Underwater Image Enhancement via Learning Water Type Desensitized Representations

Zhenqi Fu, Xiaopeng Lin, Wu Wang, Yue Huang, and Xinghao Ding

***Abstract*—For underwater applications, the effects of light absorption and scattering result in image degradation. Moreover, the complex and changeable imaging environment makes it difficult to provide a universal enhancement solution to cope with the diversity of water types. In this letter, we present a novel underwater image enhancement (UIE) framework termed SCNet to address the above issues. SCNet is based on normalization schemes across both spatial and channel dimensions with the key idea of learning water type desensitized features. Considering the diversity of degradation is mainly rooted in the strong correlation among pixels, we apply whitening to de-correlates activations across spatial dimensions for each instance in a mini-batch. We also eliminate channel-wise correlation by standardizing and re-injecting the first two moments of the activations across channels. The normalization schemes of spatial and channel dimensions are performed at each scale of the U-Net to obtain multi-scale representations. With such latent encodings, the decoder can easily reconstruct the clean signal, and unaffected by the distortion types caused by the water. Experimental results on two real-world UIE datasets show that the proposed approach can successfully enhance images with diverse water types, and achieves competitive performance in visual quality improvement.**

***Index Terms*—Underwater image enhancement, whitening, normalization, water types**

I. Introduction

UNDERWATER optical vision is a critical perception component for marine research and underwater robotics. For example, underwater surveillance systems (USS) and autonomous underwater vehicles (AUV) rely on high-quality images to fulfill their objectives. Scientists also need clear underwater images to study deteriorating coral reefs and other aquatic life [1], [2]. Unfortunately, the quality of images acquired for these applications is commonly degraded due to various influences. One of the major factors is wavelength-dependent light attenuation (i.e., absorption and scattering) over the depth of objects in the scene. The absorption effect is caused by the fact that the red light is absorbed at a higher rate than green and blue in the water. Hence, images recorded in water scenes are always dominated by bluish or greenish tint. The scattering phenomenon (including forward-scattering and backward-scattering) stems from suspending particles present, which diminishes the image quality by introducing a homogeneous background noise and haze-like appearance.

Apart from the light attenuation, another challenge in underwater image enhancement (UIE) is the diversity of water types (i.e., degradation distributions). As presented in Fig. 1, underwater scenes captured in diverse water types (e.g., shallow coastal waters, deep oceanic waters, and muddy waters) show a significant difference in appearances and styles. Normally, it is difficult for a single model to enhance underwater images with such multiple distributions. In other words, providing a universal solution for UIE is challenging.

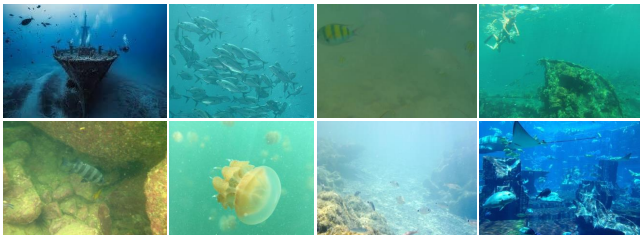

Fig. 1. Underwater scenes captured in diverse water types show a significant difference in appearances and styles.

Although many UIE approaches have been proposed, such as model-free methods [3]–[6], prior-based methods [7]–[11], and data-driven approaches [12]–[15], few of them consider the challenges of distribution diversity explicitly. To improve the image quality meanwhile handle the diversity of water types, in literature [15], the authors first synthesize ten different underwater image datasets. Then, they train UIE models for each water type. However, this approach seems inefficient and relies on the prior knowledge of the water type for the given image. Since we do not know the water type ahead of time, in literature [11], the authors try different parameter sets out of an existing library of water types (e.g., Jerlov water types). Each set leads to a different restored image and the one that best satisfies the Gray-World assumption is chosen as the final output. Similarly, this approach is also inefficient because it must perform ten times for each image to be enhanced. Additionally, it is unreliable to select the best result by a simple Gray-World based quality assessment metric. Recently, Uplavikar et al. [16] learned domain agnostic features for multiple water types and generated clean versions from those features. Concretely, they additionally utilize a network to classify the water type of a given image from U-Net's [17] latent vectors. Then, they leverage an adversarial loss to force the classifier is unsure of the possible water types. Finally, water type agnostic representations can be learned by the encoder.

In this letter, we develop a novel framework for UIE based on learning multi-scale water type desensitized representations. Generally, the degradation diversity of underwater

Zhenqi Fu, Xiaopeng Lin, Wu Wang, Yue Huang, and Xinghao Ding, are with the School of Informatics Xiamen University, Xiamen, 361005, China (e-mail: fuzhenqi@stu.xmu.edu.cn, 23320201154005@stu.xmu.edu.cn, wangwu@stu.xmu.edu.cn, yhuang2010@xmu.edu.cn, dxh@xmu.edu.cn).

images is introduced by both absorption and scattering effects. The former makes the captured image presents different colors. While the latter blurs the edges of objects and leads to various degrees of haze-like effects. To drastically eliminate the degradation diversity, we propose SCNet, which normalizes activations across both spatial and channel dimensions. Concretely, we apply instance whitening to de-correlate features across spatial dimensions. Meanwhile, we remove the first two moments of the activations across channel dimensions in the encoder, and then we transform and send them to the decoder. Note that, normalizing channel-wise activations aims at enhancing the structural information of blurred images [19]. By normalizing the encodings across both spatial and channel dimensions, SCNet can get rid of the impact of water types. As a result, the decoder can reconstruct the clean image more accurately and easily. In summary, this letter introduces the following contributions:

1) We present a normalization guide deep learning framework to address the issue of degradation diversity during the UIE. Instead of designing complex network architecture, we perform normalization schemes at each scale of a simple U-Net to learn multi-scale water type desensitized representations.
2) To obtain better enhancement performance, we not only de-correlates the activations across spatial dimensions but also normalizes and re-injects the first two moments of the activation across channel dimensions.
3) Experimental results demonstrate that our approach outperforms the previous methods significantly in both improving the visual quality and dealing with the diversity of water types.

II. Approach

Our solution is based upon normalization schemes, which standardize and whiten data using the extracted statistics. As shown in Fig. 2, we combine normalization methods with a U-Net to learn water type desensitized representations. On each scale of the U-Net, we perform spatial-wise and channel-wise normalization simultaneously. Thus activations are normalized across both spatial and channel dimensions. Normalized activations are style and appearance irrelevant. Consequently, the decoder can easily reconstruct the clear signal with better quality.

A. Spatial-wise Normalization (SN)

Spatial-wise normalization is performed via instance whitening [18] to eliminate the influence of diverse water types and discard the extracted statistics across spatial dimensions. We propose to adopt instance whitening to normalize features because the appearance of an individual image can be well encoded by the covariance matrix. In our method, SN is performed in each U-net's skip-connection. Let $\mathbf{X} \in \mathbb{R}^{C\times NHW}$ refer to the data matrix of a mini-batch, where, C, N, H, W indicate the number of channels, number of instances, height, and width respectively. Here, N, H, and W are viewed as a single dimension for convenience. Let matrix $\mathbf{X}_n \in \mathbb{R}^{C\times NHW}$ be the n-th instance in the mini-batch, where $n \in \{1, 2, \ldots, N\}$. Then the whitening transformation Γ for an instance $\mathbf{X}_n$ can be formulated as:

$$\Gamma\left(\mathbf{X}_n\right) = \mathbf{\Sigma}^{-1/2}\left(\mathbf{X}_n - \boldsymbol{\mu}\right) \tag{1}$$

where $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ denote the mean vector and the covariance matrix computed from the data. Specifically, for instance whitening, $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ are calculated within each individual sample by:

$$\boldsymbol{\mu} = \frac{1}{HW}\mathbf{X}_n \tag{2}$$

$$\mathbf{\Sigma} = \frac{1}{HW}\left(\mathbf{X}_n - \boldsymbol{\mu}\right)\left(\mathbf{X}_n - \boldsymbol{\mu}\right)^T + \alpha\mathbf{I} \tag{3}$$

where α is a small positive number to prevent a singular $\mathbf{\Sigma}$. In this way, the whitening transformation Γ whitens each instance separately (i.e., $\Gamma\left(\mathbf{X}_n\right)\Gamma(\mathbf{X}_n)^T = \mathbf{I}$). Note that, in the covariance matrix $\mathbf{\Sigma}$, the diagonal elements are the variance for each channel, while the off-diagonal elements are the correlation between channels. Therefore, Eq. (1) can not only standardize but also de-correlates activations.

To enhance the representation capacity, we add scale and shift operations for instance whitening. Thus, Eq. 1 can be rewritten as:

$$\Gamma\left(\mathbf{X}_n\right) = \mathbf{\Sigma}^{-1/2}\left(\mathbf{X}_n - \boldsymbol{\mu}\right)\gamma + \beta \tag{4}$$

where γ and β are learnable parameters denoting the scale and shift operations respectively.

B. Channel-wise Normalization (CN)

After SN, the encoded features are insensitive to water types in spatial dimensions, but channel-wise correlation still exists. Channel-wise statistics are position-dependent and reveal the structural information. Diverse water types lead to various degrees of scattering effects, which blur the edge and reduce the visibility of important objects. Therefore, it is necessary to further eliminate the degradation diversity caused by the scattering effects and improve the visibility of underwater scenes. Motivated by [19], we first remove the mean and standard deviation across channels in U-Net's encoder. Then the removed mean and standard deviation are transformed via 1×1 convolutional operator to generate optimized statistics. Finally, we re-inject them into the encoder layers to transfer channel-wise information. Similar with the notation definition in SN, let $\mathbf{X} \in \mathbb{R}^{HW\times NC}$ be the data matrix of a mini-batch. Let matrix $\mathbf{X}_n \in \mathbb{R}^{HW\times NC}$ be the n-th sample in the mini-batch, where $n \in \{1, 2, \ldots, N\}$. Then the channel-wise normalization Ω for a sample $\mathbf{X}_n$ can be calculated as:

$$\Omega\left(\mathbf{X}_n\right) = \frac{\mathbf{X}_n - \boldsymbol{\mu}}{\boldsymbol{\sigma}} \tag{5}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are the mean and standard deviation vectors. For individual instance, $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are calculated by:

$$\boldsymbol{\mu} = \frac{1}{C}\mathbf{X}_n \tag{6}$$

$$\boldsymbol{\sigma} = \sqrt{\frac{1}{C}\sum\left(\mathbf{X}_n - \boldsymbol{\mu}\right)^2} + \alpha\mathbf{I} \tag{7}$$

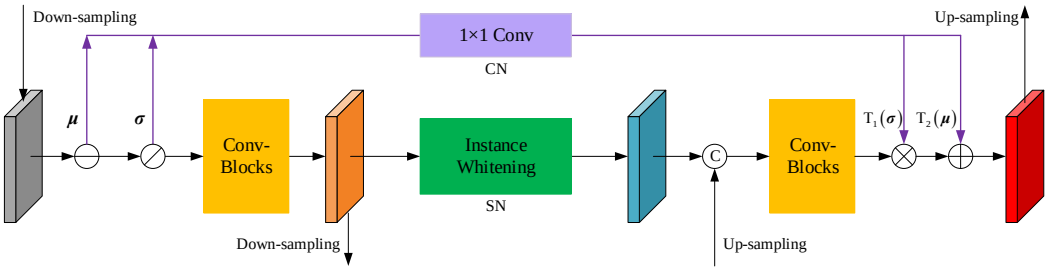

Fig. 2. Illustration of normalization across spatial and channel dimensions at a single scale of the U-Net. The normalized activations are style and appearance irrelevant. The decoder can easily reconstruct the clear signal.

where α is a small positive number to prevent a singular $\boldsymbol{\sigma}$. As mentioned before, after removing the mean and standard deviation across channel dimensions, we transform and re-inject them into the corresponding decoder layer. To be specific, the mean is added to the features, and the standard deviation is multiplied, which can be written as:

$$\mathbf{y} = \mathrm{T}_1(\boldsymbol{\sigma})\,\mathbf{x} + \mathrm{T}_2(\boldsymbol{\mu}) \tag{8}$$

where T_1 and T_2 denote the transformation functions for $\boldsymbol{\sigma}$ and $\boldsymbol{\mu}$ respectively.

C. Loss Function

Given a training set $\{\mathbf{u}_{raw}, \mathbf{u}_{gt}\}$, where $\mathbf{u}_{raw}$ indicates the raw underwater instances and $\mathbf{u}_{gt}$ refer to the corresponding clear versions. We adopt mean squared error (MSE) and perceptual similarity (PS) [20] for training our network. MSE is calculated based on pixel-wise difference:

$$\ell_{\mathrm{MSE}} = \frac{1}{n}\sum\left(\hat{\mathbf{u}}_{gt} - \mathbf{u}_{gt}\right)^2 \tag{9}$$

where $\hat{\mathbf{u}}_{gt}$ denotes the enhanced output. n is the number of pixels. The perceptual similarity assesses a solution concerning perceptually relevant characteristics. Here, the perceptual similarity is defined as the euclidean distance between the feature representations of enhanced images and clear instances. It can be formulated as follows:

$$\ell_{\mathrm{PS}} = \frac{1}{m}\sum\left(\varphi_{i,j}\left(\hat{\mathbf{u}}_{gt}\right) - \varphi_{i,j}\left(\mathbf{u}_{gt}\right)\right)^2 \tag{10}$$

where $\varphi_{i,j}$ indicates the feature map obtained by the j-th convolution (after activation) before the i-th max-pooling layer within the pr-trained VGG16 network [21]. m is the number of pixels of all feature map extracted. The overall loss function consists of two components and is minimized during the network training. It is expressed as:

$$\ell_{\mathrm{all}} = \ell_{\mathrm{MSE}} + \lambda\ell_{\mathrm{PS}} \tag{11}$$

where λ denotes the weight.

III. Experiment

We employ the real-world UIE dataset (UIEBD) [14] to train and test our model. This dataset contains 890 real-world underwater images and corresponding high-quality reference instances. We use the first 700 images for training and the rest for testing. We adopt the PyTorch framework to train our network using Adam solver with an initial learning rate of 1e-4. The mini-batch size is set as 1 empirically. The patch size is 128×128. We compare the proposed method with several state-of-the-art UIE methods including two model-free approaches (CLAHE [3] and Fusion [6]), two prior-based approaches (Histogram [10] and ULAP [22]), and two data-driven approaches (DuwieNet [14] and GLCHE [23]). We use two objective quality assessment metrics (i.e., SSIM [24] and PSNR) as the performance criteria. A better UIE approach should have higher SSIM/PSNR scores. Besides, we test the model performance on another real-world dataset (RUIE) [2] to further demonstrate the superiority of learning water type desensitized representations. Note that RUIE dataset does not contain reference images. Therefore, only qualitative results are presented on this dataset.

TABLE I
Performance of different methods on the UIEBD dataset. The best results are in bold.

Criteria	SSIM	PSNR
CLAHE	0.8181	18.1827
Fusion	0.8222	21.1849
Histogram	0.7620	18.5148
ULAP	0.7318	16.2723
DuwieNet	0.8303	19.3134
GLCHE	0.8487	21.0270
SCNet	**0.8756**	**21.9482**

TABLE II
Ablation study on the UIEBD dataset. The best results are in bold.

Criteria	SSIM	PSNR
U-Net	0.8439	19.7249
SCNet w/o SN	0.8664	21.2098
SCNet w/o CN	0.8592	21.1710
SCNet (FULL)	**0.8756**	**21.9482**

A. Performance Comparison

Quantitative results of different UIE algorithms on the UIEBD dataset are presented in Tab. I. As we can observe, SCNet achieves the best performance in terms of two full-reference image quality evaluation metrics. This is because SCNet takes the diversity of underwater degradation into account. SCNet combines normalization methods with a U-Net to learn multi-scale water type desensitized latent repre-

Fig. 3. Visual comparisons on the UIEBD dataset [14]. (a) Raw image. (b) CLAHE. (c) Fusion. (d)Histogram. (e) ULAP. (f) DuwieNet. (g) GLCHE. (h) SCNet. (i) Reference. The proposed method can properly correct colors and clearly suppress artifacts.

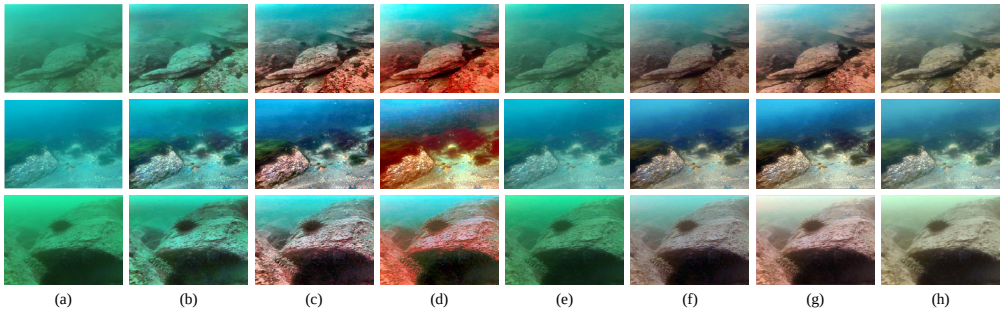

Fig. 4. Visual comparisons on the RUIE dataset [2]. (a) Raw image. (b) CLAHE. (c) Fusion. (d)Histogram. (e) ULAP. (f) DuwieNet. (g) GLCHE. (h) SCNet. The proposed method achieves better generalization performance and is able to handle diverse water types.

sentations, which can cope with the diversity of water types and recover clean images with higher quality. Prior-based methods (Histogram and ULAP) obtain low SSIM and PSNR scores. The reason may line in that prior-based methods rely on handcrafted physical imaging models and accurately prior knowledge. However, design a general and effective prior model is extremely difficult. Model-free methods (CLAHE and Fusion) and data-driven methods (DuwieNet and GLCHE) achieve higher performance compare with prior-based methods, but all of them are inferior to our method since they ignore the impact of diverse water types.

We also provide subjective comparisons in Fig. 3. From the visual results, we can make the following observations: 1) Model-free methods can deal with contrast distortions but fail to correct colors. Besides, this kind of method (e.g., Fusion) may produce a reddish color shift due to over-enhancement. 2) The enhanced results of Histogram and ULAP is poor and showing serious over-enhancement effects. 3) DuwieNet tends to generate under-enhanced results that are visually unsatisfactory. 4) The best two results are obtained by SCNet and GLCHE. Moreover, SCNet can handle the diversity of water types, and can consistently generate natural and vivid results on all testing images. This benefits from the learned latent space which is water type invariant. Fig. 4 reports the visual results on the RUIE dataset. As can be observed, the compared methods can not address the degradation diverse well, they tend to produce over/under-enhancement effects. On the contrary, SCNet enables generating enhanced images with more natural colors and fewer artifacts. This provides strong evidence that learning water type desensitized representations can help to improve the generalization performance.

B. Ablation Study

We conduct ablation studies to verify the effectiveness of proposed spatial-wise and channel-wise normalization approaches. Table II presents the test results using four different settings. From Tab. II, we can observe that directly using U-Net cannot obtain satisfactory results because it does not take the special distortions of the underwater environment into account. Normalizing representations on either spatial or channel dimensions can significantly improve enhancement performance. As expected, the best results are obtained by simultaneously normalizing features on both spatial and channel dimensions. This is because the diversity of water types not only exists in spatial dimensions but also channel dimensions.

IV. Conclusion

In this letter, we propose a novel data-driven method for underwater image enhancement. Different from most existing approaches that focus on designing complex network architectures to meet the requirement of generating high-quality enhanced outputs, we propose to combine a simple U-Net with spatial-wise and channel-wise normalization to deal with the diversity of water types meanwhile improve the visual quality of enhanced results. By normalizing activations across both spatial and channel dimensions, appearance irrelevant representations can be effectively learned. As a result, the decoder can easily reconstruct the clean signal from those latent representations. Experimental results show that SCNet achieves competitive performance on visual quality improvement and has better generation capacity for real applications.

REFERENCES

[1] G. L. Foresti, "Visual inspection of sea bottom structures by an autonomous underwater vehicle," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 31, no. 5, pp. 691–705, 2001.

[2] R. Liu, X. Fan, M. Zhu, M. Hou, and Z. Luo, "Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–1, 2020.

[3] S. M. Pizer, R. E. Johnston, J. P. Ericksen, B. C. Yankaskas, and K. E. Muller, "Contrast-limited adaptive histogram equalization: speed and effectiveness," in *[1990] Proceedings of the First Conference on Visualization in Biomedical Computing*, 1990, pp. 337–345.

[4] 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.

[5] Y. C. Liu, W. H. Chan, and Y. Q. Chen, "Automatic white balance for digital still camera," *IEEE Transactions on Consumer Electronics*, vol. 41, no. 3, pp. 460–466, 1995.

[6] C. Ancuti, C. O. Ancuti, T. Haber, and P. Bekaert, "Enhancing underwater images and videos by fusion," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 81–88.

[7] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2341–2353, 2011.

[8] J. Y. Chiang and Y. Chen, "Underwater image enhancement by wavelength compensation and dehazing," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1756–1769, 2012.

[9] 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.

[10] C. Li, J. Guo, R. Cong, Y. Pang, and B. 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.

[11] D. Berman, T. Treibitz, and S. Avidan, "Diving into haze-lines: Color restoration of underwater images," in *Proceedings of the British Machine Vision Conference*, 2017, pp. 1–12.

[12] 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*, vol. 3, no. 1, pp. 387–394, 2018.

[13] C. Li, J. Guo, and C. Guo, "Emerging from water: Underwater image color correction based on weakly supervised color transfer," *IEEE Signal Processing Letters*, vol. 25, no. 3, pp. 323–327, 2018.

[14] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, K. Sam, and D. Tao, "An underwater image enhancement benchmark dataset and beyond," *IEEE Transactions on Image Processing*, vol. 29, pp. 4376–4389, 2019.

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

[16] P. Uplavikar, Z. Wu, and Z. Wang, "All-in-one underwater image enhancement using domain-adversarial learning," in *2019 IEEE International Conference on Computer Vision Workshops*, 2019, pp. 1–8.

[17] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[18] X. Pan, X. Zhan, J. Shi, X. Tang, and P. Luo, "Switchable whitening for deep representation learning," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 1863–1871.

[19] B. Li, F. Wu, K. Q. Weinberger, and S. Belongie, "Positional normalization," in *Advances in Neural Information Processing Systems*, 2019, pp. 1622–1634.

[20] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4681–4690.

[21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[22] 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 *Pacific Rim Conference on Multimedia*. Springer, 2018, pp. 678–688.

[23] X. Fu and X. Cao, "Underwater image enhancement with global-local networks and compressed-histogram equalization," *Signal Processing: Image Communication*, p. 115892, 2020.

[24] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.