UNSUPERVISED AND UNTRAINED UNDERWATER IMAGE RESTORATION BASED ON PHYSICAL IMAGE FORMATION MODEL

Shu Chai[1,2]*, Zhenqi Fu*[1,2]*, Yue Huang*[1,2]*, Xiaotong Tu*[1,2]*, Xinghao Ding*[1,2⋆]

[1]Key Laboratory of Underwater Acoustic Communication and
Marine Information Technology, Ministry of Education, Xiamen University
[2]School of Informatics, Xiamen University, China
⋆dxh@xmu.edu.cn

ABSTRACT

Underwater images suffer from degradation caused by light scattering and absorption. Training a deep neural network to restore underwater images is challenging due to the labor-intensive data collection and the lack of paired data. To this end, we propose an unsupervised and untrained underwater image restoration method based on the layer disentanglement and the underwater image formation model. Specifically, our network disentangles an underwater image into four components, i.e., the scene radiance, the direct transmission map, the backscatter transmission map, and the global background light, which are further combined to reconstruct the underwater image in a self-supervised manner. Our method can avoid using paired training data and large-scale datasets, benefiting from the unsupervised and untrained characteristics. Extensive experiments demonstrated that our method obtains promising performance compared with six methods on three real-world underwater image databases.

***Index Terms*—** Underwater Image Restoration, Unsupervised Learning, Untrained Network, Underwater Image Formation Model, Layer Disentanglement

1. INTROUCTION

Acquiring clear images in the underwater scene is an important issue for ocean engineering and research. However, underwater images suffer from low visibility, color distortion, and blurriness due to light scattering and absorption. The existing underwater image restoration methods could be divided into three categories: non-physical model-based, physical model-based methods, and learning-based methods.

To be specific, non-physical model-based methods improve visual quality by directly adjusting the pixel value of images. For example, Li et al.[1] propose a contrast enhancement algorithm based on a histogram distribution prior and recover image color combined with a dehazing algorithm. Ancuti et al.[2] propose a multi-scale fusion strategy to fuse the color-compensated and white-balanced result of underwater images. As non-physical model-based methods neglect the imaging process, the quality of the restored image depends on the consistency between the adopted algorithms and the actual image properties. Physical model-based methods estimate the unknown parameters of the degradation model to restore underwater images. Following this paradigm, Peng et al.[3] propose a depth estimation method based on light absorption and image blurriness. Song et al.[4] propose statistical models of background lights estimation with a manually annotated background lights database. However, these model-based methods tend to produce unstable results in challenging underwater scenes, because estimating multiple underwater imaging parameters is knotty and the assumed underwater image formation model does not always hold.

In recent, advancement in deep learning has significantly improved the performance of computer vision tasks. For instance, Li et al.[5] propose a multi-color space encoder network and a medium transmission-guided decoder network for the restoration. But these learning-based methods have a major obstacle, that is, a large-scale labeled dataset is required to train the neural network. To address this, Li et al.[6] propose a WaterGAN to generate synthetic underwater images and use them to train a two-stage deep network. Li et al.[7] propose an underwater image synthesis algorithm based on underwater scene priors. Since the synthetic images are probably less informative and inconsistent with the real images, it could lead to the domain shift issue [8] when the network is applied to the real-world underwater images.

In this paper, we propose a novel end-to-end network that employs four subnetworks to disentangle an underwater image into four components. After that, these components are used in the revised underwater image formation model to reconstruct the original image. Thanks to the proposed novel loss function, our network can be trained without the ground-truth image (unsupervised) and an image collection

The study is supported partly by the National Natural Science Foundation of China under Grants 61971369, 52105126, 82172033, U19B2031, China Postdoctoral Science Foundation (No. 2021M702726), Science and Technology Key Project of Fujian Province (No. 2019HZ020009) and Fundamental Research Funds for the Central Universities 20720200003.

(untrained). Our contributions can be summarized as follows:

- This work proposes an unsupervised and untrained method for single underwater image restoration, which could avoid the labor-intensive data collection and the lack of real-world paired data.
- Our method takes the advantages of model-based and learning-based methods. The whole framework consists of four disentanglement subnetworks to capture the underwater imaging parameters, i.e., the scene radiance, the direct transmission map, the backscatter transmission map, and the global background light.
- Our method achieves comparable performance on three real-world underwater image databases, even though it is an unsupervised method.

2. METHOD

2.1. Underwater Image Formation Model

Our method is based on the revised underwater image formation model, which describes the degradation of underwater images due to light scattering and absorption. As the Beer-Lambert law [9], light propagation is associated with an attenuation factor $e^{-\beta d}$, where d is the distance from the source and β is the attenuation coefficient, which depends on the given scene. Guided by the Beer-Lambert law, Koschmieder [10] formulates the effect of light scattering as below:

$$B(d) = \left(1 - e^{-\beta d}\right) A \tag{1}$$

where A represents the global background light, and $B(d)$ is called the backscatter that degrades the image due to light reflected from suspended particles. Later, the scene radiance J is also attenuated by the same factor $e^{-\beta d}$. Thus we get:

$$D(d) = Je^{-\beta d} \tag{2}$$

where $D(d)$ is called the direct signal containing the scene information. Based on the above effects of light scattering, the underwater image formation model is usually expressed as the additive combination of $B(d)$ and $D(d)$. In addition, we need to incorporate the effects of light absorption on the attenuation coefficient for a more accurate model.

Akkaynak et al.[11] observe that in the underwater scene, the direct signal and the backscatter are governed by two distinct attenuation coefficients, i.e., RGB attenuation and backscatter coefficients. Thus we revise the underwater image formation model as follows:

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

where $\boldsymbol{I}$ represents the underwater image, $\boldsymbol{J}$ refers to the scene radiance, x is an image pixel, $d(x)$ is the scene depth at that pixel, β^D and β^B respectively denote the RGB attenuation and backscatter coefficients, which are different for each color channel due to the wavelength selective characteristics. For simplicity, we rewrite Eq. 3 as follows:

$$I(x) = J(x)T_D(x) + (1 - T_B(x))A \tag{4}$$

where $T_D(x) = e^{-\beta^D d(x)}$ is the direct transmission map and $T_B(x) = e^{-\beta^B d(x)}$ is the backscatter transmission map.

2.2. Network Architecture

Given a single underwater image as the input, we aim to recover the scene radiance. Fig. 1 shows our proposed unsupervised and untrained framework, which consists of four subnetworks, namely J-Net, TD-Net, TB-Net, and A-Net.

J-Net, TD-Net and TB-Net The three subnetworks are respectively used to estimate the scene radiance $J(x)$, the direct transmission map $T_D(x)$, and the backscatter transmission map $T_B(x)$. As the two transmission maps and the scene radiance are dependent on the input image, we adopt the same network architecture for J-Net, TD-Net and TB-Net, which consists of five convolution layers, four batch normalization layers, four ReLU activation layers and a layer of sigmoid function. In brief, this is a non-degenerate architecture [12], which can avoid the down-sampling operation to ensure that the image details are not lost.

A-Net A-Net is used to estimate the global background light A from the input image. Since the global background light is independent of the image content, it owns the global property. We reasonably assume that the global background light follows a latent Gaussian distribution, and the prediction of A becomes a variational inference problem [13]. Therefore A-Net takes a variational auto-encoder, which consists of an encoder, an intermedia block, and a symmetric decoder. The encoder produces a latent code z with the mean μ_z and the standard deviation σ_z. Then the intermedia block transforms the latent code to the latent Gaussian distribution $\mathcal{N}(\mu_z, \sigma_z^2)$ and generates the reconstruction of the latent code $\hat{z}$ sampled from the Gaussian model. After that, $\hat{z}$ is fed into the decoder to obtain the reconstruction of the global background light.

2.3. Loss Function

To enable unsupervised learning in our network, we propose a set of differentiable losses.

Self-supervised Reconstruction Loss Self-supervised reconstruction loss is proposed to constrain the layer disentanglement. Our method disentangles an underwater image into four components which are further used to reconstruct the input image at the top layer. We aim to minimize the loss $\mathcal{L}_{Rec}$ as below:

$$\mathcal{L}_{Rec} = \|I(x) - x\|_2^2 \tag{5}$$

where x denotes the input image and $I(x)$ is the reconstructed image.

Contrast Enhancement Loss The contrast of underwater images is reduced due to the haze effect caused by light scattering. Zhu et al.[14] observe that the difference between the brightness and the saturation is close to zero in a haze-free region. To supervise J-Net, the contrast enhancement loss $\mathcal{L}_{Con}$ is designed as:

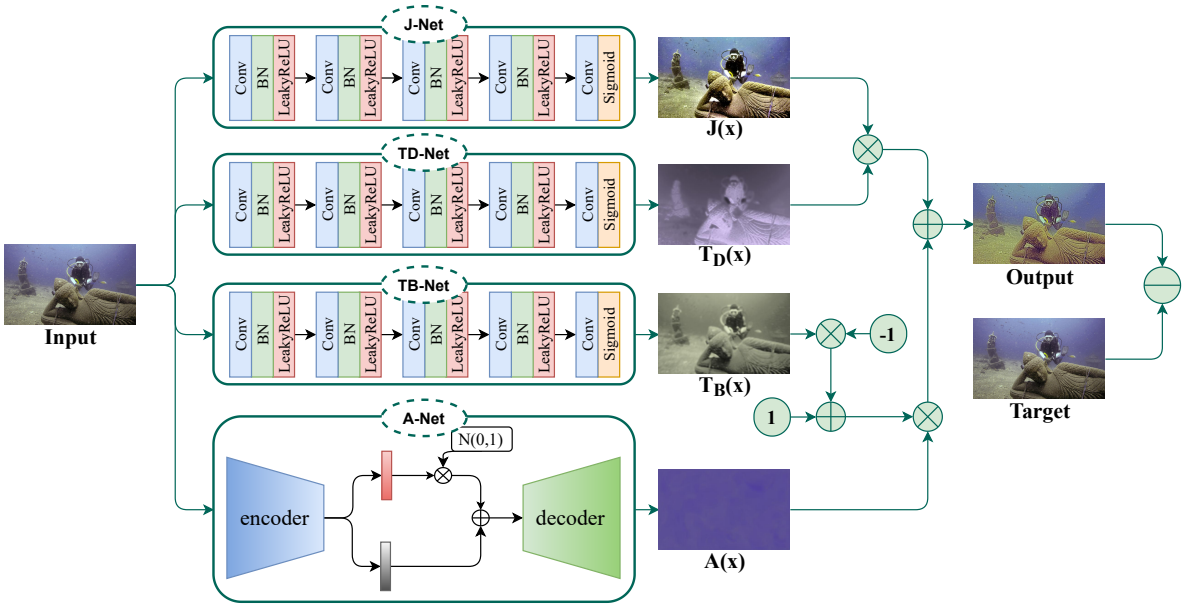

Fig. 1. The network architecture of our method, which consists of four subnetworks, i.e., the scene radiance estimation network (J-Net), the direct transmission map estimation network (TD-Net), the backscatter transmission map estimation network (TB-Net), and the global background light estimation network (A-Net). Taking a single underwater image as the input, the four subnetworks disentangle the input into four components, which are further combined to reconstruct the input underwater image based on the revised underwater image formation model. At the top layer, our network outputs the reconstructed image and takes the input as the target to supervised the layer disentanglement.

$$\mathcal{L}_{Con} = \|V(J(x)) - S(J(x))\|_2^2 \tag{6}$$

where $V(J(x))$ denotes the brightness of the estimated scene radiance $J(x)$, and $S(J(x))$ denotes the saturation of $J(x)$.

Color Constancy Loss Following the Gray-World color constancy hypothesis [15], we design a color constancy loss to correct the potential color deviations of the restored image. The loss $\mathcal{L}_{Col}$ is expressed as follows:

$$\mathcal{L}_{Col} = \sum_{c\in\Omega} \|\mu(J^c) - 0.5\|_2^2, \Omega = \{R, G, B\} \tag{7}$$

where $\mu(J^c)$ represents the average intensity value of color channel c in the estimated scene radiance.

Light Global Property Loss Light global property loss is designed for variational inference, which aims to minimize the difference between the latent code z and the reconstruction of the latent code $\hat{z}$ in A-Net. $\mathcal{L}_{KL}$ is calculated as below:

$$\mathcal{L}_{KL} = KL(\mathcal{N}(\mu_z, \sigma_z^2) \| \mathcal{N}(0, I)) \tag{8}$$

where $KL(\cdot)$ denotes the Kullback-Leibler divergence between two distributions, $\mathcal{N}(\mu_z, \sigma_z^2)$ denotes the learned latent Gaussian distribution, and $\mathcal{N}(0, I)$ refers to the standard normal distribution.

Transmission Consistency Loss Since the backscatter coefficients only depend on the optical properties of the water [11], they should be constants in the backscatter transmission map. We propose a transmission consistency loss to supervise TB-Net. The loss $\mathcal{L}_T$ is defined as:

$$\mathcal{L}_T = \sum_{(c_1,c_2)\in\varepsilon} \|\frac{\log T^{c_1}}{\log T^{c_2}} - \mu(\frac{\log T^{c_1}}{\log T^{c_2}})\|_2^2 \tag{9}$$

where $\varepsilon = \{(R,G),(R,B),(G,B)\}$ is a set of color pairs, T^c denotes the estimated backscatter transmission map of c channel, and μ is the average factor. As the RGB attenuation coefficients vary strongly with the distance d [11], we do not constrain TD-Net here.

Total Loss The total loss of our method is as below:

$$\mathcal{L} = \lambda_1\mathcal{L}_{Rec} + \lambda_2\mathcal{L}_{Con} + \lambda_3\mathcal{L}_{Col} + \lambda_4\mathcal{L}_{KL} + \lambda_5\mathcal{L}_T \tag{10}$$

where λ is the weight. We set $\lambda_1 = 1$, $\lambda_2 = 1$, $\lambda_3 = 1$, $\lambda_4 = 1$ and $\lambda_5 = 0.1$ empirically to get the best performance.

3. EXPERIMENT

3.1. Experimental Settings

We conduct our experiments on three real-world underwater image datasets, U45 [16], Challenging-60 [17], and Stereo [18]. For comprehensive comparisons, we compare our method with 6 methods which are divided into two groups, namely, two supervised methods and five unsupervised methods. To be specific, the supervised methods are Water-Net [19] and UWCNN [7]. The unsupervised methods are IBLA [3], ColorFusion [2], Statistical [4], and DDIP [20]. It should be pointed out that IBLA, ColorFusion and Statistical are the

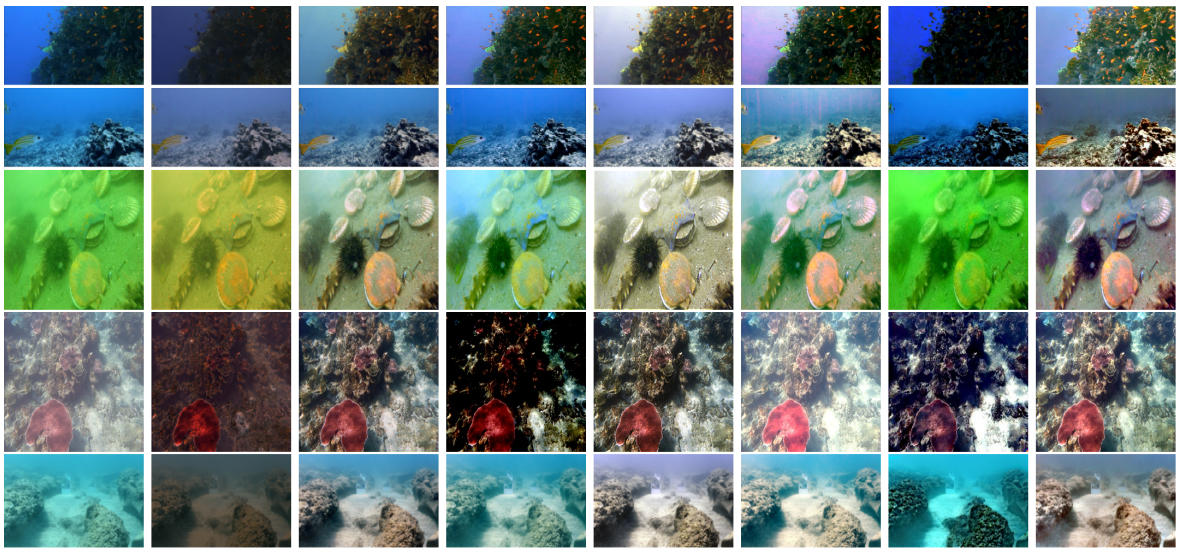

Fig. 2. Visual comparisons of the different restoration methods on real-world underwater images. Our method has a remarkable effect on color restoration and contrast enhancement.

Datasets	Measures	Techniques							
			Supervised		Unsupervised				
		Input	UWCNN	Water-Net	IBLA	ColorFusion	Statistical	DDIP	Ours
U45	UIQM/UCIQE	2.34/24.58	4.30/23.80	4.36/28.64	2.74/30.65	4.69/31.22	3.19/28.08	2.52/30.56	5.17/31.24
Challenging-60	UIQM/UCIQE	0.67/25.82	1.93/24.04	2.23/ 28.79	1.59/20.02	2.31/31.05	1.89/29.90	3.90/30.18	4.51/30.23
Stereo	UIQM/UCIQE	-0.09/20.66	2.67/23.39	2.86/31.85	0.74/25.01	4.08/35.73	1.81/26.00	1.74/29.52	4.10/30.37

Table 1. Quantitative comparisons of the different underwater image restoration methods on three real-world underwater image datasets. The best result is in red while the second best one is in blue.

shallow models whereas DDIP and our method are based on deep neural networks. Since the used datasets only contain underwater images without the reference images, we select two no-reference quality measurements: UIQM [21] and UCIQE [22] to evaluate the restored images. Higher values indicate the image has better objective quality. In addition to quantitative comparisons, visual comparisons are considered to analyze the performance of different methods. During training, we use the ADAM optimizer with 500 training iterations to optimize our network. The learning rate is set to 0.001. All experiments were implemented with PyTorch on an Nvidia RTX 2080Ti GPU.

3.2. Visual and Perceptual Comparisons

We present the visual comparisons with competitive methods on three datasets in Fig. 2. As can be seen, the input underwater images suffer from poor visibility caused by light scattering and absorption. The methods of UCWNN, DDIP, and IBLA cannot achieve satisfactory results due to the obvious color deviation. Statistical produces over-restoration artifacts and over-exposure phenomenon. Compared with Water-Net and ColorFusion, our method outperforms them in color restoration and contrast enhancement.

3.3. Quantitative Comparisons

Table 1 presents the quantitative results of six methods on three datasets. As can be seen, IBLA and Statistical employ handcrafted features to estimate the parameters of the image formation model, their performance is limited. ColorFusion is a non-physical model-based method, whose result is not stable. Compared with the supervised methods, our method achieves better performance, despite that it is an unsupervised method. As a result, our method reaches the best performance over other methods visually and metrically.

4. CONCLUSION

In this paper, we propose an unsupervised and untrained network for underwater image restoration based on the revised underwater image formation model. Our network consists of four joint disentanglement subnetworks, which are capable of estimating the four latent components in the imaging process to reconstruct the underwater image. We adopt the efficient loss function and train our network on a single underwater image. Extensive experiments demonstrate that our contributions lead our network to achieve promising performance in the quantitative and qualitative comparisons.

5. REFERENCES

[1] 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," *IEEE Transactions on Image Processing*, vol. 25, no. 12, pp. 5664–5677, 2016.

[2] Codruta O Ancuti, Cosmin Ancuti, Christophe De Vleeschouwer, and Philippe Bekaert, "Color balance and fusion for underwater image enhancement," *IEEE Transactions on image processing*, vol. 27, no. 1, pp. 379–393, 2017.

[3] Yan-Tsung Peng and Pamela C Cosman, "Underwater image restoration based on image blurriness and light absorption," *IEEE transactions on image processing*, vol. 26, no. 4, pp. 1579–1594, 2017.

[4] Wei Song, Yan Wang, Dongmei Huang, Antonio Liotta, and Cristian Perra, "Enhancement of underwater images with statistical model of background light and optimization of transmission map," *IEEE Transactions on Broadcasting*, vol. 66, no. 1, pp. 153–169, 2020.

[5] Chongyi Li, Saeed Anwar, Junhui Hou, Runmin Cong, Chunle Guo, and Wenqi Ren, "Underwater image enhancement via medium transmission-guided multi-color space embedding," *IEEE Transactions on Image Processing*, vol. 30, pp. 4985–5000, 2021.

[6] 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*, vol. 3, no. 1, pp. 387–394, 2017.

[7] Chongyi Li, Saeed Anwar, and Fatih Porikli, "Underwater scene prior inspired deep underwater image and video enhancement," *Pattern Recognition*, vol. 98, pp. 107038, 2020.

[8] Shai Ben-David, John Blitzer, Koby Crammer, Fernando Pereira, et al., "Analysis of representations for domain adaptation," *Advances in neural information processing systems*, vol. 19, pp. 137, 2007.

[9] Donald F Swinehart, "The beer-lambert law," *Journal of chemical education*, vol. 39, no. 7, pp. 333, 1962.

[10] H Koschmieder, "Theorie der horizontalen sichtweite, beitrage zur physik der freien atmosphare," *Meteorologische Zeitschrift*, vol. 12, pp. 33–53, 1924.

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

[12] Runde Li, Jinshan Pan, Zechao Li, and Jinhui Tang, "Single image dehazing via conditional generative adversarial network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8202–8211.

[13] Diederik P Kingma and Max Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[14] Qingsong Zhu, Jiaming Mai, and Ling Shao, "A fast single image haze removal algorithm using color attenuation prior," *IEEE transactions on image processing*, vol. 24, no. 11, pp. 3522–3533, 2015.

[15] Gershon Buchsbaum, "A spatial processor model for object colour perception," *Journal of the Franklin institute*, vol. 310, no. 1, pp. 1–26, 1980.

[16] Hanyu Li, Jingjing Li, and Wei Wang, "A fusion adversarial underwater image enhancement network with a public test dataset," *arXiv preprint arXiv:1906.06819*, 2019.

[17] Chongyi Li, Chunle Guo, Wenqi Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao, "An underwater image enhancement benchmark dataset and beyond," *IEEE Transactions on Image Processing*, vol. 29, pp. 4376–4389, 2019.

[18] Dana Berman, Deborah Levy, Shai Avidan, and Tali Treibitz, "Underwater single image color restoration using haze-lines and a new quantitative dataset," *IEEE transactions on pattern analysis and machine intelligence*, 2020.

[19] Chongyi Li, Chunle Guo, Wenqi Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao, "An underwater image enhancement benchmark dataset and beyond," *IEEE Transactions on Image Processing*, vol. 29, pp. 4376–4389, 2019.

[20] Yosef Gandelsman, Assaf Shocher, and Michal Irani, ""' Double-Dip": Unsupervised image decomposition via coupled deep-image-priors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11026–11035.

[21] Karen Panetta, Chen Gao, and Sos Agaian, "Human-visual-system-inspired underwater image quality measures," *IEEE Journal of Oceanic Engineering*, vol. 41, no. 3, pp. 541–551, 2015.

[22] Miao Yang and Arcot Sowmya, "An underwater color image quality evaluation metric," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 6062–6071, 2015.