Underwater image enhancement using symmetrical autoencoder with synthesis datasets

Do Gyun Kim[1] · Seok Je Cho[†]

Abstract: The underwater images obtained using optical sensors, such as cameras, have low visibility and color distortion due to underwater environments. To enhance the visibility and color of underwater images, autoencoder-based convolutional neural networks are used recently; however, a basic autoencoder is not effective in color distortion and makes the result images noisy. In this study, we propose an autoencoder with skip-connection called symmetrical autoencoder (SAE) to improve the visibility and color distortion of underwater images. The proposal of this study is twofold: (i) to symmetrically add skip-connections that connect encoders to decoders throughout the network for reconstructing ability of decoders and (ii) to synthesize underwater datasets using an underwater image formation model to train the autoencoder effectively. Through the comparison with other approaches, we show that the proposed autoencoder outperforms them in PSRN, SSIM, and color difference for test datasets. In addition, the proposed autoencoder can well generalize the actual underwater images.

Keywords: Underwater image enhancement, Deep learning, Convolutional neural network, Auto-encoder, Skip-connection

1. Introduction

As interest in marine resources has recently increased, a lot of development of underwater robots is underway to research and explore the underwater environment. Underwater robots are equipped with various sensors to check the underwater environment. Among them, the optical sensor is one of the basic sensors used to intuitively check the surrounding environment of the robot. **Figure 1** shows the image acquisition processing in underwater. It is difficult to capture underwater images due to turbidity on organisms such as plankton and inorganic precipitate. In addition, light scattering and absorption in water cause color distortion that varies depending on the wavelength of the light. Furthermore, the background light makes the underwater images hazy. To improve the visibility and color of underwater images using fundamental image enhancement methods based on image formation model (IFM), the convolutional neural networks (CNNs) are mainly used.

Fundamental image enhancement methods improve underwater images by applying gamma correction for color correction [1], histogram equalization (HE) for contrast enhancement [2], and
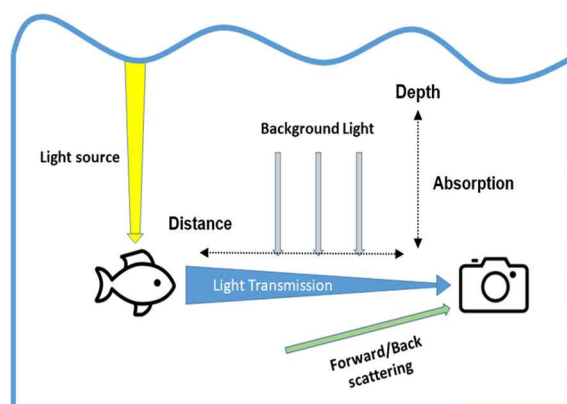

Figure 1: Image acquisition processing in underwater

etc. Although HE produces high-contrast images by evenly distributing the pixel values of the image, its results lose the original contrast. To overcome the weakness of HE, the contrast-limited adaptive histogram equalization (CLAHE) divides images into several patches and performs adaptive HE by considering each patch's original contrast [3]. However, it is difficult to improve low visibility and color distortion of the underwater environment using single fundamental image enhancement method.

† Corresponding Author (ORCID: http://orcid.org/0000-0001-9979-2252): Professor, Department of Control and Instrument Engineering, Korea Maritime & Ocean University, 727, Taejong-ro, Yeongdo-gu, Busan 49112, Korea, E-mail: sjcho@kmou.ac.kr, Tel: 051-410-4344

1 M. S. Candidate, Department of Control and Instrument Engineering, Korea Maritime & Ocean University, E-mail: ttjk2q@gmail.com, Tel: 051-410-4929
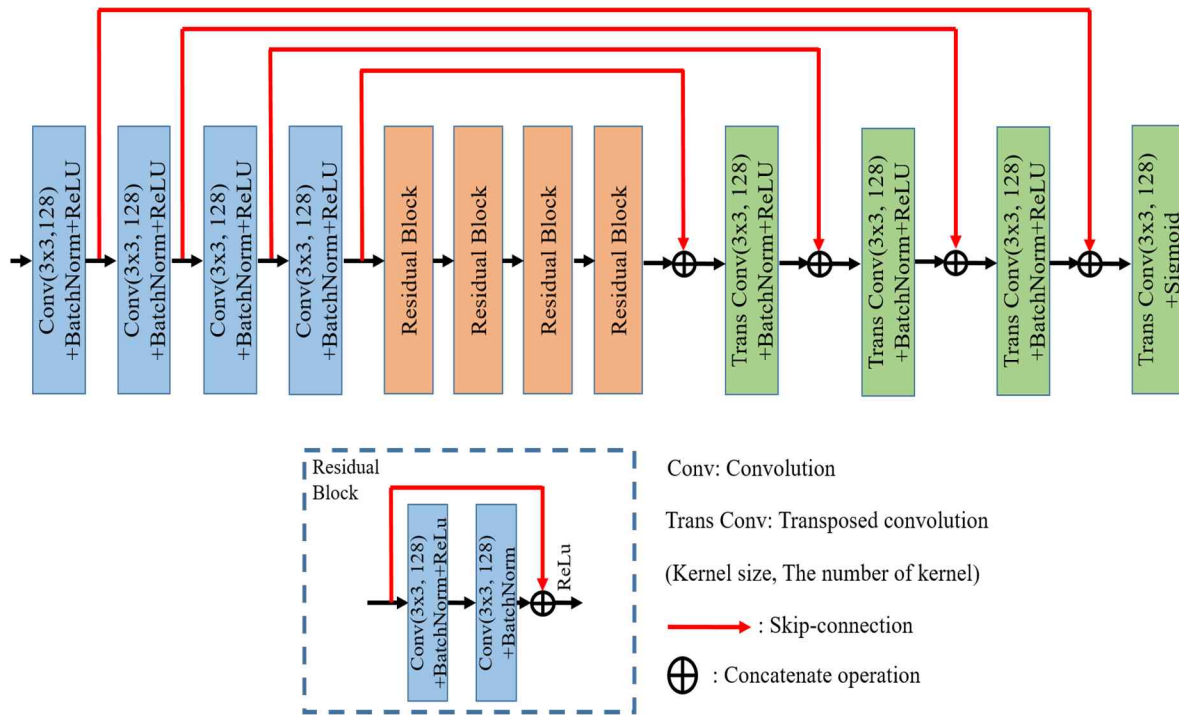
Do Gyun Kim · Seok Je Cho

Figure 2: Proposed architecture called symmetrical autoencoder

Thus, fundamental image enhancement methods are used in series or parallel. Ancuti *et al.* proposed a method called image fusion that combines three fundamental image enhancement methods, white balancing, gamma correction, and sharpening, to enhance color and visibility **[4]**. However, image fusion is not effective in improving color cast or over-enhanced red color due to white balancing.

The methods based on IFM are used to obtain the radiance of images without distortion. The dark channel prior (DCP) uses IMF to obtain haze-free images **[5]**. The DCP-based methods are frequently used for underwater images because the underwater images also have haze. Underwater DCP (UDCP) improves DCP by considering underwater environments in which the amount of transmission of red light is less than blue and green light **[6]**.

CNN, one of the deep learning methods used for image processing, is actively used for underwater image enhancements. It is used to get additional information for IMF **[7]** or to reconstruct enhanced images directly. One of the CNN network models, an autoencoder (AE), is mainly used for image enhancement. Basic AEs for colorizing was used for color restoration of outdoor images **[8]**. A basic AE, however, is not effective for underwater images that have both color distortion and low visibility.

In this study, we propose an autoencoder with skip-connection called symmetrical autoencoder (SAE) to improve the visibility

and color distortion of underwater images. The proposal of this study is twofold: (1) to symmetrically add skip-connections that connect encoders to decoders throughout the network for reconstructing ability of decoders and (ii) to synthesize underwater datasets using the underwater image formation model to train the autoencoder effectively.

Underwater datasets for training are made of indoor image datasets by means of distance information and underwater image formation models that are considered light scattering and absorption. To demonstrate the performance of the proposed SAE, we compare it with other approaches, such as CLAHE, Image fusion, DCP, UDCP, Colorizing AE, using test datasets and actual underwater images. We show that the SAE outperforms them not only in full-reference metrics such as peak signal to noise ratio (PSNR), structural similarity (SSIM), and color difference, but also in terms of visibility.

This paper is organized as follows: Section 2 introduces the proposed architecture of the SAE. Section 3 presents the underwater dataset synthesis for training, and our experimental results. Finally, we conclude this paper in Section 4.

2. Underwater image enhancement using SAE

2.1 SAE Architecture

The architecture of SAE is shown in **Figure 2**. It consists of

Table 1: Light scattering and absorption coefficient according to each water type in Jerlov water types. The water types from I to III are oceanic types. The unit of $\alpha(\lambda)$ and $\beta(\lambda)$ is m^{-1}

Wavelength (nm)	I		IA		IB		II		III	
	$\alpha(\lambda)$	$\beta(\lambda)$	$\alpha(\lambda)$	$\beta(\lambda)$	$\alpha(\lambda)$	$\beta(\lambda)$	$\alpha(\lambda)$	$\beta(\lambda)$	$\alpha(\lambda)$	$\beta(\lambda)$
650 (Red)	0.334	0.0009	0.334	0.0023	0.334	0.393	0.334	0.27	0.336	0.74
525 (Blue)	0.046	0.0021	0.047	0.0040	0.0047	0.078	0.047	0.387	0.051	1.06
450 (Green)	0.018	0.0038	0.022	0.0063	0.024	0.062	0.024	0.504	0.039	1.38

encoders, residual blocks, and decoders. The encoders consist of four convolution layers and obtain the compressed feature map of an input image using two or more strides. The residual blocks are made up of four residual blocks, of which each contains two convolution layers with a skip-connection. The skip-connection in residual block prevents gradients from being zero or near-zero when updating kernel coefficients. In other words, the problem of gradient vanishing can be avoided. The decoders are composed of four transposed convolution layers, the same as the encoders, and reconstruct the compressed feature map to the enhanced image with R-G-B channel.

In addition, to solve the complex problems of underwater images, we try to increase the depth of the decoders and the encoders. However, the more the encoders compress, the smaller the size of the feature map for the input image is. Accordingly, the shape and details of the original images are lost **[9]**, and the reconstructed images contain a lot of noise and color distortion. To overcome this problem, we add skip-connections to send the feature map of the encoders to the encoders directly.

Kernel size is one of important parameters in CNN. The smaller the size of kernel, the more effective the preserving details (e.g., edge information **[10]**) of the original image and the lower the computational cost of network. However, the extreme kernel size of 1×1 extracts local information from an image without considering spatial relationship of pixel. Therefore, we set the size of kernel to 3×3 for both considering the spatial relationship of pixel and reducing computational cost. The number of kernel is set to 128. Batch normalization (BN) and rectified logical unit (ReLU) is processed at each layer, except for the last layer. In the last layer, we use sigmoid function so that the result image has a pixel value of 0 to 1.

2.2 Updating Kernel Coefficient

We train SAE using underwater datasets consists of underwater images and corresponding clean images. The loss function of SAE is mean squared error (MSE) expressed in **Equation (1):**

$$MSE = \frac{\sum_{\lambda \in \{Red, Green, Blue\}}\left(I_\lambda^{resut}\,(i,j) - I_\lambda^{clean}\,(i,j)\right)^2}{3N} \qquad (1)$$

where $I^{resut}\,(i,j)$ and $I^{clean}\,(i,j)$ are the result image reconstructed by SAE and the clean image in the underwater datasets, respectively. i and j are the locations of current pixel in an image. N is the total number of pixels in the image. To update kernel coefficient in convolution, we use Adam optimizer in this study.

3. Experiments and Results

3.1 Synthesizing Training Datasets

To train the proposed SAE, we need pairs of an underwater image and its corresponding clean image. However, it is difficult to obtain these pairs in real underwater environments. To get the pairs effectively, we add noise to NYU datasets using the underwater image formation model (UIFM) **[11]** that is defined as **Equation (2)**:

$$I_\lambda^{syn}\,(i,j) = C_\lambda(i,j)E_\lambda^{in}\,(i,j)T_\lambda(i,j) + \left(1 - T_\lambda(i,j)\right)B_\lambda$$
$$= e^{-\alpha(\lambda)D(i,j)}E_\lambda^{in}\,(i,j)e^{-\beta(\lambda)d(i,j)}$$
$$+ \left(1 - e^{-\beta(\lambda)d(i,j)}\right)B_\lambda \qquad (2)$$

where λ is the element of $\{Red, Green, Blue\}$, and i and j are the locations of current pixel in an image. $I^{syn}\,(\cdot)$ is the synthesized underwater image. $D(\cdot)$ is the depth between the water surface and objects. $C_\lambda(\cdot)$ is color distortion that is affected by $D(\cdot)$. $\alpha(\cdot)$ is the light absorption coefficient. $E_\lambda^{in}\,(\cdot)$ is the no-distorted image substituted into NYU R-G-B images. $T_\lambda(\cdot)$ is the amount of light that is reflected from the object to the optical sensor. It is affected by the distance from the scene to the optical equipment, $d(\cdot)$, and light scattering coefficient, $\beta(\cdot)$. $d(\cdot)$ is substituted into NYU distance information.

To express distortions such as real underwater images, Jerlov

Do Gyun Kim · Seok Je Cho

Table 2: Mean values of PSNR, SSIM, and color difference of each method about all test datasets

	CLAHE	Image fusion	DCP	UDCP	Coloring AE	Proposed
PSNR	12.5092	14.0734	10.2442	10.5885	17.4380	22.0598
SSIM	0.5920	0.7199	0.4320	0.4726	0.8263	0.8797
Color diff	0.3989	0.3525	0.5074	0.4760	0.1902	0.1404

light coefficients are substituted into $\alpha(\cdot)$ and $\beta(\cdot)$. It is estimated through the transmittance of light measured in various types of ocean and coastal water. Although Jerlov offers five types of coastal and five types of ocean water of light coefficients, we consider only ocean water in this study because images taken in coastal water are very turbid and have severe color distortion, that is, there is limited information about scene. To capture the information of image in coastal environments, additional hardware devices are required. **Table 1** shows five types of ocean water of light coefficients used in this study. Other parameters are set to $D(\cdot) \in \{0.5, 15\}$, $B_{Red} \in \{0.2, 0.5\}$, $B_{Red} \in \{0.5, 0.75\}$, and $B_{Red} \in \{0.5, 0.8\}$. Consequently, we generate 10,000 synthetic training datasets randomly mixed with five types of ocean water. **Figure 3** shows the samples of training datasets used to train SAE.

NYU datasets consist of 1,449 R-G-B images taken indoor and include distance information in each image. We randomly draw 1,000 images from NYU datasets for training datasets and the other for test datasets.
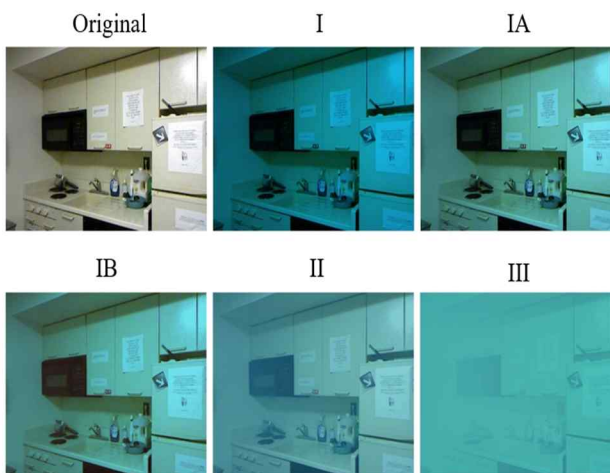
Figure 3: Samples of training sets using UIFM and NYU datasets

3.2 Performance Evaluation

We evaluate CLAHE, image fusion, DCP, UDCP, Colorizing AE, and SAE for test datasets. SAE and Colorizing AE run on 3.7 GHz CPU, 16G RAM, and Nvidia Geforce RTX 2070 SUPER GPUs using python and pytorch framework. Training parameters are the learning rate of 0.0005 and the batch size of 8. SAE and Colorizing AE are trained using our training datasets for 50 epochs. Each 449 NYU dataset for the test dataset is made of five types of underwater samples, as shown in **Figure 3**. To quantitatively evaluate the performance of each method, we use full-reference metrics such as PSNR, SSIM, and color difference of the R-G-B color channel. Color difference is computed as **Equation (3)**:

$$Cobr \quad diff \quad = \frac{\sqrt{\sum_{\lambda \in \{Red, Green, Blue\}}\left(I_\lambda^{result}\ (i,j) - I_\lambda^{clean}\ (i,j)\right)^2}}{N} \tag{3}$$

where $I^{clean}\ (\cdot)$ and $I^{clean}\ (\cdot)$ are the result image and the clean image in test datasets, respectively. i and j are the locations of current pixel in the image. N is the total number of pixels in the image. **Figure 4** shows the test dataset samples and the result images generated using each method. **Table 2** shows the mean values of PSNR, SSIM, and color difference of each method for all test datasets. The closer the result images to the original, the higher the PSNR and SSIM, and the lower the color difference. As shown in **Figure 4**, although CLAHE enhances the contrast, but it does not correct bluish and greenish tone. Image fusion is effective in visibility enhancement, but it is not effective in color correction. In addition, some images of fusion results of test datasets have artificial noise, as shown in the column **IA** of **Figure 4**. DCP darkens the bluish and greenish tone to create a dark image. Although UDCP compensates more for the red channel, the result shows a similar pattern as DCP. Colorizing AE is relatively effective for color correction than another methods, but it has noise as shown in columns **II** and **III** of **Figure 4**. The proposed SAE shows best performance for both actual result images and metrics. Additionally, **Figure 5** shows the result images of the actual underwater images obtained from Google.

4. Conclusion

We proposed SAE that is a CNN-based autoencoder with skip-connection throughout the network to enhance underwater images. To train SAE, we use synthesized underwater images
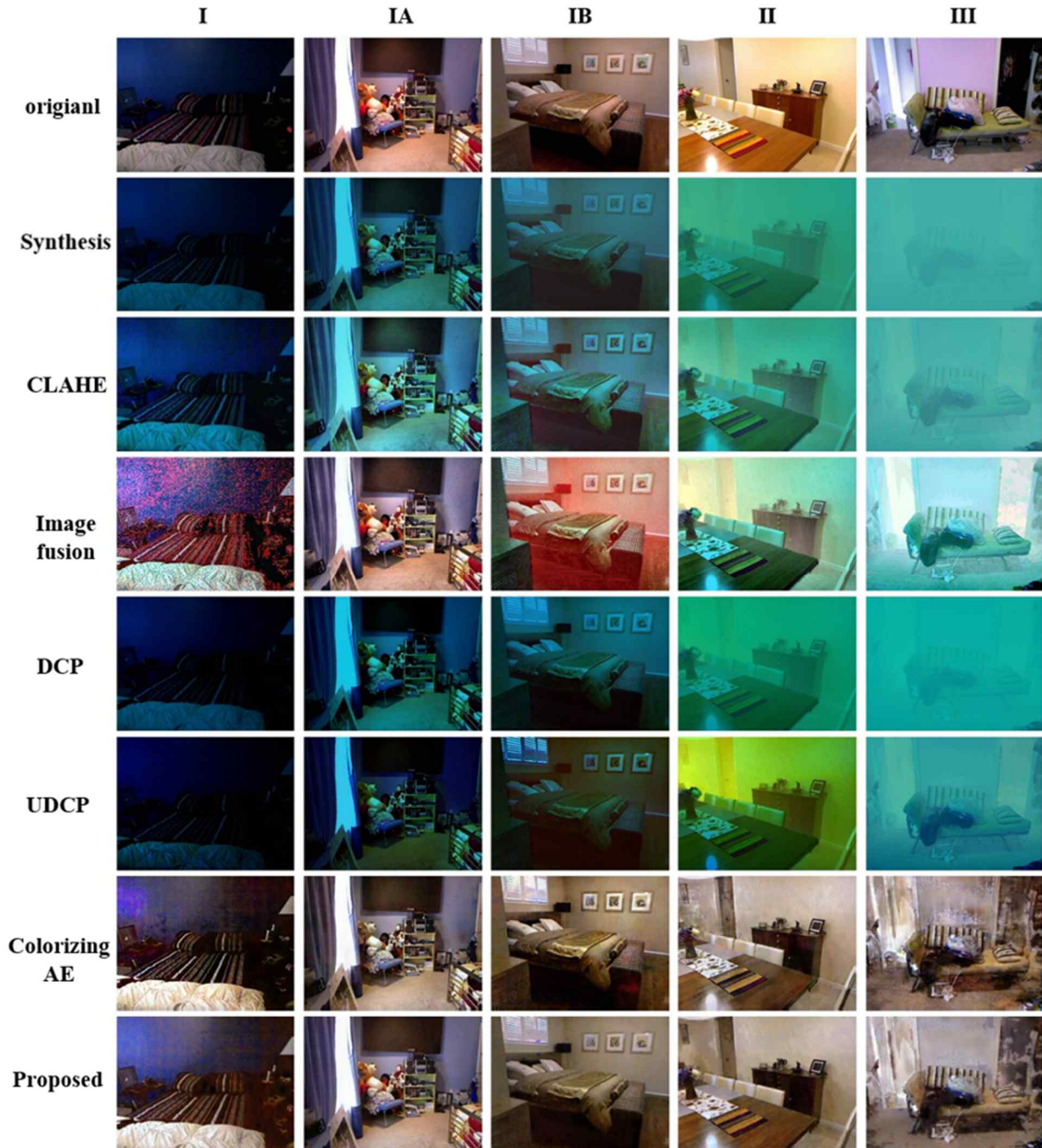
Figure 4: Samples of test datasets and their result images generated using each method. Synthesis images are generated in the same way as the training datasets. I ~ III are fives types of ocean water.

datasets that are mage out of NYU datasets using UIFM and Jerlov light coefficients to add actual color distortion and haze to underwater image datasets. Although SAE showed outstanding effectiveness visually and quantitatively in test datasets, the results of actual underwater images were less effective with respect to color correction and contrast enhancement than test datasets. Therefore, in future study, we will need to train the model by synthesizing a training dataset that

explains the actual underwater images well.

Author Contributions

Conceptualization, S. J. Cho and D. G. Kim; Methodology, D. G, Kim; Software, D. G. Kim; Formal Analysis, D. G. Kim; Investigation, D. G. Kim; Resources, D. G. Kim; Data curation D. G. Kim; Writing-Original Draft Preparation, D. G. Kim; Writing-Review & Editing, S. J. Cho; Visualization, D. G. Kim; Supervision,
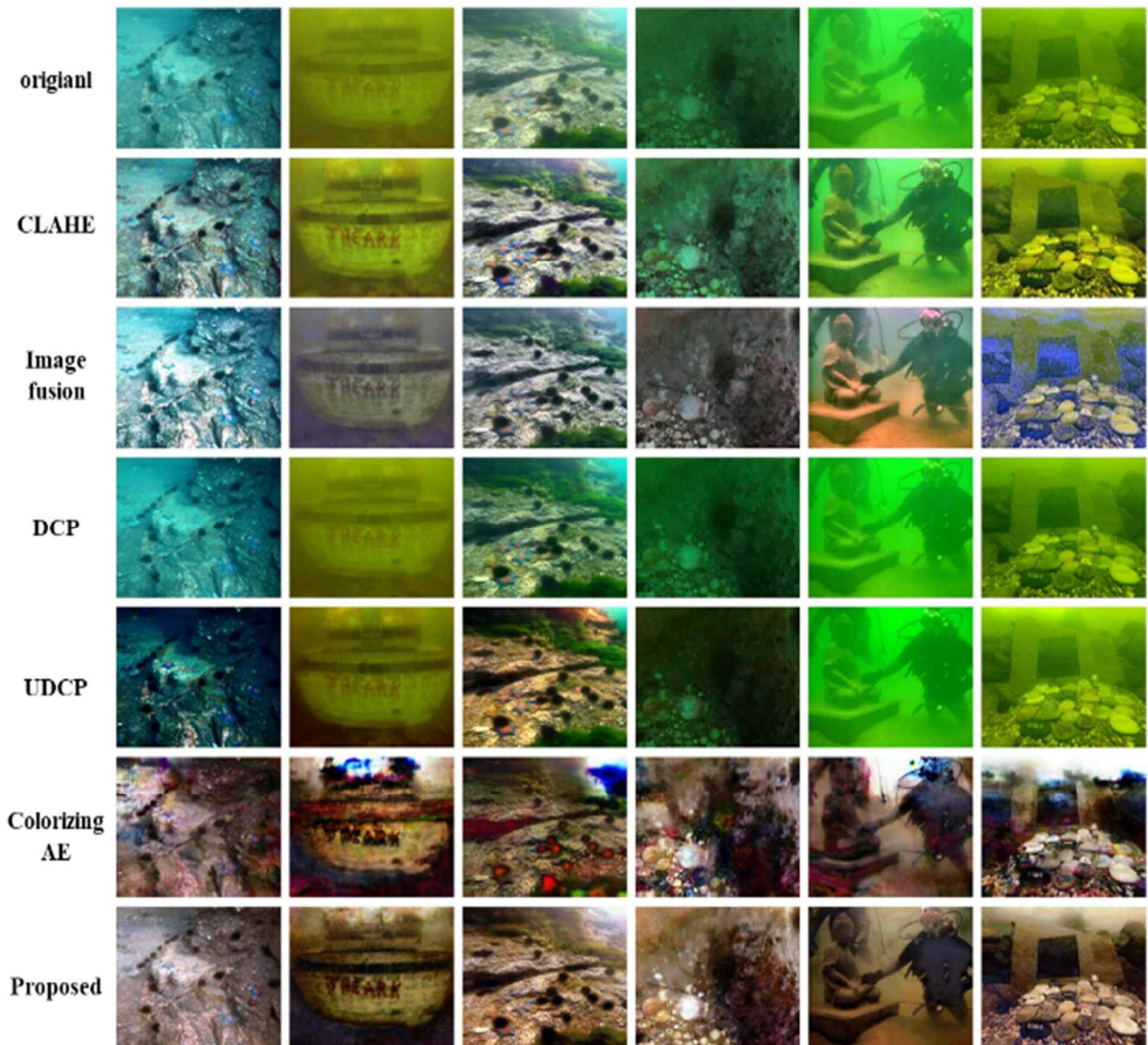
Do Gyun Kim ・ Seok Je Cho

Figure 5: Result images of the actual underwater images generated using the proposed method and other approaches

References

[1] G. Bianco, M. Muzzupappa, F. Bruno, R. Garcia, and L. Neumann, "A new color correction method for underwater imaging," ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 40 pp. 25-32, 2015.

[2] O. Deperlioglu, U. kose, and G. Guraksin, "Underwater image enhancement with HSV and histogram equalization," 7th International Conference on Advanced Technologies (ICAT'18), pp. 461-465, 2018.

[3] S. Anilkumar, P. R. Dhanya, A. A. Balakrishnan, and M. H. Supriya, "Algorithm for underwater cable tracking using CLAHE based enhancement," Proceedings of the International Symposium on Ocean Technology (SYMPOL), pp. 129-137, 2019.

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

[5] 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.

[6] P. Drews Jr, E. Nascimento, F. Moraes, S. Botelho, and M. Campos, "Transmission estimation in underwater single images," Proceeding of the IEEE International Conference on Computer Vision Workshops, pp. 825–830, 2013.

[7] M. Hou, R. Liu, X. Fan, and Z. Luo, "Joint residual learning for underwater image enhancement," Proceedings of the IEEE International Conference on Image Processing, pp. 4043-4047, 2018.

[8] L. Zhu and B. Funt, "Colorizing color images," Proceedings of the Electronic Imaging Symposium EI 2018 Conference on Human Vision and Electronic Imaging Science and Technology, 2018.

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 234-241, 2015.

[10] J. Yan, C. Li, Y. Zheng, S. Xu, and X. Yan, "MMP-Net: A multi-scale feature multiple parallel fusion network for single image haze removal," IEEE Access, vol. 8, pp. 25431-25441, 2020.

[11] X. Ding, Y. Wang, Y. Yan, Z. Liang, Z. Mi, and X. Fu, "Jointly adversarial network to wavelength compensation and dehazing of underwater images," arXiv preprint arXiv:1907.05595, 2019.