A WAVELET-BASED DUAL-STREAM NETWORK FOR UNDERWATER IMAGE ENHANCEMENT

Ziyin Ma and Changjae Oh

School of Electrical Engineering and Computer Science, Queen Mary University of London, UK

ABSTRACT

We present a wavelet-based dual-stream network that addresses color cast and blurry details in underwater images. We handle these artifacts separately by decomposing an input image into multiple frequency bands using discrete wavelet transform, which generates the downsampled structure image and detail images. These sub-band images are used as input to our dual-stream network that incorporates two sub-networks: the multi-color space fusion network and the detail enhancement network. The multi-color space fusion network takes the decomposed structure image as input and estimates the color corrected output by employing the feature representations from diverse color spaces of the input. The detail enhancement network addresses the blurriness of the original underwater image by improving the image details from high-frequency sub-bands. We validate the proposed method on both real-world and synthetic underwater datasets and show the effectiveness of our model in color correction and blur removal with low computational complexity.

Index Terms— Underwater image enhancement, wavelet decomposition, multi-color space, dual-stream network

1. INTRODUCTION

Underwater images and videos suffer from degradation due to the effects of light absorption and scattering, which causes blur and color casts. Image filtering tasks, e.g. restoration and enhancement, can improve the visual quality of underwater images, which can be divided into physics-based and learning-based methods [1].

Physics-based methods build a model based on the physical and optical properties of the images taken in water [2, 3, 4]. These methods investigate the physical mechanism of the degradation caused by color cast or scattering and compensate them to improve the underwater images. However, a single physics-based model cannot address all the complex physical and optical factors underlying the underwater scenes. This limitation leads to poor generalization and causes the results with over- or under-enhancement.

Learning-based methods employ the representation power of deep neural networks (DNNs) [5, 6, 7, 8]. These methods commonly suffer from a lack of training data as paired underwater and clean images are difficult to collect. This issue can be addressed by generative adversarial network (GAN) based style transfer that does not require the paired training data and can transfer the appearance of a clean image to an underwater image [9]. The training data can be augmented by generating synthetic images that include various color types, turbidity, and the illumination of water [10]. This training strategy can further consider the diversity of water types in DNNs by adding a classifier and learning domain agnostic features to restore images [11]. Multi-color space can be employed to analyze the input image in various color domains, which allows the DNN to produce
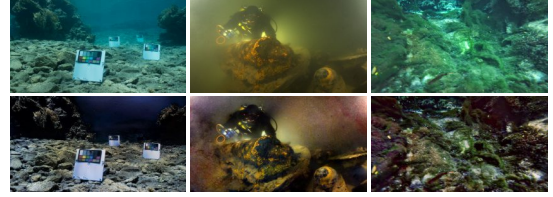
Fig. 1. (Top) Underwater images, which suffer from color cast and image blur, and (bottom) the results produced by ours.

diverse feature representations [12]. Existing learning-based methods commonly focus on elaborating the network or training strategy, but they give less attention to the underlying complex artifacts to address in underwater images.

In this paper, we propose a wavelet-based dual-stream network that performs the color correction and the detail enhancement separately by decoupling the problems of color casts and image blur in underwater images. To this end, we first decompose an input image into sub-images with multiple frequency sub-bands using discrete wavelet transform. A sub-image with low-frequency contains the image structure at half-resolution while the rest of the sub-images contain image details. To deal with these sub-band images, we present a dual-stream network that addresses the color correction to the sub-image with low-frequency and the detail enhancement to the rest of the sub-images with higher frequency, respectively. Specifically, we present a multi-color space fusion network that considers the advantage of various color spaces of input for color-cast removal and a detail enhancement network to improve the missing details. We validate our model on NYU-v2 [13], UIEB [14] and underwater color checker images (ColorChecker) [15], which shows competitive results to existing methods. The main contributions of our work are as follows[1]:

- We use wavelet decomposition to input underwater images, enabling our dual-stream network to perform the color correction and the detail enhancement, respectively.

- We present a multi-color space fusion network that incorporates the various color representations for the color cast removal.

- Our model has the advantages in avoiding the artificial color and blurry outputs with low computational complexity.

2. PROPOSED METHOD

Given an underwater image, I, we aim to learn a network, $f(\cdot)$, to generate an enhancement output, $f(I)$, that removes the color cast from I while enhancing the image details. We first use discrete

---

[1]More results are available here: `https://zziyin.github.io/UIE-WD/demo.html`
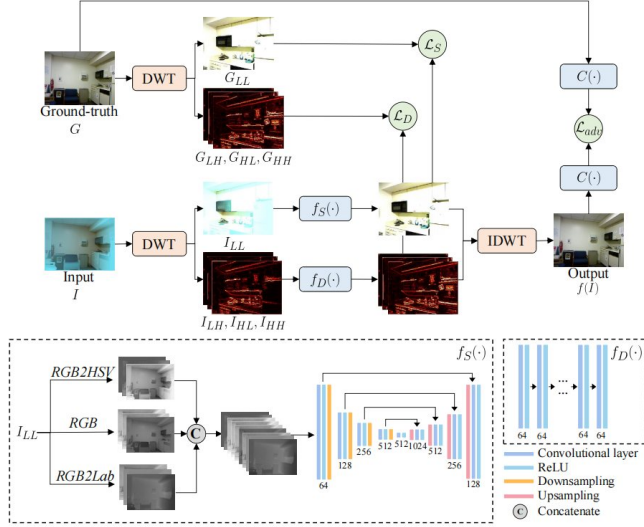
Fig. 2. Overview of our method, trained on the synthetic dataset. The sub-band images with multiple frequency bands are obtained by discrete wavelet transform (DWT), which facilitates to decouple the color cast and blurry details in underwater images and separately address these artifacts with $f_S(\cdot)$ and $f_D(\cdot)$, respectively. $f_S(\cdot)$ and $f_D(\cdot)$ are constrained by the structure loss, $\mathcal{L}_S$, and the detail loss, $\mathcal{L}_D$, respectively. The output, $f(I)$, after inverse DWT (IDWT), is constrained by the adversarial loss, $\mathcal{L}_{adv}$, with the generative adversarial network (GAN) discriminator, $C(\cdot)$.

wavelet transform (DWT) to decompose I into sub-band images that consist of an approximated original image at half-resolution, I_{LL}, and the images with high-frequency components into vertical I_{LH}, horizontal I_{HL}, and diagonal I_{HH} directions. These sub-band images are then fed to the dual-stream network, i.e. the multi-color space fusion network, $f_S(\cdot)$, and the detail enhancement network, $f_D(\cdot)$, in which the color-cast removal and detail enhancement are separately processed. $f_S(\cdot)$ aims to remove the color cast in I_{LL} by representing the input in the multiple color spaces, while $f_D(\cdot)$ aims to generate the detail enhancement output from I_{LH}, I_{HL}, and I_{HH}. These sub-band estimations are then integrated and reconstructed to the original size using inverse DWT (IDWT). We jointly train our model with the structure loss, $\mathcal{L}_S$, detail loss, $\mathcal{L}_D$, and adversarial loss, $\mathcal{L}_{adv}$. Due to the lack of clean-underwater image pairs, the model is trained on the synthetic dataset [8, 13]. Fig. 2 shows the overall pipeline of our method.

2.1. Discrete wavelet decomposition

Wavelet transform has been applied to various image processing tasks such as image super-resolution [16] and denoising [17]. Several physics-based methods in underwater image enhancement have used DWT to decompose the images and process them in the frequency domain [18, 19] to improve the contrast and resolution. In our framework, we use DWT to decompose an input image into multiple frequency sub-bands so that the color correction and detail enhancement can be separately addressed.

We decompose an input using Haar wavelets that consist of the low-pass filter, L, and the high-pass filter, H, as follows:

$$L = \frac{1}{\sqrt{2}}[1,1]^T, H = \frac{1}{\sqrt{2}}[1,-1]^T. \quad (1)$$
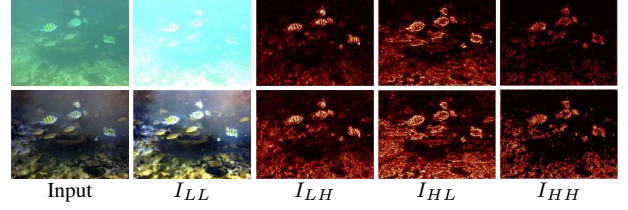
Fig. 3. Decomposition results of the (top) underwater image and (bottom) enhancement output after 1-scale DWT. Note that input images are downsampled to the half resolution, and the normalized absolute values are color-coded for I_{LH}, I_{HL} and I_{HH}, where the pixel value increases from black to red to yellow.

We obtain four sub-band images $I_{LL}, I_{LH}, I_{HL}, I_{HH}$ by conducting convolution and downsampling on image I. I_{LL} is obtained by using a low-pass filter LL^T to horizontal and vertical directions. The other three sub-band images, I_{LH}, I_{HL}, I_{HH}, are obtained by using the filters LH^T, HL^T, HH^T, that captures the high-frequency components in vertical, horizontal, and diagonal directions. The sub-band images are downsampled to half-resolution of the original input but do not result in information loss due to the biorthogonal property of DWT. DWT can be seen as a convolution process on the input I, using four 2×2 convolution kernels with fixed weights and with the stride of 2, and IDWT can be seen as the transposed convolution. The proposed network thus can be trained end-to-end.

Fig. 3 shows the original images and their sub-band images obtained by DWT. We can observe that I_{LL} includes most of the global image structure with the color cast, at the half-resolution, while I_{LH}, I_{HL}, I_{HH} contain image details captured from different directions. To fully exploit these sub-band images by considering their property, we present a dual-stream network that separately processes the sub-band images for enhancement.

2.2. Dual-stream network

The proposed dual-stream network consists of two sub-networks, multi-color space fusion network, $f_S(\cdot)$, and detail enhancement network, $f_D(\cdot)$, which separately estimate the structure and detail component of wavelet sub-images, respectively.

For $f_S(\cdot)$, we adopt U-net [20] as our base architecture. Considering that the diverse optical factors in underwater cause the various color casts, we present a multi-color space fusion module to analyze the input image with various color representations. In addition to using the common RGB color space for an input image, we transform the RGB input to HSV and Lab color spaces and then concatenate them into a 9-channel image to further extract diverse feature representations. As shown in Fig. 4, the HSV color space can directly reflect the brightness and contrast of the image. The Lab color space approximates the human visual system and provides perceptual uniformity [12]. We then pass the concatenated image with the multi-color spaces to the network to estimate the clean image without color cast. Moreover, such non-linear color transforms can greatly improve the performance of the network without making the network deeper [12]

For $f_D(\cdot)$, we aim to address the image blur by decoupling the input from the color cast and focusing only on the image details to improve. The inputs with high-frequency bands, I_{LH}, I_{HL}, and I_{HH}, pass $f_D(\cdot)$ that consists of 10 convolution layers followed by the ReLU function. The filter size and the depth of each layer are 3×3 and 64, respectively. With $f_D(\cdot)$, the missing details can be improved as shown in Fig. 3.
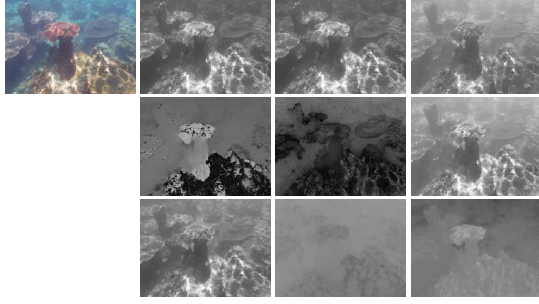
Fig. 4. Images from different color spaces. (Top) An RGB image and its R, G, B, (middle) H, S, V, and (bottom) L, a, b components. The darker the intensity, the smaller the value.

After reconstructing the final output $f(I)$ from $f_S(I_{LL})$, $f_D(I_{LH})$, $f_D(I_{HL})$ and $f_D(I_{HH})$ using IDWT, we further perform the adversarial learning for higher accuracy [21]. We add a GAN discriminator, $C(\cdot)$, which is trained to distinguish the ground truth G from $f(I)$. This GAN framework encourages our network to generate more realistic images to mislead $C(\cdot)$.

2.3. Loss functions

We train the network with loss functions that consist of three parts: structural loss, $\mathcal{L}_S$, the detail loss $\mathcal{L}_D$, and the Wasserstein GAN loss, $\mathcal{L}_{adv}$ as:

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_S + \lambda_2 \mathcal{L}_D + \lambda_3 \mathcal{L}_{adv}, \qquad (2)$$

where λ_1, λ_2, and λ_3 control the effect of the loss functions.

Considering that the proposed dual-stream architecture has two branches, we adopt individual loss functions for each network to update them respectively. For the multi-color space fusion network, $f_S(\cdot)$, we use the Multi-scale Structural Similarity [22] (MS-SSIM) loss, $\mathcal{L}_{\text{MS-SSIM}}$, and ℓ_1 loss, $\mathcal{L}_1$, that can effectively capture the contrast and color errors, respectively:

$$\mathcal{L}_{\text{MS-SSIM}}(I_{LL}) = 1 - \text{MS-SSIM}(f_S(I_{LL}), G_{LL}), \qquad (3)$$

$$\mathcal{L}_1(I_{LL}) = \|f_S(I_{LL}) - G_{LL}\|_1, \qquad (4)$$

where G_{LL} is the LL image obtained by applying DWT to the ground truth. The structure loss $\mathcal{L}_S$ is then defined by combining $\mathcal{L}_1$ and $\mathcal{L}_{\text{MS-SSIM}}$ with the hyperparameter α:

$$\mathcal{L}_S(I_{LL}) = \alpha \cdot \mathcal{L}_{\text{MS-SSIM}} + (1 - \alpha) \cdot \mathcal{L}_1. \qquad (5)$$

For the detail enhancement network, $f_D(\cdot)$, we measure the ℓ_2 error between the estimated detail sub-image, $f_D(I_{(i)})$, from $I_{(i)}$ with component $i \in \{LH, HL, HH\}$ and its ground truth, $G_{(i)}$:

$$\mathcal{L}_D(I_{(i)}) = \|f_D(I_{(i)}) - G_{(i)}\|_2. \qquad (6)$$

We further use an adversarial loss, $\mathcal{L}_{adv}$, for the final output, $f(I)$. We employ the loss in Wasserstein GAN [21] that shows stable performance in training. Our network, as a generator, aims to minimize the Wasserstein distance between the real and the generated distribution calculated by the discriminator $C(\cdot)$.

3. VALIDATION

3.1. Setup

Our network, implemented with PyTorch, was trained and tested on Intel Core i5-7200 CPU and NVIDIA Geforce RTX 2070 GPU.

Table 1. UIQM, UCIQE, and CIE2000 scores of physics-based (P) and learning-based (L) methods and the average test time per image. The 1st and 2nd best results are in bold and underline, respectively. ↑: The higher, the better, ↓: the lower, the better.

Methods		UIEB [14]			ColorChecker [15]	
Name	P/L	UIQM↑	UCIQE↑	Time(s)	CIE2000↓	Time(s)
ULAP [2]	P	7.213	5.832	5.09	39.98	5.69
IBLA [3]	P	7.385	6.129	38.71	41.35	38.54
UDCP [4]	P	5.648	**8.523**	15.23	39.93	17.98
UWCNN [7]	L	7.791	4.769	3.31	<u>38.21</u>	4.40
WaterNet [8]	L	**8.818**	4.465	1.56	38.63	2.25
UIE-DAL [11]	L	<u>8.376</u>	6.254	<u>0.10</u>	40.94	<u>0.27</u>
Ours	L	8.032	<u>6.341</u>	**0.08**	**37.22**	**0.26**

Fig. 5. Visual samples of underwater image enhancement from (a-c) UIEB [14] and (d) ColorChecker [15]. UIQM and UCIQE scores (the higher, the better) are presented on the top-left and the CIE2000 score (the lower, the better) is presented on the bottom-right in (d).

We used RMSProp [23] as an optimizer with a batch size of 4 and set learning rate to 0.0005 for $f_S(\cdot)$, and 0.00002 for $f_D(\cdot)$. The weights, λ_1, λ_2, λ_3, α, were set to 0.5, 1, 1, 0.5, respectively. We first trained the dual-stream network and then added the discriminator into the training process. The discriminator was trained 5 epochs for each epoch of the training for the dual-stream network.

Due to the lack of access to the ground truth for real underwater images, we followed the training protocol in [8] to train our model. We used a synthetic underwater dataset produced from NYU-v2 dataset [13] with 1449 indoor images. We augmented each image to 36 images with 6 different Jelov water types and 6 different levels of background light, and selected 20k images as our training dataset. For testing, we used both synthetic and real-world underwater images, including the NYU-v2 with 3000 synthetic images, the UIEB

Table 2. Ablation study on the loss functions.

Loss functions			NYU-v2 [13]	UIEB [14]	
$\mathcal{L}_{\text{MS-SSIM}}$	$\mathcal{L}_1$	$\mathcal{L}_{Adv}$	SSIM↑	UIQM↑	UCIQE↑
✓			0.882	6.551	6.084
	✓		0.884	6.604	**6.812**
✓	✓		0.896	7.341	6.061
✓	✓	✓	**0.918**	**8.032**	6.341

| $\mathcal{L}_1$ | $\mathcal{L}_{\text{MS-SSIM}}$ | $\mathcal{L}_1+\mathcal{L}_{\text{MS-SSIM}}$ | $\mathcal{L}_{total}$ |

Fig. 6. Results from the models trained with different loss functions.

dataset [14] with 890 real underwater images, and the ColorChecker dataset [15] collected by Olympus Tough 6000 with 7 images.

For quantitative comparisons of underwater images, we used UIQM [24], UCIQE [25], and CIE2000 [26] metrics. UIQM considers the colorfulness, sharpness, and contrast while UCIQE evaluates the chroma, saturation, and contrast of images. CIE2000 measures the color difference between two images, which can be used for the color checker images. We also employ the SSIM to evaluate the effectiveness of our network in the ablation studies, which can be measured on the synthetic dataset [13] with the ground-truth images.

3.2. Results and comparisons

Fig. 5 shows visual results generated by existing methods and ours on the UIEB dataset [8] and the underwater ColorChecker [15]. Physics-based methods, ULAP [2], IBLA [3], UDCP [4] improve the brightness and contrast well, but fail to remove severe color cast and blurring. Existing learning-based methods, Water-Net [8], UWCNN [7], UIE-DAL [11] can deal with the color cast and blurring issues, but the artifacts are observed in the color checkers (UWCNN, UIE-DAL) or the contrast is low (UWCNN, WaterNet). Our network produces the images by effectively removing the color cast with the improved details while avoiding unnatural artifacts.

Table 1 shows the quantitative results, where the learning-based approaches generally show better scores than the physics-based approaches. For UCIQE scores, except for UDCP, ours shows higher scores than other methods. Although UDCP has a higher UCIQE value, the visual results show low intensities in green and blue channels (Fig. 5(a)-(c)), which makes the global tone relatively red (Fig. 5(d)). The physics-based methods show limited performance in heavily blurred images, e.g., Fig. 5(b), thus resulting in a low sharpness. Our method outperforms others in detail enhancement, which makes our UIQM values generally higher in heavily blurred images. To better compare the color correction, we analyze the performance of color constancy using the images taken underwater with the standard color checker [15]. Our model obtains the lowest CIE2000 score, indicating that the colors in the color checker are the closest color to the real ones. We also measure the average testing time for each method on the UIEB dataset and ColorChecker. The physics-based methods take a longer time to enhance the image, while the learning-based methods take less time. Our network, thanks to image decomposition, uses the smaller size of inputs and thus shows low computational complexity.

3.3. Ablation study

Loss functions. We examine the loss functions by training $f_S(\cdot)$ with different settings: $\mathcal{L}_1$, $\mathcal{L}_{\text{MS-SSIM}}$, and $\mathcal{L}_1+\mathcal{L}_{\text{MS-SSIM}}$. Table 2 and

Table 3. Ablation study on the effect of discrete wavement transform (DWT), $f_D(\cdot)$, multi-color space (MCS), and generative adversarial network (GAN).

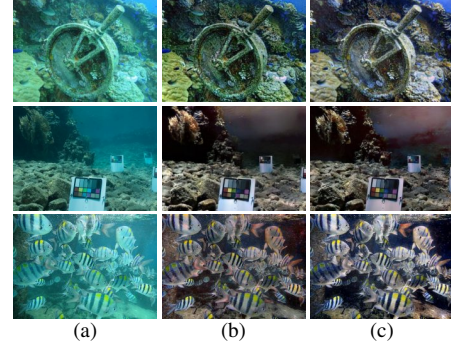| Modules				NYU-v2 [13]	UIEB [14]	
DWT	$f_D(\cdot)$	MCS	GAN	SSIM↑	UIQM↑	UCIQE↑
✓				0.841	7.006	5.283
	✓			0.852	7.125	5.647
✓	✓			0.854	7.095	5.426
✓	✓	✓		0.872	7.341	6.061
✓	✓	✓	✓	**0.911**	**8.032**	**6.341**

(a)   (b)   (c)

Fig. 7. Ablation study for each module. (a) Inputs, (b,c) the results without and with (top) discrete wavelet transform, (middle) $f_D(\cdot)$, and (bottom) multi-color space.

Fig. 6 show the results of the ablation study. $\mathcal{L}_1+\mathcal{L}_{\text{MS-SSIM}}$ shows less visual artifacts than the results from $\mathcal{L}_1$ and $\mathcal{L}_{\text{MS-SSIM}}$. The result from $\mathcal{L}_{\text{MS-SSIM}}$ shows high contrast, which leads to a high UCIQE score. However, the contrast enhancement is only focused near the object. Finally, the result from $\mathcal{L}_{\text{total}}$ shows a less noisy and more colorful output. The quantitative results also show that $\mathcal{L}_{\text{total}}$ achieves the highest SSIM and UIQM.

Network configurations. We examine each component of our model. Fig. 7 and Table 3 show the results of the ablation study. We compare our results with a dual-stream network [27] that uses an RGB image for both sub-networks. As shown in the 1st row of Fig. 7, our network can address the underwater image enhancement better by separately addressing the color correction and detail enhancement. To see the effect of $f_D(\cdot)$ for detail enhancement, we compare our results, $f(I)$, with the images reconstructed by $f_S(I_{LL})$ and the original details, I_{LH}, I_{HL}, and I_{HH}. We can observe the advantage of $f_D(\cdot)$ as shown in the 2nd row of Fig. 7 and Table 3. We further test $f_S(\cdot)$ by replacing this model with the vanilla U-net [20], which shows the advantage of multi-color space in $f_S(\cdot)$ with the better contrast and color appearance.

4. CONCLUSION

We presented a method to decouple the various artifacts in underwater image enhancement by using DWT, which can separately perform color correction and detail enhancement with the proposed dual-stream network. The key idea is that the sub-band image with low-frequency contains the color cast and the rest of sub-band images with high-frequency contain blurry details. We showed that our model can effectively remove the color cast and improve the blurry details. The test time is reduced by decomposing the input into smaller sizes. Future work includes a subjective evaluation of our results and investigating a metric for visual quality assessment of underwater images.

5. REFERENCES

[1] C.Y. Li, R. Mazzon, and A. Cavallaro, "Underwater image filtering: methods, datasets and evaluation," *arXiv preprint arXiv:2012.12258*, 2020.

[2] W. Song, Y. Wang, D. Huang, and D. Tjondronegoro, "A rapid scene depth estimation model based on underwater light attenuation prior for underwater image restoration," in *Proceedings of the Pacific Rim Conference on Multimedia*, 2018.

[3] Y.T. Peng and P.C. Cosman, "Underwater image restoration based on image blurriness and light absorption," *IEEE Transactions on Image Processing (TIP)*, vol. 26, no. 4, pp. 1579–1594, 2017.

[4] P. Drews, E. Nascimento, F. Moraes, S. Botelho, and M. Campos, "Transmission estimation in underwater single images," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2013.

[5] N. Wang, Y. Zhou, F. Han, H. Zhu, and J. Yao, "Uwgan: underwater gan for real-world underwater color restoration and dehazing," *arXiv preprint arXiv:1912.10269*, 2019.

[6] J. Li, K.A. Skinner, R.M Eustice, and M. Johnson-Roberson, "Watergan: Unsupervised generative network to enable real-time color correction of monocular underwater images," *IEEE Robotics and Automation letters (RA-L)*, vol. 3, no. 1, pp. 387–394, 2017.

[7] C. Li, S. Anwar, and F. Porikli, "Underwater scene prior inspired deep underwater image and video enhancement," *Pattern Recognition (PR)*, vol. 98, pp. 107038, 2020.

[8] S. Anwar, C. Li, and F. Porikli, "Deep underwater image enhancement," *arXiv preprint arXiv:1807.03528*, 2018.

[9] C. Fabbri, M.J. Islam, and J. Sattar, "Enhancing underwater imagery using generative adversarial networks," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2018.

[10] C.Y. Li and A. Cavallaro, "Cast-gan: Learning to remove colour cast from underwater images," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 2020.

[11] P.M. Uplavikar, Z. Wu, and Z. Wang, "All-in-one underwater image enhancement using domain-adversarial learning.," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019.

[12] C. Li, S. Anwar, J. Hou, R. Cong, C. Guo, and W. Ren, "Underwater image enhancement via medium transmission-guided multi-color space embedding," *IEEE Transactions on Image Processing (TIP)*, vol. 30, 2021.

[13] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from rgbd images," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2012.

[14] 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 (TIP)*, vol. 29, pp. 4376–4389, 2019.

[15] C.O. Ancuti, C. Ancuti, C. De Vleeschouwer, and P. Bekaert, "Color balance and fusion for underwater image enhancement," *IEEE Transactions on Image Processing (TIP)*, vol. 27, no. 1, pp. 379–393, 2018.

[16] H. Huang, R. He, Z. Sun, and T. Tan, "Wavelet-srnet: A wavelet-based cnn for multi-scale face super resolution," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017.

[17] E. Kang, W. Chang, J. Yoo, and J.C. Ye, "Deep convolutional framelet denosing for low-dose ct via wavelet residual network," *IEEE Transactions on Medical Imaging (T-MI)*, vol. 37, no. 6, pp. 1358–1369, 2018.

[18] S.R. Singh, "Enhancement of contrast and resolution of gray scale and color images by wavelet decomposition and histogram shaping and shifting," in *Proceedings of the International Conference on Medical Imaging, m-Health and Emerging Communication Systems*, 2014.

[19] R. Priyadharsini, T.S. Sharmila, and V. Rajendran, "A wavelet transform based contrast enhancement method for underwater acoustic images," *Multidimensional Systems and Signal Processing*, vol. 29, no. 4, pp. 1845–1859, 2018.

[20] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-assisted Intervention (MICCAI)*, 2015.

[21] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proceedings of the International Conference on Machine Learning (ICML)*, 2017.

[22] Z. Wang, E.P. Simoncelli, and A.C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proceedings of The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers*, 2003.

[23] T. Tieleman and G. Hinton, "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude," *COURSERA: Neural networks for machine learning*, vol. 4, no. 2, pp. 26–31, 2012.

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

[25] M. Yang and A. Sowmya, "An underwater color image quality evaluation metric," *IEEE Transactions on Image Processing (TIP)*, vol. 24, no. 12, pp. 6062–6071, 2015.

[26] G. Sharma, W. Wu, and E.N. Dalal, "The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations," *Color Research & Application*, vol. 30, no. 1, pp. 21–30, 2005.

[27] J. Pan, S. Liu, D. Sun, J. Zhang, Y. Liu, J. Ren, Z. Li, and et al., "Learning dual convolutional neural networks for low-level vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.