Severely Degraded Underwater Image Enhancement with a Wavelet-based Network

Shunsuke Takao, Tsukasa Kita, Taketsugu Hirabayashi
Port and Airport Research Institute,
Department of Infrastructure
Digital Transformation Engineering,
Yokosuka, Japan

Abstract—Underwater images are important in marine science and ocean engineering fields owing to giving color information, low cost, and compact. Yet obtained underwater images are often degraded and restoring and enhancing wavelength selective signal attenuation of underwater images depending on complex underwater physical process is essential in practical application. While recently developed deep learning is a promising choice, constructing sufficiently large dataset covering whole real images is challenging, peculiar to underwater image processing. In order to supplement relatively small dataset, previous studies alternatively construct an artificial underwater image dataset based on a physical model or Generative Adversarial Network. Also, incorporating traditional signal processing methods into the network architecture has shown promising success, though enhancement of severely degraded underwater images remains to be a big issue. In this paper, we tackle underwater image enhancement based on an encoder-decoder based deep learning model incorporating discrete wavelet transform and whitening and coloring transform. We also construct a severely degraded real underwater image dataset. The presented model shows excellent results both qualitatively and quantitatively in the artificial and real image dataset. Constructed dataset is available at https://github.com/tkswalk/2022-IJACSA.

Keywords—*Underwater image enhancement; deep learning; discrete wavelet transform; whitening and coloring transform*

I. INTRODUCTION

Underwater optical images are essential in sensing vast ocean environment. Optical cameras beneficially capture high resolution color information, as well as relatively low cost and compact compared to other acoustic devices. While underwater optical images are essential especially in tasks requiring color information, such as ocean monitoring, maintenance of port facilities, and resource development, serious image degradation is obstacle in efficient utilization. Specifically, wavelength selective color distortion which displays blueish, greenish, and yellowish appearances, or decreased contrast caused by complex underwater physical process worsens the visibility of an underwater image [1], [2], as shown in the upper part of Fig. 1.

To overcome the low visibility of underwater images and expand the scope of application, underwater image enhancement methods based on deep learning have rapidly improved by refining model architecture and training dataset. In underwater image enhancement, deep learning models are mainly trained by mapping degraded images to the corresponding clear images. However, collecting clear and degraded real underwater image pairs is high cost or inherently difficult

Fig. 1. An Example of a Severely Degraded Underwater Image (Above) and an Enhanced Result by our Model (Below).

especially in turbid water in a coastal region. Alternatively, artificial underwater image datasets constructed with a simplified physical model or Generative Adversarial Network (GAN) are employed for training, yet their effectiveness are limited because real underwater images depend on complex physical process and many physical parameters like water body or ambient light and may be apart from artificial images [3], [4]. Subsequently, an artificial underwater image dataset based on the revised underwater image formation model [4] is recently proposed which more reflects real underwater environment [5].

Under the constraint of limited amount of data, incorporating traditional signal processing methods into the network architecture is also effective in underwater image enhancement. While the shallow CNN based model incorporating white balance, histogram equalization, and gamma correction has

shown measurable success, enhancement of severely degraded underwater images remains to be a challenging issue [6].

In this paper, we tackle severely degraded underwater image enhancement with an encoder-decoder based network combining discrete wavelet transform and whitening and coloring transform (WCT). The high frequency components of an input image is structurally extracted with discrete wavelet transform in the encoder part, and is preserved by passing them to the decoder part, thereby obtaining a sharp output. Also, as underwater images are quite diverse and display various tones of color and degrees of blurriness, input image features are whitened and mapped to style image features with WCT to stabilize training. The presented model is trained with the recently proposed physically revised artificial underwater image dataset [5] and an elaborated loss function. Also, we present a seriously degraded real underwater image dataset taken in Okinawa, Japan. The constructed dataset includes blueish and greenish images of divers, an underwater construction machine, and rubble mounds of port structures. Our underwater image enhancement model is evaluated with the artificial image dataset and the constructed real image dataset, showing fine results both qualitatively and quantitatively. Our main contributions are summarized as follows:

- We present an underwater image enhancement model combining discrete wavelet transform and whitening and coloring transform.

- We construct a real underwater image dataset including severely degraded blueish or greenish underwater images.

- The presented model successfully removes overall blueish tones of seriously degraded underwater images, mainly outperforming state-of-the-art underwater image enhancement methods both in real and artificial datasets.

II. Related Work

A. Previous Underwater Image Enhancement Methods

Supervised underwater image enhancement models based on Convolutional Neural Network (CNN), Generative Adversarial Network (GAN), and recently appeared Vision Transformer (ViT), have rapidly improved. As models mainly learn pixel transformation tasks, skip connection is often employed not to apart from the original input image. Also, encoder-decoder process is adopted to mitigate the input noise. To be specific, UWCNN is a densely connected CNN model where an input is injected to the different layers with no pooling layers or batch normalization steps [7]. FUnIE-GAN is a fully convolutional conditional GAN model. The generator has five encoder-decoder pairs with several skip connections to enable real time inference [8]. The above two models are either trained with an artificial underwater image dataset. Recently proposed ViT based model is also equipped with several skip connections to stabilize training. To cope with wavelength selective and spatially variant signal attenuation of underwater images, channel-wise attention and spatial-wise attention are incorporated into the architecture [9]. As the difficulty of covering whole real underwater images, incorporating traditional signal processing methods to the network process

is effective in underwater image processing. For example, Water-Net is a simple CNN based network which fuses the results of white balance, gamma correction, and histogram equalization [6]. First, three results of each signal processing methods and the original input are fed to the network to predict the three fusion coefficient maps. The predicted three coefficient maps are multiplied by the enhanced results which are obtained by passing through the three independent feature transformation units to reduce the artifacts introduced from the signal processing methods. The refined output is finally obtained by fusing the above three results. Also, discrete wavelet transform is employed to preserve fine image structure [10], [11]. Other than learning based methods, many unsupervised underwater image enhancement methods assume physical model and correct color distortion by imposing white balance, which often requires the estimation of ambient light or average color [12], [13].

B. Previous Underwater Image Datasets

As obtaining sufficient real underwater image pairs is challenging, construction of the dataset itself is important in underwater image processing. Based on a simplified underwater image formation model, [7] constructed an artificially deteriorated underwater image dataset to which visually matches real underwater images by setting the attenuation coefficient to a constant and neglecting other related physical parameters. More recently, based on the revised underwater image formation model [4], an artificial underwater image dataset is proposed which clearly takes into account the dependency of water types, lightning conditions, and camera sensors. The constructed dataset is implied to be more real compared to the previous one [5]. GAN-based approaches generate artificial underwater images by converting initially clear underwater images to degraded ones to cheat the classifier. The model is trained with an unpaired dataset by minimizing Cycle-Consistency loss [8], [14]. Other than artificial underwater images, clearly enhanced real underwater images are collected among results of many conventional enhancement methods by scoring human ranking by hand. This approach is expected to reflect human perceptions, yet is laborious and the sample size is limited to at most a few thousand [6], [9].

III. Methodology

Presented underwater image enhancement model is based on a simple encoder-decoder network architecture with several skip connections, similar to well known U-Net in image segmentation task [15], as shown in Fig. 2. Pooling and up-sampling layers are respectively replaced with discrete wavelet transform and inverse discrete wavelet transform to maintain structural information. Whitening and Coloring Transform (WCT) mainly employed in style transfer task is also incorporated into the model to mitigate covariate shift between training data distribution and test data distribution. Brief introduction of discrete wavelet transform and WCT is described followed by the details of model architecture.

A. Signal Reconstruction with Wavelet Transform

The power of discrete wavelet transform (DWT) especially using Haar wavelet is shown in style transfer and inverse problems by generalizing conventional pooling operations
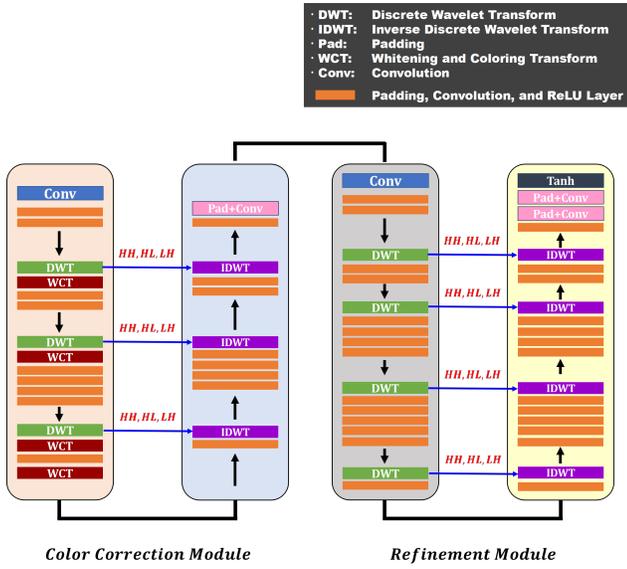
Fig. 2. An Overview of the Model Architecture. An Input Image goes through Several Convolution, Padding, and ReLU Layers Followed by WCT Layers in the Color Correction Module. High Frequency Components, HH, HL, LH, Extracted with Discrete Wavelet Transform in the Encoder are Passed to the Decoder to Preserve Detailed Signal. Compared to Color Correction Module, the Refinement Module is Simply Implemented by Removing the WCT Layers and Adding Several Convolutional Layers to Mitigate Noise.

Fig. 3. Our Model Recovers the Severely Degraded Artificial Underwater Image Better (Right) Compared with the Baseline (Left).

TABLE I. PSNR AND SSIM OF OUR MODEL AND BASELINE PER 10 WATER TYPES.

Water-type		I	IA	IB	II	III	1C	3C	5C	7C	9C
PSNR	Propose	16.61	16.316	16.568	15.124	15.8	16.987	15.94	16.214	15.485	15.559
	baseline	16.031	15.514	15.369	13.896	14.351	15.838	14.53	15.124	13.999	14.175
SSIM	Propose	0.702	0.686	0.684	0.608	0.649	0.7	0.652	0.662	0.626	0.629
	baseline	0.684	0.66	0.65	0.566	0.606	0.672	0.609	0.625	0.578	0.582

WCT is incorporated in our model to mitigate the covariate shift between training data and testing data.

C. Network Architecture

The network architecture shown in Fig. 2 is a simple encoder-decoder based model with several skip connections and no pooling layers. In order to preserve detailed image signal, high frequency components extracted with discrete wavelet transform in the encoder part, $\{LH^T, HL^T, HH^T\}$, are passed to the inverse discrete wavelet transform in the decoder part. WCT is incorporated in the color correction module to normalize feature maps.

Input images are first passed through a convolutional layer followed by several convolution, padding, and ReLU activation layers in the color correction module. Then, encoded features go through the discrete wavelet transform layer and low frequency component, LL^T, is processed with WCT and subsequent deeper layers. The remaining high frequency components, LH^T, HL^T, HH^T, are skipped to the decoder part to preserve detailed signal. The encoded features and the passed high frequency components are up-sampled with inverse discrete wavelet transform followed by several convolution, padding, and ReLU activation layers. The subsequent refinement module is similar to the color correction module, but WCT is removed and several convolution, padding, and ReLU activation layers and the last layers of padding, convolution, and hyperbolic tangent activation layer are added to mitigate input noise. Such repeated structure is designed to extract local image structure. Kernel size and stride of all convolutional layers are set to 3 and 1, respectively. We use pre-trained model on photo-realistic style transfer method [11], denoted as baseline, to normalize the complex input distribution caused by the complicated real underwater environment. Our model is similar to [11], but one DWT and IDWT layers, a few convolution, padding, and ReLU activation layers, and the last layers of padding, convolution, and hyperbolic tangent activation layer are added. Here, Fig. 3 shows that

like average pooling or max pooling, which simply subsamples and summarizes the neighboring pixel information [16], [11], [17]. Haar wavelet operation consists of four kernels, $\{LL^T, LH^T, HL^T, HH^T\}$, where L and H are respectively defined as $L^T := \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\end{bmatrix}, H^T := \frac{1}{\sqrt{2}}\begin{bmatrix}-1 & 1\end{bmatrix}$. Frequency information is efficiently retained and extracted with L and H, and low frequency signal is captured by L while high frequency signal is captured by H. Inverse discrete wavelet transform (IDWT) is the mirror operation of discrete wavelet transform and is employed for structural reconstruction in the decoder part with minimal noise amplification.

B. Whitening and Coloring Transform (WCT)

The aim of WCT in style transfer is to obtain a stylized image preserving content features [18]. After feature extraction with a pre-trained network, covariance matrix of high dimensional feature maps f_c of content image is first made to be an identity matrix (whitening), followd by singular value decomposition. The whitened content feature $\hat{f}_c$ is then projected onto the eigenspace of the style feature f_s (coloring), described as following procedure:

1) Whitening: Obtain whitened feature $\hat{f}_c = E_c D_c^{-\frac{1}{2}} E_c^T f_c$

2) Coloring: Obtain colored feature $\hat{f}_{cs} = E_s D_s^{\frac{1}{2}} E_s^T \hat{f}_c$

where E_c and D_c are respectively an orthogonal matrix of eigenvectors and a diagonal matrix of the covariance matrix of f_c, and E_s and D_s are that of f_s. Here, $f_c f_c^T = E_c D_c E_c^T$ is satisfied. Colored feature $\hat{f}_{cs}$ satisfies $\hat{f}_{cs} \hat{f}_{cs}^T = f_s f_s^T$, preserving higher order feature correlation which reflects style information. As underwater image distribution is complex and various,

Fig. 4. We Collect Real Underwater Images around Rubble Mounds of Port Structures taken in Okinawa, Japan. The Collected Underwater Images are Significantly Degraded, Showing Blueish or Greenish Appearances.

our model recovers the severely degraded artificial underwater image better (right) compared with the baseline (left). As for quantitative metric, Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM), full reference metrics of image quality which reflects human perception, are computed. PSNR and SSIM of our method and the baseline per 10 water types classified by [19], are shown in Table I. Scores of PSNR and SSIM are improved approximately 1 and 0.3, respectively, compared to the baseline in all water types.

D. Loss Function

We combine three loss terms, reconstruction loss L_{rec}, Laplacian pyramid Lap1 loss L_{lap}, and luminance loss L_{lum} between a correct image I_c and an estimated image I_e for training our network, defined as follows:

$$Loss = \alpha L_{rec} + \beta L_{lap} + \lambda L_{lum} \tag{1}$$

where α, β, and λ are hyper parameters.

Reconstruction loss L_{rec} means the pixel wise l1 distance between I_c and I_e, denoted as follows:

$$L_{rec}(I_c, I_e) = |I_c - I_e|_1 \tag{2}$$

Laplacian pyramid Lap1 loss L_{lap} measures differences between I_c and I_e in Laplacian pyramid representation to take into account various frequency components and get a structural image [20], [21], defined as:

$$L_{lap}(I_c, I_e) = \sum_i^{2^{2i}} |L^i(I_c) - L^i(I_e)|_1 \tag{3}$$

Here, $L_i(I)$ means the i-th level of the Laplacian pyramid representation of an image I [22].

Also, we propose luminance loss L_{lum}. The luminance loss measures pixel wise difference between I_c and I_e of their luminance components after transforming to YCbCr color space, described as follows:

$$L_{lum} = |Y(I_c) - Y(I_e)|_1 \tag{4}$$

where luminance component Y can be defined as:

$$Y = 0.299R + 0.587G + 0.114B$$

Here, R, G, and B mean the red, green and blue channels of the original image, respectively. The luminance loss is proposed to facilitate training, as luminance components are less susceptible to color tones of an underwater image.

IV. EXPERIMENTS

A. Construction of Real Underwater Image Dataset

We collected real underwater images around rubble mounds of port structures in Okinawa, Japan. The constructed dataset contains significantly degraded underwater images of an underwater construction machine, a diver, and rubble mounds, which are taken with GoPro HERO4. As shown in Fig. 4, the underwater images are directly taken by a diver or a camera mounted with the upper part of the construction machine, showing blueish or greenish appearances. The constructed dataset is available at https://github.com/tkswalk/2022-IJACSA.

B. Experimental Setting

We train the model with a recently proposed artificial underwater image dataset [5] based on the revised underwater image formation model which more reflects real underwater environment. The model clearly considers the dependencies of related physical parameters, such as water types, lightning conditions, and camera sensors [4]. In the dataset, wavelength data of 10 water types classified by [19], two camera sensors, and the three light spectrum data are employed, namely 60 kinds of artificial images are generated per one image. Clear indoor RGB-D images from NYU Depth Dataset V2 [23] containing depth information are transformed based on the underwater physical model, resulting in 86940 image pairs in total [4], [5]. Among the 1449 original images from NYU Depth Dataset V2, first 1000 images are used for the training data, next 300 images are used for the validation data, and the last 149 images are used for the test data.

A degraded input image is first resized to 256×256 resolutions and mapped to an enhanced image. The coefficients of the loss function, α, β, and λ, are respectively set to 1, 10, 1. Adam optimizer [24] is adopted and the learning rate is set to 0.0001. The training epoch is 80 and the model is implemented with PyTorch and GeForce RTX 2080 Ti GPU.

C. Results and Discussions of Artificial Underwater Images

We qualitatively and quantitatively compare the restoration results with available state-of-the-art underwater image enhancement methods. As shown in Fig. 5, FUnIE-GAN (4th row) [8], UWCNN [7] (7th row), Water-Net [6] (8th row), and U-Transformer [9] (9th row) are evaluated for the deep learning based approaches, while results of retinex-based theory (5th row, denoted as Retinex) [13] and underwater dark channel prior (6th row, denoted as UDCP) [12] are also compared for the unsupervised methods. The first row of Fig. 5 shows the

TABLE II. RESULTS OF PSNR PER 10 WATER TYPES

PSNR	I	IA	IB	II	III	1C	3C	5C	7C	9C
UWCNN	14.74	13.95	13.31	11.03	11.17	11.31	11.3	10.37	10.89	11.11
FUnIE-GAN	17.3	15.8	14.99	13.54	13.45	15.95	13.57	13.86	12.44	12.17
Water-Net	17.02	15.45	14.6	13.73	13.14	16.52	13.24	12.65	12.4	12.5
Retinex	15.86	14.87	13.97	12.87	12.93	14.52	13.01	13.48	12.39	12.49
UDCP	13.38	12.55	11.69	11.03	10.78	12.4	10.81	10.59	10.36	10.54
U-Transformer	16.63	15.15	14.46	13.11	13.09	15.5	13.21	13.54	12.4	12.59
Ours	16.61	16.32	16.57	15.12	15.8	16.99	15.94	16.21	15.48	15.56

TABLE III. RESULTS OF SSIM PER 10 WATER TYPES

SSIM	I	IA	IB	II	III	1C	3C	5C	7C	9C
UWCNN	0.695	0.655	0.622	0.511	0.511	0.541	0.534	0.508	0.462	0.48
FUnIE-GAN	0.684	0.64	0.611	0.55	0.561	0.644	0.564	0.567	0.53	0.528
Water-Net	0.765	0.683	0.655	0.592	0.599	0.706	0.603	0.586	0.57	0.5793
Retinex	0.715	0.677	0.652	0.579	0.603	0.676	0.606	0.624	0.575	0.5791
UDCP	0.658	0.607	0.544	0.479	0.476	0.588	0.479	0.462	0.452	0.469
U-Transformer	0.651	0.617	0.593	0.538	0.549	0.616	0.553	0.565	0.53	0.536
Ours	0.702	0.686	0.684	0.608	0.649	0.7	0.652	0.662	0.626	0.629

employed test data of indoor images from [23], and they are artificially converted based on a underwater image formation model, as shown in the second row. PSNR and SSIM, full reference metrics measuring the image quality, are calculated for the quantitative evaluation.

The artificial underwater image dataset contains various colors and degrees of degradation which reflects water types or lightning conditions [5]. In qualitative evaluation in Fig. 5, many restoration results are not sufficiently well recovered because of the severe image degradation of an input. While our model relatively well restored blueish, greenish, and yellowish artificial underwater images (3rd row), previous methods hardly improve the visibility (4th [8], 6th [12], and 7th row [7]) or insufficiently output whitish images (5th [13], 8th [6], and 9th row [9]). Also, PSNR and SSIM per 10 water types classified by [19] are respectively shown in Table II and Table III. Our model achieves better performance compared to other methods in 9 out of 10 water types. While our model mainly outperforms other methods in almost all water types, output images are sometimes decolored as shown in the 4th column of the 3rd row.

D. Results and Discussions of Real Underwater Images

Next, we proceed to restoration results of real underwater images, as shown in the 1st row of Fig. 6. Real underwater images of 1st to 3rd column come from the constructed dataset collected in Okinawa, Japan, and the remains come from [6] which contains severely degraded underwater images. Our model (2nd row) restores significantly degraded blueish (1st, 2nd, 3rd, 6th column), greenish (4th column), and yellowish (5th column) underwater images. The output images contain less overall blueish tones compared to results of other methods. Among the results of previous methods, Water-Net [6] (7th row) combining white balance, gamma correction, and histogram equalization, are better also in the yellowish and greenish inputs, yet failed to restore the severely degraded input (1st column). The performance of Water-Net is mainly dominated by the signal processing results as Water-Net fuses outputs of them. FUnIE-GAN [8] (3rd row), GAN based model, hardly improves the visibility and adds grid artifact in severely degraded inputs. UWCNN [7] (6th row), CNN based model, introduces color bias as shown in the 4th and 5th column. Vision transformer based U-Transformer
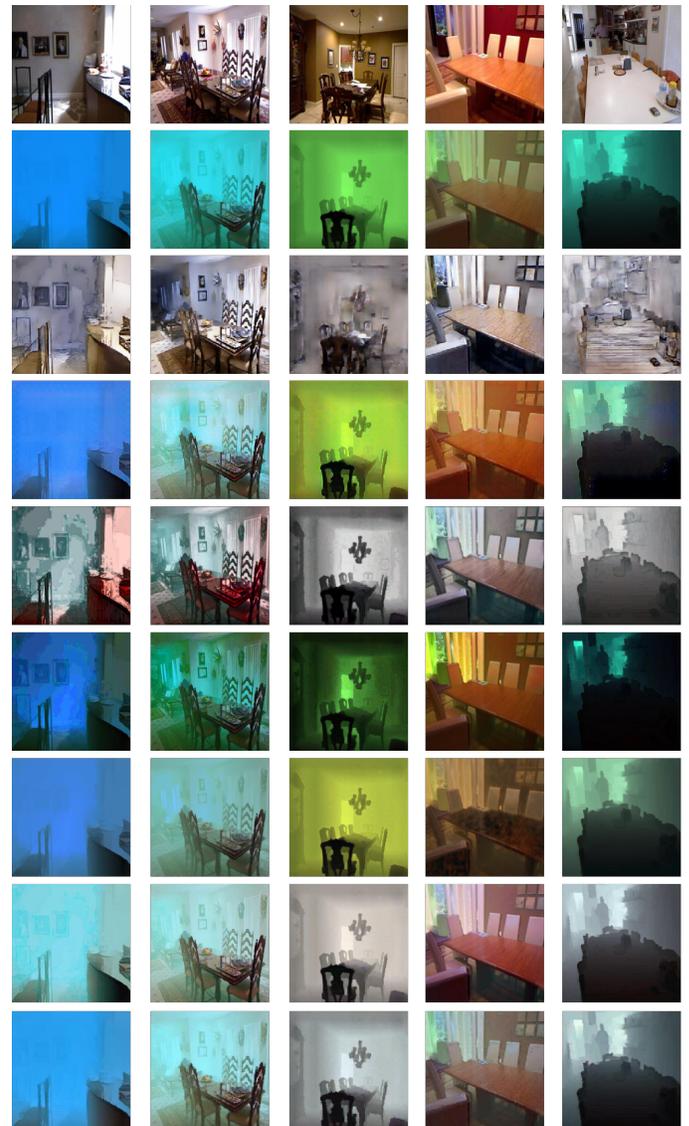

Fig. 5. Restoration Results of Artificial Underwater Images. 1st Row Shows Original Indoor Images, 2nd Row Shows Transformed Input Images, 3rd Row Shows Results of Proposed Model, 4th Row Shows FUnIE-GAN [8]. 5th Row Shows Retinex [13], 6th Row Shows UDCP [12], 7th Row Shows UWCNN [7], 8th Row Shows Water-Net [6], and 9th Row Shows U-Transformer [9].

[9] (8th row) failed to recover greenish and yellowish inputs, respectively shown in the 4th and 5th column. Among the non-learning based methods, Retinex [13] (4th row) corrects a greenish image (4th column), yet also adds reddish color bias in other images (2nd, 3rd, and 6th column). UDCP [12] (5th row), statistical method, hardly improves the overall visibility. As no ground truth is available in real underwater images, PSNR and SSIM are not computed. As real underwater images are tremendously diverse, many supervised models fail to enhance severely degraded underwater images. Among results of previous methods, better results are obtained with Water-Net [6]. Compared to Water-Net trained with a dataset less than 1000 real underwater images, our training dataset is about 100 times larger than that of Water-Net. Also, large amount of severely degraded underwater images are included [5], thus

perceptually better results are tend to be obtained with our model, as shown in the 1st column of Fig. 6.

E. Ablation Study of Loss Function

Ablation study of loss function in Eq. (1) is shown in this section. PSNR and SSIM scores per 10 water types are computed in Table IV. Results of employing only reconstruction loss L_{rec} are denoted as $L1$, plus luminance loss L_{lum} are denoted as $L1 + lum$, and all loss are denoted as ALL. Each loss functions contribute the scores in almost all water types. While proposed luminance loss L_{lum} in Eq. (4) improves less, we observe that the luminance loss stabilizes the training, as it doesn't depend on input color.

TABLE IV. ABLATION STUDY OF PROPOSED LOSS FUNCTION

Water-type		I	IA	IB	II	III	1C	3C	5C	7C	9C
PSNR	L1	16.37	16.081	16.269	14.832	15.485	16.659	15.62	15.873	15.174	15.265
	L1+lum	16.392	16.1	16.251	14.877	15.551	16.651	15.665	15.857	15.24	15.318
	ALL	16.61	16.316	16.568	15.124	15.8	16.987	15.94	16.214	15.485	15.559
SSIM	L1	0.694	0.675	0.672	0.591	0.633	0.689	0.635	0.646	0.607	0.609
	L1+lum	0.695	0.677	0.673	0.593	0.635	0.69	0.637	0.647	0.609	0.611
	ALL	0.702	0.686	0.684	0.608	0.649	0.7	0.652	0.662	0.626	0.629

V. CONCLUSION

This study tackles significantly degraded underwater image enhancement with a deep learning model incorporating discrete wavelet transform and whitening and coloring transform. The presented model is trained with the elaborated loss function and recently proposed physically revised artificial underwater image dataset. We also construct real underwater image dataset taken near the rubble mounds of port structures. The dataset characteristically includes severely degraded blueish or greenish underwater images. The presented model outperforms previous state-of-the-art underwater image enhancement models in 9 out of 10 water types in the evaluation employing an artificial underwater image dataset. Also, our model successfully removes blueish tints from real underwater images, showing splendid results qualitatively and quantitatively.

REFERENCES

[1] Y. Wang, W. Song, G. Fortino, L. Qi, W. Zhang, and A. Liotta, "An experimental-based review of image enhancement and image restoration methods for underwater imaging," *IEEE Access*, vol. 7, pp. 140 233–140 251, 2019.

[2] K. A. Skinner, J. Zhang, E. A. Olson, and M. Johnson-Roberson, "Uwstereonet: Unsupervised learning for depth estimation and color correction of underwater stereo imagery," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 7947–7954.

[3] S. Anwar and C. Li, "Diving deeper into underwater image enhancement: A survey," *Signal Processing: Image Communication*, vol. 89, p. 115978, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0923596520301478

[4] D. Akkaynak and T. Treibitz, "A revised underwater image formation model," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[5] S. Takao, "Underwater image sharpening and color correction with a dataset based on correct underwater physical model," in *Submitted*.

[6] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao, "An underwater image enhancement benchmark dataset and beyond," *IEEE Transactions on Image Processing*, vol. 29, pp. 4376–4389, 2020.

[7] C. Li, S. Anwar, and F. Porikli, "Underwater scene prior inspired deep underwater image and video enhancement," *Pattern Recognition*, vol. 98, p. 107038, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320319303401

[8] M. J. Islam, Y. Xia, and J. Sattar, "Fast underwater image enhancement for improved visual perception," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 3227–3234, 2020.

[9] L. Peng, C. Zhu, and L. Bian, "U-shape transformer for underwater image enhancement," *arXiv preprint arXiv:2111.11843*, 2021.

[10] T. Williams and R. Li, "Wavelet pooling for convolutional neural networks," in *International Conference on Learning Representations*, 2018. [Online]. Available: https://openreview.net/forum?id=rkhlb8lCZ

[11] A. Jamadandi and U. Mudenagudi, "Exemplar-based underwater image enhancement augmented by wavelet corrected transforms," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019.

[12] P. L. J. Drews, E. R. Nascimento, S. S. C. Botelho, and M. F. Montenegro Campos, "Underwater depth estimation and image restoration based on single images," *IEEE Computer Graphics and Applications*, vol. 36, no. 2, pp. 24–35, 2016.

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

[14] C. Fabbri, M. J. Islam, and J. Sattar, "Enhancing underwater imagery using generative adversarial networks," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 7159–7165.

[15] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.

[16] M. Fu, H. Liu, Y. Yu, J. Chen, and K. Wang, "Dw-gan: A discrete wavelet transform gan for nonhomogeneous dehazing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 203–212.

[17] J. Yoo, Y. Uh, S. Chun, B. Kang, and J.-W. Ha, "Photorealistic style transfer via wavelet transforms," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

[18] Y. Li, C. Fang, J. Yang, Z. Wang, X. Lu, and M.-H. Yang, "Universal style transfer via feature transforms," *Advances in neural information processing systems*, vol. 30, 2017.

[19] N. G. Jerlov, *Marine optics*. Elsevier, 1976.

[20] P. Bojanowski, A. Joulin, D. Lopez-Pas, and A. Szlam, "Optimizing the latent space of generative networks," in *Proceedings of the 35th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, J. Dy and A. Krause, Eds., vol. 80. PMLR, 10–15 Jul 2018, pp. 600–609. [Online]. Available: http://proceedings.mlr.press/v80/bojanowski18a.html

[21] H. Wu, J. Liu, Y. Xie, Y. Qu, and L. Ma, "Knowledge transfer dehazing network for nonhomogeneous dehazing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020.

[22] Haibin Ling and K. Okada, "Diffusion distance for histogram comparison," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 1, 2006, pp. 246–253.

[23] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from rgbd images," in *Computer Vision – ECCV 2012*, A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 746–760.

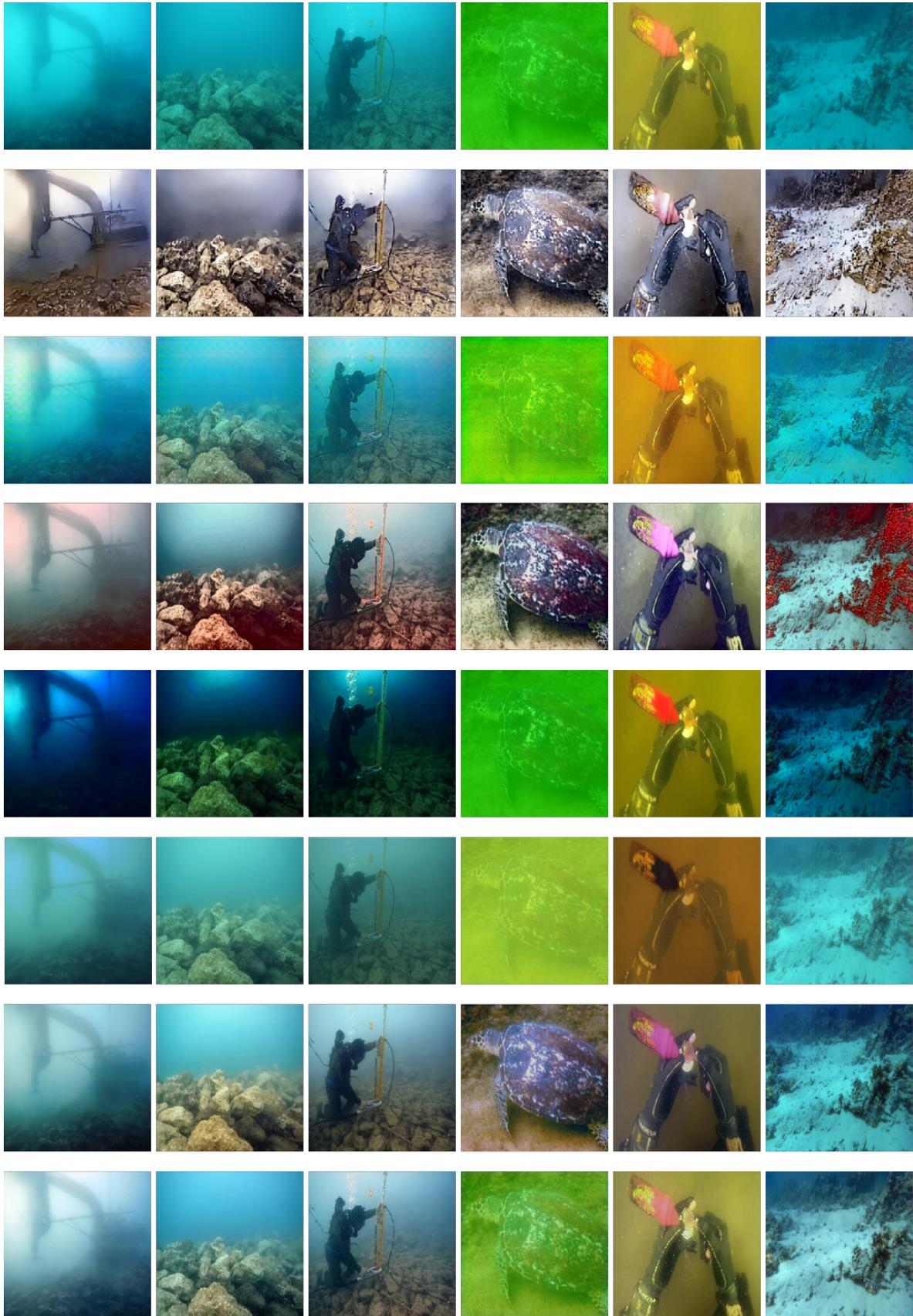[24] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *ICLR (Poster)*, 2015.

Fig. 6. Restoration Results of Real Underwater Images. 1st Row Shows Input Underwater Images, 2nd Row Shows Proposed Model, 3rd Row Shows FUnIE-GAN [8], 4th Row Shows Retinex [13], 5th Row Shows UDCP [12], 6th Row Shows UWCNN [7], 7th Row Shows Water-Net [6], and 8th Row Shows U-Transformer [9].