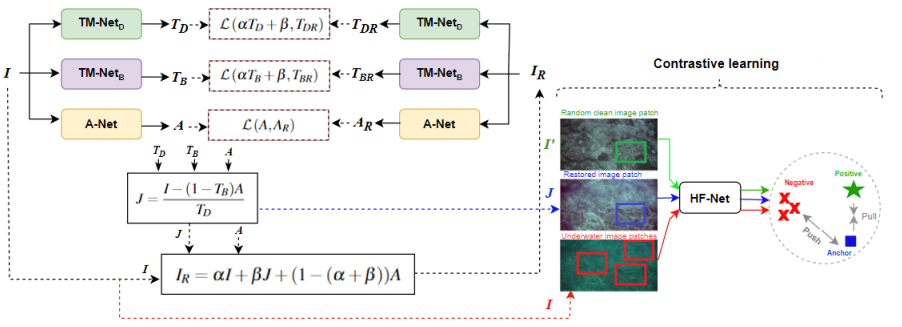


Figure 1: Schematic of our UWZR. Input UW image I is disentangled into global background light (A), transmission maps corresponding to direct signal (T_D) and back-scatter (T_B) using A-Net, TM-Net$_D$, and TM-Net$_B$, respectively. Restored image J is estimated analytically, and the re-degraded image I_R is obtained by mixing I, J, and A with suitable scale factors. UWZR enforces the relationship between the disentangled components of I and I_R. The contrastive learning section makes the restored image patch in the representation space (anchor) moves away from the hazy image representations (negatives) and moves closer to the clean image representation (positive).

3.1 Re-Degradation strategy

We rewrite Eq. 1 as follows:

$$I = JT_D + (1 - T_B)A \tag{2}$$

We further degrade I to generate another image I_R (we call it as the re-degraded image) by combining I, J, and A in a known ratio (α, β): $\alpha \in (0,1)$, $\beta \in (0,1)$, and $(\alpha+\beta) < 1$. Our re-degradation scheme can be written as:

$$I_R = \alpha I + \beta J + (1 - (\alpha + \beta))A \tag{3}$$

Substituting I from Eq. 2 in Eq. 3, and simplifying, we get:

$$I_R = J(\alpha T_D + \beta) + (1 - (\alpha T_B + \beta))A \tag{4}$$

By comparing Eq. 2 and Eq. 4, it can be seen that Eq. 4 is the UW image formation model for the formation of the re-degraded image I_R from the original clean image J but with different transmission maps: $T_{DR} = \alpha T_D + \beta$, $T_{BR} = \alpha T_B + \beta$, and with the same global background light $A_R = A$. Hence, the UW image formation model for I_R can be written as:

$$I_R = JT_{DR} + (1 - T_{BR})A_R \tag{5}$$

Now, the re-degraded image I_R and the raw UW image I follow a similar UW image formation model with the same global atmospheric light, and as the mixing ratio (α, β) is fixed, the relation between the transmission maps $T_* : \{* = (D,B)\}$ and $T_{*R} : \{* = (D,B)\}$ are known. These constraints between I and I_R can be leveraged for our zero-shot training. Concretely, our UWZR is optimized to generate outputs that satisfy the relationship between the components of the UW image and the re-degraded UW image.

A detailed block diagram of our approach is given in Fig. 1. From the input UW image I, transmission map estimation networks TM-Net$_D$ and TM-Net$_B$ estimate transmission maps corresponding to direct signal (T_D) and back-scatter (T_B), respectively. Global background light estimation network A-Net outputs global background light A. Restored image J is estimated analytically from the disentangled components T_D, T_B, and A. The re-degraded image I_R is generated using Eq. 3. Transmissions maps (T_{DR} and T_{BR}) and global background light (A_R) of I_R are estimated from the same networks TM-Net$_D$, TM-Net$_B$, and A-Net. The relationship between the disentangled components of I and I_R is used for the zero-shot training of UWZR. The network structure of each block is provided in the supplementary.

3.2 Contrastive learning

Our aim in contrastive learning is to make the restored output from our UWZR close to the distribution of clean images and move it away from the hazy UW image distribution. For extracting the contrastive features corresponding to the haze content, we use a haze feature extraction network (HF-Net) which outputs a 64-dimensional feature from the input image. From the restored image of UWZR, a random clean image, and the input UW image, we select patches of size 256×256 to form anchor, positive sample, and negative samples, respectively. It should be noted that the same patch from the UW hazy image corresponding to the restored image patch should be taken as one of the negative samples since our aim is to learn haze features. By leveraging contrastive learning, UWZR should be able to output a restored image that is closer to a clean image. The network structure of HF-Net is given in the supplementary.

3.3 Loss functions

The relationship between the disentangled components of I and I_R guides the training of our zero-shot network. Hence, we mainly employ two losses corresponding to global background light and the transmission maps of I and I_R. In order to facilitate contrastive learning, we employ contrastive loss also. In addition to the above three primary losses, two other no-reference losses are used to improve the visual quality of the restored image.

Transmission map relation loss ($\mathcal{L}_{TM}$):

As derived in Sec. 3.1, the network should be able to output transmission maps (both for direct signal and back-scatter) from I and I_R such that their relationship is maintained. If $\hat{T}_* : \{* = (D,B)\}$ and $\hat{T}_{*R} : \{* = (D,B)\}$ are the transmission maps obtained from I and I_R respectively, transmission map relation loss $\mathcal{L}_{TM}$ can written as:

$$\mathcal{L}_{TM} = ||(\alpha\hat{T}_D + \beta) - \hat{T}_{DR}||_2^2 + ||(\alpha\hat{T}_B + \beta) - \hat{T}_{BR}||_2^2 \tag{6}$$

Global background light equivalence loss ($\mathcal{L}_A$):

Global background light equivalence loss constrains global background light estimated from I and I_R to be the same. If $\hat{A}$ and $\hat{A}_R$ are the disentangled global background light from I and I_R respectively, $\mathcal{L}_A$ can be written as:

$$\mathcal{L}_A = ||\hat{A} - \hat{A}_R||_2^2 \tag{7}$$

Color loss ($\mathcal{L}_{CLR}$):

Based on the gray-world assumption of natural image statistics, a color loss [12] is added to correct the potential color deviations in the restored image. The color loss $\mathcal{L}_{CLR}$ is given as:

$$\mathcal{L}_{CLR} = \sum_{c\in\Omega} ||\mu(J^c) - 0.5||_2^2 \tag{8}$$

where μ represents mean, $\Omega = \{R,G,B\}$ is the set of color channels in the restored image J.

Pixel saturation avoidance loss ($\mathcal{L}_{PS}$):

Pixel saturation avoidance loss prevents saturation of pixels so that there is no hindrance to gradient flow [23] and is given by:

$$\mathcal{L}_{PS} = \sum_{x,c\in\Omega} (\max(J^c(x),1) + \max(J_R^c(x),1)) - \sum_{x,c\in\Omega} (\min(J^c(x),0) + \min(J_R^c(x),0)) \tag{9}$$

where J and J_R are respectively the scene radiance estimated from I and I_R, x is the pixel location, and $\Omega = \{R,G,B\}$.

Contrastive loss ($\mathcal{L}_{CL}$):
For the anchor image patch feature X_A, positive sample feature X^+, and negative sample features X_i^- $\{i = 1, \cdots, K\}$, contrastive loss is given as,

$$\mathcal{L}_{CL} = -\log \frac{\exp(X_A \cdot X^+/\tau)}{\exp(X_A \cdot X^+/\tau) + \sum_{i=1}^{K} \exp(X_A \cdot X_i^-/\tau)} \tag{10}$$

where τ is a temperature hyper-parameter, and K is the number of negative samples used. Anchor, positive and negative samples are taken as explained in Sec. 3.2.

Total Loss
The total loss of our network is given by

$$\mathcal{L} = a\mathcal{L}_{TM} + b\mathcal{L}_A + c\mathcal{L}_{CLR} + d\mathcal{L}_{PS} + e\mathcal{L}_{CL} \tag{11}$$

where a, b, c, d, and e are the weights corresponding to different losses. We empirically set $a = b = c = d = 1$, and $e = 1000$.

There are important differences between UWZR and SoTA zero-shot methods. Unlike [23] and [39], UWZR is based on the revised UW image formation model [1] where the transmission maps corresponding to direct-scatter and back-scatter are distinct. Hence our transmission map relation loss contains two terms corresponding to both T_D and T_B. The re-degradation strategy in UWZR is different from [23, 44], and is more general. If the simple UW image formation model [10] is considered (i.e., $T_D = T_B$), then our re-degradation equation (Eq. 3) becomes the re-degradation strategy in [23] and [44] when $\beta = 0$, and $\beta = 1 - \alpha$, respectively. Similar to [5, 23], UWZR does not use any prior-related losses for training unlike [26, 39, 44]. Only our method introduces a contrastive learning scheme along with zero-shot learning for UW image restoration.

4 Experiments

In this section, we first discuss the datasets used for comparison along with the details of implementation, followed by quantitative and qualitative evaluations of our results. We consider state-of-the-art methods of UW image restoration for comparison. Ablation studies are included to verify the effectiveness of our contributions.

4.1 Datasets and implementation details

For comparison, we have used four standard real-world UW image datasets, UIEB [30], HICRD [20], RUIE [33], and SQUID [4]. For quantitative evaluation, we have used 190, 120, 100, and 72 images from UIEB [30], HICRD [20], RUIE [33], and SQUID [4] datasets, respectively. UIEB [30] and HICRD [20] datasets have clean reference images. RUIE [33] and SQUID [4] datasets contain only raw UW images. Training is done for 500 epochs using Adam optimizer with a learning rate of 0.001. We conduct our experiments on a PC with Intel Xeon CPU, 24 GB RAM, and an NVIDIA GeForce RTX3090 GPU. The selection of α and β is discussed in the supplementary. For contrastive learning, we have used patches of size 256×256 as the inputs to HF-Net. We have taken a random clean image from UIEB [30] dataset as positive samples for contrastive learning.

4.2 Performance comparison

We compared UWZR with three traditional UW image restoration methods: GDCP [38], IBLA [37], Histogram prior [27]; one unsupervised data-driven method for UW restoration: USUIR [12]; three supervised UW restoration networks: WaterNet [30], UIEC2-Net [43],

Dataset		UIEB [30]		HICRD [20]		RUIE [33]		SQUID [4]	
Method	Class	PSNR↑	SSIM↑	PSNR↑	SSIM↑	UIQM↑	UCIQE↑	UIQM↑	UCIQE↑
GDCP [38]	Trad.	13.3	0.55	11.3	0.33	2.62	0.53	-0.91	0.51
IBLA [37]		14.3	0.57	12.8	0.16	1.73	0.51	0.21	0.47
Hist. prior [27]		18.5	0.59	14.4	0.60	4.15	**0.67**	**5.71**	0.59
USUIR [12]	Unsup.	20.3	0.84	13.7	0.55	3.04	0.57	3.06	0.62
WaterNet [30]	Sup.	19.3	0.83	20.8	0.67	3.51	0.56	2.08	0.55
UIEC2-Net [43]		**23.6**	**0.85**	15.8	0.76	3.42	0.57	2.91	0.59
PUIE-Net [13]		**21.8**	**0.87**	**21.0**	**0.78**	4.01	0.53	2.32	0.50
Double-DIP [14]	Zero-shot	12.4	0.38	13.9	0.17	1.85	0.47	1.29	0.41
Chai *et al.* [6]		10.1	0.19	14.6	0.18	4.04	0.52	1.66	0.54
Kar *et al.* [23]		13.7	0.51	11.3	0.49	**6.60**	**0.66**	**3.84**	**0.66**
Ours: UWZR		19.4	0.77	**21.4**	**0.79**	**4.51**	**0.67**	3.20	**0.64**

Table 1: Quantitative comparisons of enhanced image quality on datasets UIEB [30], HICRD [20], RUIE [33], and SQUID [4] using image quality assessment metrics. Higher values are better. PSNR is in dB. (Best: Red bold highlight, Second best: Blue bold highlight). Trad.: traditional methods, Unsup.: unsupervised network, Sup.: supervised network. Note that, supervised/unsupervised networks are trained with UIEB [30] dataset (Best among untrained methods: Red non-bold highlight).

PUIE-Net [13]; and three other zero-shot restoration methods: DDIP [14], an untrained network for UW restoration by Chai *et al.* [6], and the image restoration network by Kar *et al.* [23]. For testing the unsupervised [12] and three supervised [13, 30, 43] networks, we have taken their weights trained on UIEB [30] datasets. All the other 7 methods (traditional and zero-shot methods), including our UWZR, do not need training on a large-scale dataset.

4.2.1 Quantitative evaluation

Performance on image restoration is evaluated using the full-reference image quality assessment metrics PSNR and SSIM for datasets UIEB [30] and HICRD [20]. For datasets without ground truth images, RUIE [33] and SQUID [4], we use two no-reference UW image quality assessment metrics UIQM [36] and UCIQE [46]. In Table 1, average metric values calculated for different methods are given. The higher the metric values, the better the performance. It can be observed that, for the UIEB dataset, supervised methods UIEC2-Net [43] and PUIE-Net [13] perform better than other methods. It should be noted that all the supervised/unsupervised methods are trained with the UIEB dataset itself. Among all the untrained methods (traditional as well as zero-shot), our UWZR has the best score for UIEB. For HICRD [20] dataset, UWZR has the best score among all the methods. Unlike in the case of UIEB dataset, supervised networks are not the best for other datasets which reveals their weak generalization capability. The method of Kar *et al.* [23] and our UWZR have better scores for RUIE [33] and SQUID [4] datasets than other methods. The traditional histogram prior method [27] also has the best UCIQE value for RUIE and the best UIQM value for SQUID. It can be inferred from Table 1 that, our approach performs equally well for both UIEB [30] and HICRD [20] datasets, unlike the case of supervised/unsupervised networks, which work the best on UIEB on which they are trained but exhibit inferior performance for other datasets. For RUIE [33] and SQUID [4] datasets also, our method has good scores.

4.2.2 Qualitative evaluation

For qualitatively evaluating the performance on image restoration, in Fig. 2, we give comparison results of restored images from different methods for images (one image each) from HICRD [20], UIEB [30], RUIE [33], and SQUID [4] datasets. We have included qualitative results of more UW images from all the methods in supplementary. The restoration quality of GDCP [38], IBLA [37], and Double-DIP [14] is visually poor for all the datasets. His-

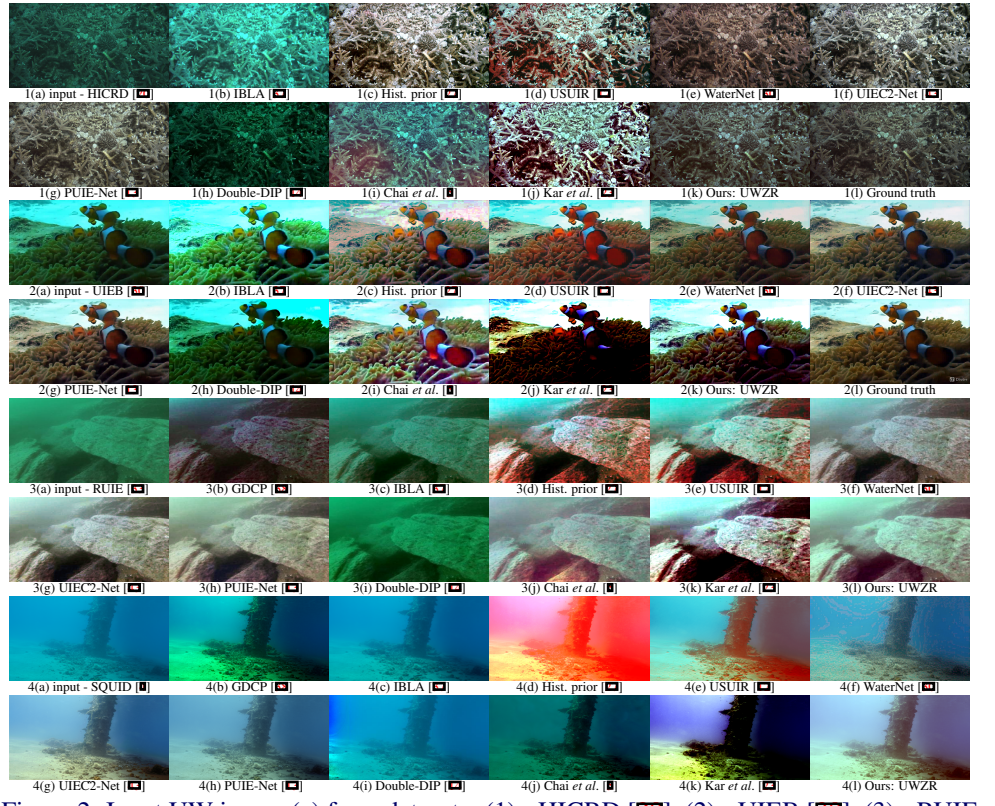
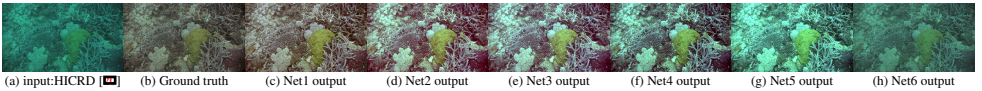


Figure 2: Input UW image (a) from datasets: (1) - HICRD [20], (2) - UIEB [30], (3) - RUIE [33], (4) - SQUID [4] with ground truth (1(l) and 2(l) for HICRD and UIEB) and the restored images from different methods. Note that our results are visually good, and for HICRD and UIEB datasets, our output (1(k) or 2(k)) is quite close to ground truth (1(l) or 2(l)).

togram prior method [27] has high contrast outputs, but there are color artifacts due to over-restoration (Fig. 2.3(d) and 4(d)) even though it has higher quantitative metrics for RUIE and SQUID. The unsupervised method USUIR [42] also has color deviations for all the datasets. The outputs of supervised networks (Waternet [30], UIEC2-Net [43], and PUIE-Net [13]) for UIEB [30] dataset (on which they are trained) are close to ground truth but for HICRD [20] dataset, the outputs are quite distant from the ground truth. For RUIE [33] dataset, their output is visually of low-contrast, and for SQUID [4], the restoration quality is poor. Outputs of other zero-shot methods ([14], [5], and [23]) are over/under-saturated or have color deviations. For the method of Kar *et al*. [23], even though the no-reference metrics are higher for RUIE [33] and SQUID [4], their outputs are low-dynamic range images and most of the image portions are weakly illuminated (Fig. 2.2(j), 3(k), and 4(k)). For HICRD [20] dataset, the restored output of only our UWZR is close to ground truth. Our outputs do not have any color deviations. For all four images, UWZR delivers very good visual quality.

Our method performs consistently well (both visually and quantitatively) on all four datasets. Among untrained methods (both traditional and zero-shot), our UWZR performs the best. Despite being a zero-shot approach, our approach outperforms or performs as par with supervised/unsupervised UW image restoration networks which need a huge dataset for training. The time analysis of UWZR is discussed in the supplementary.



(a) input:HICRD [20] (b) Ground truth (c) Net1 output (d) Net2 output (e) Net3 output (f) Net4 output (g) Net5 output (h) Net6 output

Figure 3: Ablation studies on HICRD dataset [20] for different configurations.

N/w	$\mathcal{L}_{CLR}$	$\mathcal{L}_{PS}$	$\mathcal{L}_{A}$	$\mathcal{L}_{CL}$	$\mathcal{L}_{TM}$	PSNR↑ /SSIM↑
Net1	✓	✓	✓	✓	✓	21.4/0.79
Net2	×	✓	✓	✓	✓	21.1/0.76
Net3	✓	×	✓	✓	✓	20.8/0.69
Net4	✓	✓	×	✓	✓	18.5/0.67
Net5	✓	✓	✓	×	✓	17.2/0.67
Net6	✓	✓	✓	✓	×	15.3/0.59

(1) Ablation study

Clean image for $\mathcal{L}_{CL}$		PSNR↑ /SSIM↑
Ref1-UIEB [30]		21.4/0.79
Ref2-HICRD [20]		21.5/0.78
Ref3-UIEB [30]		21.3/0.78

(2) Dependency of different clean images

Table 2: (1) Ablation studies on HICRD dataset [20] for different configurations. (2) Different reference images (from both UIEB [30] and HICRD [20]) for the positive patches for contrastive learning, and the corresponding metric values. PSNR is in dB.

4.2.3 Ablation studies

To study the effectiveness of our contributions and different losses, we conduct ablation studies on HICRD dataset [20]. We formed 6 networks, Net1 to Net6 as shown in Table 2(1). For each network, we calculated the average PSNR and SSIM values which are given in Table 2(1). Outputs from each network for an image from HICRD dataset are given in Fig. 3. Net1 is our network UWZR. Net2 and Net3 are without color loss ($\mathcal{L}_{CLR}$) and pixel saturation avoidance loss ($\mathcal{L}_{PS}$), respectively. From the metric values and qualitative results, it is evident that the contribution of $\mathcal{L}_{CLR}$ and $\mathcal{L}_{PS}$ to the overall performance of UWZR is less. Net4, without global background light equivalence loss ($\mathcal{L}_{A}$), has lesser metric values and poor visual quality. Net5 and Net6 are without the contrastive loss ($\mathcal{L}_{CL}$) and transmission map relation loss ($\mathcal{L}_{TM}$), respectively, and their visual quality and the metric values are quite less. It is evident that $\mathcal{L}_{CL}$ and $\mathcal{L}_{TM}$ play an important role in our zero-shot UW restoration strategy. Our proposed $\mathcal{L}_{TM}$ has the highest contribution to the overall restoration quality.

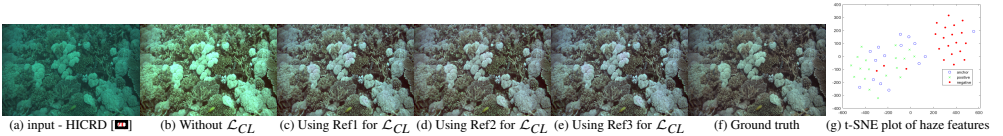
Underwater model	**Simple UW model**	**Revised UW model**	**Revised UW model**	**Revised UW model**
(α, β)	General (α, β)	$\beta = 0$	$\beta = 1 - \alpha$	General (α, β) (ours)
PSNR(dB)/SSIM	20.0/0.73	19.3/0.67	18.9/0.64	21.4/0.79

Table 3: Additional ablations on underwater image formation model and the re-degradation strategy.

To study the effect of UW image formation model on our network, we modified our network to include only one TM-Net (to mimic the classical UW model) and trained it. The PSNR/SSIM obtained on HICRD [20] dataset is given in Table 3. It can be seen that the classical UW model gives lower metric values. This reveals the importance of using the refined UW model. As additional ablations, we have included quantitative metrics of our network for $\beta = 0$ and $\beta = 1 - \alpha$ in Table 3 (on HICRD [20] dataset). It can be seen that our re-degradation strategy has the highest score.

4.2.4 Analysis on contrastive learning

In this section, we analyze the dependency of the type of clean images used for contrastive learning, on the restoration performance of UWZR. Three different reference images are taken as shown in Table 2(2). Ref1 and Ref3 are from UIEB [30], and Ref2 is from HICRD [20] dataset. Average PSNR and SSIM values obtained for HICRD [20] dataset while tak-



(a) input - HICRD [20] (b) Without $\mathcal{L}_{CL}$ (c) Using Ref1 for $\mathcal{L}_{CL}$ (d) Using Ref2 for $\mathcal{L}_{CL}$ (e) Using Ref3 for $\mathcal{L}_{CL}$ (f) Ground truth (g) t-SNE plot of haze features

Figure 4: Outputs of UWZR (b) without using $\mathcal{L}_{CL}$ and (c,d,e) using positive patches from different reference images for contrastive learning for an image (a) from HICRD [20] dataset. (g): t-SNE plot of haze features extracted by HF-Net from anchor, +ve, and -ve samples.

ing positive sample patches for contrastive learning from three different clean images (from Ref1, Ref2, and Ref3) are also given in Table 2(2). The qualitative results of UWZR without $\mathcal{L}_{CL}$, and with $\mathcal{L}_{CL}$ using Ref1, Ref2 and Ref3 are given in Fig. 4. It is evident that dependency of image restoration quality on the choice of reference image is negligible. From Fig. 4, it can be seen that without using $\mathcal{L}_{CL}$, restoration quality becomes poor.

HF-Net is trained to learn the haze content in the image patches by encouraging the positives (patches of clean image) closer while keeping the negatives (patches of input UW image) further away from the restored output. To understand the effect of $\mathcal{L}_{cl}$, we use t-SNE [40] (t-distributed stochastic neighbor embedding) to visualize learned features by HF-Net from the patches of restored image (anchors), random clean image (positives), and input UW image (negatives). The t-SNE plot is plotted in Fig. 4(g) where each point denotes the 2D representation of the 64-dimensional haze feature. By using $\mathcal{L}_{cl}$, the features of the anchor come closer to the clean image patches but are far away from the haze features of input UW image patches. This helps HF-Net to improve the restoration quality.

5 Conclusion

In this work, we proposed a zero-shot underwater image restoration method in which we introduced a re-degradation strategy to generate a secondary UW image, and theoretically showed the relationship between the latent components of the input UW image and the re-degraded UW image, that we utilized to formulate our training strategy. We introduced a contrastive learning scheme also to improve the restoration quality. Experiments on real UW datasets demonstrate that UWZR is superior to SoTA methods for UW image restoration.

Acknowledgement

Support provided by the Department of Science and Technology, India through project No. EE1920271DSTX005001 is gratefully acknowledged.

References

[1] Derya Akkaynak and Tali Treibitz. A revised underwater image formation model. In *CVPR*, pages 6723–6732, 2018. doi: 10.1109/CVPR.2018.00703.

[2] Derya Akkaynak and Tali Treibitz. Sea-thru: A method for removing water from underwater images. In *CVPR*, pages 1682–1691, 2019. doi: 10.1109/CVPR.2019.00178.

[3] Cosmin Ancuti, Codruta Orniana Ancuti, Tom Haber, and Philippe Bekaert. Enhancing underwater images and videos by fusion. In *CVPR*, pages 81–88, 2012. doi: 10.1109/CVPR.2012.6247661.

[4] Dana Berman, Deborah Levy, Shai Avidan, and Tali Treibitz. Underwater single image color restoration using haze-lines and a new quantitative dataset. *IEEE PAMI*, 43(8): 2822–2837, 2021. doi: 10.1109/TPAMI.2020.2977624.

[5] Shu Chai, Zhenqi Fu, Yue Huang, Xiaotong Tu, and Xinghao Ding. Unsupervised and untrained underwater image restoration based on physical image formation model. In *ICASSP*, pages 2774–2778, 2022. doi: 10.1109/ICASSP43922.2022.9746292.

[6] John Y. Chiang and Ying-Ching Chen. Underwater image enhancement by wavelength compensation and dehazing. *TIP*, 21(4):1756–1769, 2012. doi: 10.1109/TIP.2011.2179666.

[7] Chaitra Desai, Badduri Sai Sudheer Reddy, Ramesh Ashok Tabib, Ujwala Patil, and Uma Mudenagudi. Aquagan: Restoration of underwater images. In *CVPRW*, pages 295–303, 2022. doi: 10.1109/CVPRW56347.2022.00044.

[8] Chaitra Desai, Sujay Benur, Ramesh Ashok Tabib, Ujwala Patil, and Uma Mudenagudi. Depthcue: Restoration of underwater images using monocular depth as a clue. In *WACVW*, pages 196–205, 2023. doi: 10.1109/WACVW58289.2023.00025.

[9] Paulo L.J. Drews, Erickson R. Nascimento, Silvia S.C. Botelho, and Mario Fernando Montenegro Campos. Underwater depth estimation and image restoration based on single images. *IEEE CG&A*, 36(2):24–35, 2016. doi: 10.1109/MCG.2016.26.

[10] P. Drews Jr, E. do Nascimento, F. Moraes, S. Botelho, and M. Campos. Transmission estimation in underwater single images. In *ICCVW*, pages 825–830, 2013. doi: 10.1109/ICCVW.2013.113.

[11] Xueyang Fu, Peixian Zhuang, Yue Huang, Yinghao Liao, Xiao-Ping Zhang, and Xinghao Ding. A retinex-based enhancing approach for single underwater image. In *ICIP*, pages 4572–4576, 2014. doi: 10.1109/ICIP.2014.7025927.

[12] Zhenqi Fu, Huangxing Lin, Yan Yang, Shu Chai, Liyan Sun, Yue Huang, and Xinghao Ding. Unsupervised underwater image restoration: From a homology perspective. *AAAI*, 36(1):643–651, Jun. 2022.

[13] Zhenqi Fu, Wu Wang, Yue Huang, Xinghao Ding, and Kai-Kuang Ma. Uncertainty inspired underwater image enhancement. In *Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XVIII*, page 465–482, Berlin, Heidelberg, 2022. Springer-Verlag. ISBN 978-3-031-19796-3. doi: 10.1007/978-3-031-19797-0_27. URL https://doi.org/10.1007/978-3-031-19797-0_27.

[14] Yosef Gandelsman, Assaf Shocher, and Michal Irani. "double-dip": Unsupervised image decomposition via coupled deep-image-priors. In *CVPR*, pages 11018–11027, 2019. doi: 10.1109/CVPR.2019.01128.

[15] Ahmad Shahrizan Abdul Ghani and Nor Ashidi Mat Isa. Underwater image quality enhancement through rayleigh-stretching and averaging image planes. *Int. J. Nav. Archit. Ocean Eng.*, 6(4):840–866, 2014. ISSN 2092-6782. doi: https://doi.org/10.2478/IJNAOE-2013-0217. URL https://www.sciencedirect.com/science/article/pii/S2092678216302588.

[16] Clément Godard, Oisin Mac Aodha, and Gabriel J. Brostow. Unsupervised monocular depth estimation with left-right consistency. In *CVPR*, pages 6602–6611, 2017. doi: 10.1109/CVPR.2017.699.

[17] Yecai Guo, Hanyu Li, and Peixian Zhuang. Underwater image enhancement using a multiscale dense generative adversarial network. *IEEE J. Ocean. Eng.*, 45(3):862–870, 2020. doi: 10.1109/JOE.2019.2911447.

[18] Praful Hambarde, Subrahmanyam Murala, and Abhinav Dhall. Uw-gan: Single-image depth estimation and image enhancement for underwater images. *IEEE Trans. Instrum. Meas.*, 70:1–12, 2021. doi: 10.1109/TIM.2021.3120130.

[19] Junlin Han, Mehrdad Shoeiby, Tim Malthus, Elizabeth Botha, Janet Anstee, Saeed Anwar, Ran Wei, Lars Petersson, and Mohammad Ali Armin. Single underwater image restoration by contrastive learning. In *IGARSS*, pages 2385–2388, 2021. doi: 10.1109/IGARSS47720.2021.9553857.

[20] Junlin Han, Mehrdad Shoeiby, Tim Malthus, Elizabeth Botha, Janet Anstee, Saeed Anwar, Ran Wei, Mohammad Ali Armin, Hongdong Li, and Lars Petersson. Underwater image restoration via contrastive learning and a real-world dataset. *Remote Sensing*, 14(17), 2022. ISSN 2072-4292. doi: 10.3390/rs14174297. URL https://www.mdpi.com/2072-4292/14/17/4297.

[21] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *PAMI*, 33(12):2341–2353, 2011. doi: 10.1109/TPAMI.2010.168.

[22] Kashif Iqbal, Michael Odetayo, Anne James, Rosalina Abdul Salam, and Abdullah Zawawi Hj Talib. Enhancing the low quality images using unsupervised colour correction method. In *2010 IEEE International Conference on Systems, Man and Cybernetics*, pages 1703–1709, 2010. doi: 10.1109/ICSMC.2010.5642311.

[23] Aupendu Kar, Sobhan Kanti Dhara, Debashis Sen, and Prabir Kumar Biswas. Zero-shot single image restoration through controlled perturbation of koschmieder's model. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16200–16210, 2021. doi: 10.1109/CVPR46437.2021.01594.

[24] Harald Koschmieder. Theorie der horizontalen sichtweite. *Beitrage zur Physik der freien Atmosphare*, pages 33–53, 1924.

[25] Victor Lempitsky, Andrea Vedaldi, and Dmitry Ulyanov. Deep image prior. In *CVPR*, pages 9446–9454, 2018. doi: 10.1109/CVPR.2018.00984.

[26] Boyun Li, Yuanbiao Gou, Jerry Zitao Liu, Hongyuan Zhu, Joey Tianyi Zhou, and Xi Peng. Zero-shot image dehazing. *IEEE Transactions on Image Processing*, 29: 8457–8466, 2020. doi: 10.1109/TIP.2020.3016134.

[27] Chong-Yi Li, Ji-Chang Guo, Run-Min Cong, Yan-Wei Pang, and Bo Wang. Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior. *TIP*, 25(12):5664–5677, 2016. doi: 10.1109/TIP.2016.2612882.

[28] Chongyi Li, Jichang Guo, and Chunle Guo. Emerging from water: Underwater image color correction based on weakly supervised color transfer. *SPL*, 25(3):323–327, 2018. doi: 10.1109/LSP.2018.2792050.

[29] Chongyi Li, Saeed Anwar, and Fatih Porikli. Underwater scene prior inspired deep underwater image and video enhancement. *Pattern Recognition*, 98: 107038, 2020. ISSN 0031-3203. doi: https://doi.org/10.1016/j.patcog.2019.107038. URL https://www.sciencedirect.com/science/article/pii/S0031320319303401.

[30] Chongyi Li, Chunle Guo, Wenqi Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao. An underwater image enhancement benchmark dataset and beyond. *TIP*, 29:4376–4389, 2020. doi: 10.1109/TIP.2019.2955241.

[31] Chongyi Li, Saeed Anwar, Junhui Hou, Runmin Cong, Chunle Guo, and Wenqi Ren. Underwater image enhancement via medium transmission-guided multi-color space embedding. *TIP*, 30:4985–5000, 2021. doi: 10.1109/TIP.2021.3076367.

[32] Jie Li, Katherine A. Skinner, Ryan M. Eustice, and Matthew Johnson-Roberson. Watergan: Unsupervised generative network to enable real-time color correction of monocular underwater images. *IEEE Robotics and Automation Letters*, 3(1):387–394, 2018. doi: 10.1109/LRA.2017.2730363.

[33] Risheng Liu, Xin Fan, Ming Zhu, Minjun Hou, and Zhongxuan Luo. Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light. *CSVT*, 30(12):4861–4875, 2020. doi: 10.1109/TCSVT.2019.2963772.

[34] M. Ludvigsen, G. Johnsen B. Sortland, and H. Singh. Applications of geo-referenced underwater photo mosaics in marine biology and archaeology. In *Oceanography*, volume 20, page 140–149, 2007.

[35] Charles H. Mazel. In situ measurement of reflectance and fluorescence spectra to support hyperspectral remote sensing and marine biology research. In *OCEANS 2006*, pages 1–4, 2006. doi: 10.1109/OCEANS.2006.307001.

[36] Karen Panetta, Chen Gao, and Sos Agaian. Human-visual-system-inspired underwater image quality measures. *IEEE J. Ocean. Eng.*, 41(3):541–551, 2016. doi: 10.1109/JOE.2015.2469915.

[37] Yan-Tsung Peng and Pamela C. Cosman. Underwater image restoration based on image blurriness and light absorption. *TIP*, 26(4):1579–1594, 2017. doi: 10.1109/TIP.2017.2663846.

[38] Yan-Tsung Peng, Keming Cao, and Pamela C. Cosman. Generalization of the dark channel prior for single image restoration. *TIP*, 27(6):2856–2868, 2018. doi: 10.1109/TIP.2018.2813092.

[39] Xie Q., Z. Gao X., Liu, and H. Huang. Underwater image enhancement based on zero-shot learning and level adjustment. *Heliyon, 9(4)*, 2023.

[40] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(86):2579–2605, 2008. URL http://jmlr.org/papers/v9/vandermaaten08a.html.

[41] Nisha Varghese and A. N. Rajagopalan. Self-supervised monocular underwater depth recovery, image restoration, and a real-sea video dataset. In *ICCV*, 2023.

[42] Yi Wang, Hui Liu, and Lap-Pui Chau. Single underwater image restoration using adaptive attenuation-curve prior. *IEEE Trans. Circuits Syst. I Regul. Pap.*, 65(3):992–1002, 2018. doi: 10.1109/TCSI.2017.2751671.

[43] Yudong Wang, Jichang Guo, Huan Gao, and Huihui Yue. Uiec$\hat{2}$-net: Cnn-based underwater image enhancement using two color space. *Signal Process. Image Commun.*, 96: 116250, 2021.

[44] Jianchong Wei, Yi Wu, Liang Chen, Kunping Yang, and Renbao Lian. Zero-shot remote sensing image dehazing based on a re-degradation haze imaging model. *Remote Sensing*, 14(22), 2022. ISSN 2072-4292. URL https://www.mdpi.com/2072-4292/14/22/5737.

[45] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. Contrastive learning for compact single image dehazing. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10546–10555, 2021. doi: 10.1109/CVPR46437.2021.01041.

[46] Miao Yang and Arcot Sowmya. An underwater color image quality evaluation metric. *TIP*, 24(12):6062–6071, 2015. doi: 10.1109/TIP.2015.2491020.

[47] J. Yuh and M. West. Underwater robotics. *Advanced Robotics*, 15(5):609–639, 2001. doi: 10.1163/156855301317033595. URL https://doi.org/10.1163/156855301317033595.

[48] Shu Zhang, Ting Wang, Junyu Dong, and Hui Yu. Underwater image enhancement via extended multi-scale retinex. *Neurocomputing*, 245:1–9, 2017. ISSN 0925-2312. doi: https://doi.org/10.1016/j.neucom.2017.03.029. URL https://www.sciencedirect.com/science/article/pii/S0925231217305246.

[49] Weidong Zhang, Peixian Zhuang, Hai-Han Sun, Guohou Li, Sam Kwong, and Chongyi Li. Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement. *TIP*, 31:3997–4010, 2022. doi: 10.1109/TIP.2022.3177129.