

Enhanced Color Sensing and Recognition of Underwater Color Using Robust Adaptive Tone Mapping

Kuo-Jui Hu,¹ Meng-Yi Chen,² Yuh-Shihng Chang,^{3*} and Sheng-Long Kao⁴

¹Graduate Institute of Color and Illumination Technology, National Taiwan University of Science and Technology,
No. 43, Keelung Rd., Sec. 4, Da'an Dist., Taipei City 106335, Taiwan (R.O.C.)

²Graduate Institute of Applied Science Technology, National Taiwan University of Science and Technology, Taipei,
No. 43, Keelung Rd., Sec. 4, Da'an Dist., Taipei City 106335, Taiwan (R.O.C.)

³Dept. of Information Management, National Chin-Yi University of Technology,
No. 57, Sec. 2, Zhongshan Rd., Taiping Dist., Taichung 411030, Taiwan (R.O.C.)

⁴Dept. of Transportation Science, National Taiwan Ocean University,
No. 2, Beining Rd., Zhongzheng Dist., Keelung City 202301, Taiwan (R.O.C.)

(Received May 2, 2023; accepted October 26, 2023)

Keywords: underwater color sensing, deep learning, image enhancement, Y'CbCr color space, adaptive tone mapping

Sensing images in the underwater environment is a significant issue in ocean engineering. Acquiring clear underwater images involves many challenges, such as climate, environment, and human factors. The most important problems are the fogging effect caused by the dispersion of light and the energy of each light wavelength when it propagates in water. Then, a color cast is caused by inconsistent attenuation. A common issue is the dispersion of light that occurs in underwater photography, which can impact the overall color balance of an imager. While current research can make use of good approaches for obtaining good visual quality and quantitative indicators, having a wider color gamut space and a dynamic image range can improve visible details. Therefore, we propose a module for enhancing underwater color image sensing with robust adaptive tone mapping for inferring degradation models using deep learning models and with adaptive tone mapping for further improving the image dynamic range. We address issues with limited dynamic range and brightness in underwater image sensing and recognition using a robust adaptive tone mapping method. Quantitative and qualitative results show that our method performs relatively well in the Underwater Image Enhancement Benchmark dataset compared with other recent methods that apply appropriate tone mapping to the large-scale layers of the image to preserve details and avoid over-enhancement. Therefore, the color gamut of our augmented image has a large scale and is evenly distributed when visualized in the Y'CbCr color space. In the future, our research method is expected to be applied to different types of underwater work and environment, and to reduce the severe degradation problems that usually occur in underwater images.

*Corresponding author: e-mail: eric_chang@ncut.edu.tw
<https://doi.org/10.18494/SAM4642>

1. Introduction

Images captured underwater often suffer from severe degradation issues such as color cast and background light scattering. Our solution to this problem consists of two stages: (1) an enhanced color recognition of underwater color sensing using robust adaptive tone mapping and (2) improved clear underwater scenes by enhanced color sensing. Substances such as tiny particles in water absorb most of the light energy and scatter the reflected light in the scene before it reaches the camera. Therefore, images tend to have low contrast. Owing to the high requirement for capturing high-quality underwater images, underwater image imprinting and enhancement have become active research areas in recent years.

Various algorithms have been developed to perform this task using traditional image processing and computer vision methods. These methods can be divided into enhancement and restoration methods. Enhancement methods aim to improve image quality purely by redistributing pixel intensities or using image pixel statistics, such as adopting color sensing and correction, restoration filters, histogram equalization, or a fusion of the above methods,^(1,2) and applying them to improve the image quality. Restoration methods rely on models of underwater image formation, which describe the relationship between the captured scene and the underwater environment, and are aimed at restoring the image assumed to be degraded.

In recent years, convolutional neural networks (CNNs) have proven effective in solving computer vision tasks such as image segmentation and object recognition. Therefore, many novel CNN-based methods have been proposed to solve the problem of underwater image restoration. The lack of training data for underwater images is a major challenge, so generative adversarial networks (GANs) are also widely used^(3–7) to generate realistic underwater images that can assist in supervising the learning process or performing restoration tasks. Still, there is a large gap between synthetic datasets and images captured underwater, and these models tend to ignore the characteristics of the environment. An example of a supervised learning model for underwater image enhancement is the U-color network.⁽⁵⁾ Li *et al.* improved the strength of feature representations using the so-called “multicolor space encoder network” by analyzing the input in different color spaces (such as RGB, HSV, and Lab) before passing RGB/Lab/HSV color space through a multilayer deep residual encoder network.^(5,6) An attention mechanism⁽⁷⁾ is applied to capture the most important features in all feature maps, and the decoder network is media-oriented to address the degradation of underwater image quality. The learning process combines a perceptual loss⁽⁸⁾ and a mean squared error (MSE) loss so that the loss function can represent both low-level and high-level feature differences between the output and the ground truth image. Using this as a baseline, we will try replacing the HSV color positions with Y’CbCr, since HSV has been shown to be useful in the representation of color properties.

Unlike past experiments, no ablation study is provided on the relationship between the contribution of perceptual loss in the loss function and the visual quality of the results and the effectiveness of the training process. In this study, we examine the choice of hyperparameters between the perceptual loss and the MSE loss. For many underwater applications, sensing visible details in large areas of the image, including dark and bright areas, is highly desirable in addition to visual quality. We use an augmentation step in the sensing process. A method using

bilateral filters and adaptive tone mapping⁽⁹⁾ is used to address this specific problem of enhancing the dynamic range of images. The input is first decomposed into large-scale and detailed layers using bilateral filters. Then, each large-scale layer is divided into three regions in accordance with the brightness level: dark, midtone, and light. Appropriate tone mapping is applied to each of these regions individually on the basis of the properties of each region. Adaptive tone mapping was performed on large-scale layers, thus preserving details in the sensing image. This approach can significantly improve the dynamic range and avoid side effects such as color separation caused by traditional methods such as histogram equalization. Our method achieves visually pleasing results while still maintaining a large amount of details in regions of different brightness levels. Figure 1 shows our experimental results.

2. Related Works

Inspired by the blurring problem of outdoor images, a key issue of an optical model, we used the image formation model (IFM) to describe the formation of blurred or degraded images. The formula for IFM is

$$I^c = J^c(x)T(x) + A^c(1 - T(x)), \quad c \in \{r, g, b\}, \quad (1)$$

where I^c is the observed intensity, J^c is the scene radiance, A^c is the homogeneous ambient or background light, and $T(x)$ is the medium transmission map indicating the amount of light that

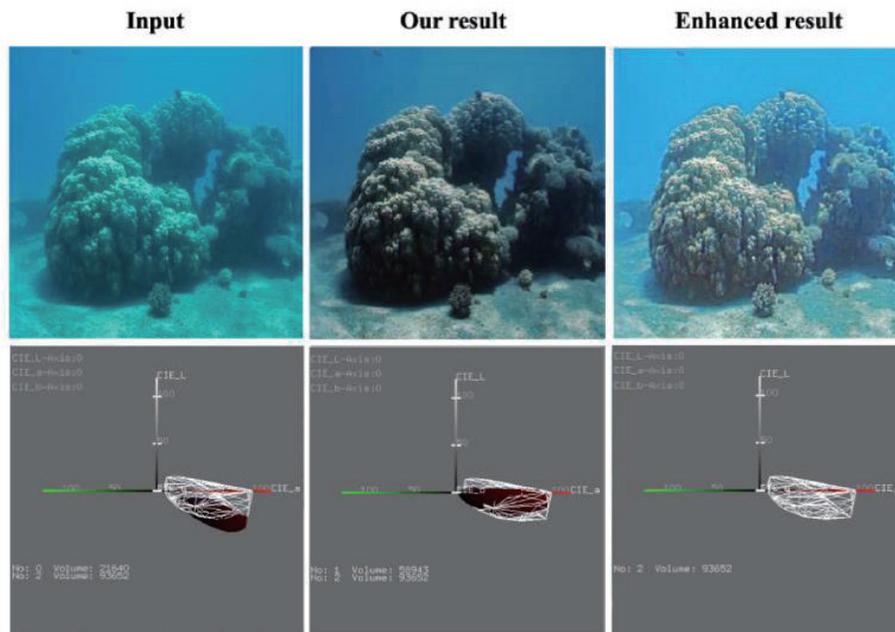


Fig. 1. (Color online) Visual demonstration of our underwater color sensing and restoration method. (a) is an original image, (b) is a translated image reconstructed by the network using the inferred causes, and (c) is a result of enhancement using adaptive tone mapping. Their corresponding visualization color gamut CIELAB color space with an enhanced result is shown in the white box, whereas the other color space shown as a solid line shows the gamut volume. The gamut volume is used as an objective function to optimize the imaging system.

reaches the camera. Most of the current methods use IFM as the starting point where the goal is to retrieve J^c through inverse operation if other variables are known.

Much intensive research has been performed in the field of underwater image processing methods. In general, we divide some state-of-the-art quality improvement methods into conventional improvement and deep-learning-based methods.

2.1 Conventional improvement methods

As mentioned in Sect. 1, the above two methods are aimed at enhancing images, while our approach is aimed at restoring degraded scenes. The main difference is whether the restoration method relies on the physical model of the image to make assumptions.

Underwater image enhancement algorithms focus on reallocating pixels to improve image quality using traditional methods. Owing to the limited number of scenes captured underwater, there are often severe distortions and artifacts. In the past, a fusion-based approach was developed to address this problem. On the basis of four fusion weights computed by the Laplacian, two inputs are combined using color sensing contrast, local contrast, saliency, and exposure to produce an output image with significantly improved visual quality. The visual quality of the image is further improved by incorporating a multiscale fusion strategy to reduce artifacts in the low-frequency parts of images enhanced by earlier algorithms.

Underwater image restoration methods are based on assumptions about how images are physically degraded by factors such as light absorption and scattering. The most famous of these is dark channel prior (DCP), which is used for outdoor image dephasing. Underwater DCP (UDCP), which calculates the dark channel on the basis of only the blue and green channels, is tailored for specific uses in underwater environments. Peng *et al.* proposed the red channel prior, where the dark channel is computed from the cyan and inverse red channels, to generalize these assumptions. Peng *et al.* proposed a generic dark channel before covering the different properties of all these different environments (not just underwater).⁽¹⁰⁾

2.2 Deep-learning-based methods

With the success of deep learning, Li *et al.* proposed to use a synthetic underwater database to train a CNN-based model to regenerate sharp and restored images.⁽¹¹⁾ There are still large discrepancies between synthetic data and real underwater images. To solve the problem of insufficient underwater training datasets, Li *et al.* also constructed an underwater image enhancement benchmark dataset for the comprehensive study of various enhancement algorithms.⁽¹²⁾ Furthermore, the power of CNN is fused in order to combine the advantages of restoration and enhancement methods. Li *et al.* proposed Ucolor's network,⁽⁵⁾ which uses a multicolor space embedding and attention mechanism as the basis of the encoder network. Features in color spaces other than RGB can be considered and selectively enhanced. At the same time, the media transport graph is used to effectively guide the decoder network to the area of quality degradation.

3. Proposed Method

We propose a method based on deep learning while utilizing traditional augmentation methods to solve the problem of underwater image restoration. The architecture design is divided into a learning network for quality improvement and an adaptive tone mapping module to further increase the dynamic range. We used the network architecture proposed in this paper as the network design. Tone mapping is used with bilateral filters to sensor the visibility of details, especially in dark areas, without overexposing and distorting the image. Furthermore, we conducted a study to analyze the effect of hyperparameters on learning the loss function used in the network.

3.1. Network architecture

The main components include the Residual Learning Module (RES-MOD), Channel Attention Module (CA-MOD), and Transfer Guidance Module (TRANS-MOD). Gray boxes in Fig. 2 indicate input or output quantities and are dimensionalized. All convolutional layers used in the network have filters of size 3×3 and stride 1. The residual module is written as, for example, RES-MOD 128, indicating that the number of filters for the convolution operation in the module is 128. Downsampling is performed by max pooling operation and upsampling by bilinear interpolation. The encoding module architecture of the above method is shown in Fig. 2(a) and the decoding module architecture in Fig. 2(b).

The flowchart and detailed architecture of the network are shown in Figs. 2(a), 2(b), and 3. The network architecture consists of a color space encoder (Sect. 3.1.1), residual and attention modules (Sects. 3.1.2 and 3.1.3), a medium transmission guidance module (Sect. 3.1.4), and a loss function (Sect. 3.1.5). The adaptive tone mapping module is described in a separate section (Sect. 3.2).

In the encoder network, the input passes through a color space encoder involving color space conversion from RGB to CIELAB and Y'CbCr color spaces. In each encoding, input passes through three residual learning modules, and $2\times$ downsampling is applied between them. For the color space encoding path, there will be three levels of feature representations of $\times 1$, $\times 1/2$, and $\times 1/4$ sizes. For each level, the outputs of the three color spaces are concatenated into volumes of different features and recalibrated according to the weights of each channel by a channel attention module.

For the decoder network, three different levels of feature quantities $\times 1$, $\times 1/2$, and $\times 1/4$ are combined with correspondingly sized reverse medium transfer (RMT) maps in the transfer guiding module. RMT maps of three different sizes can also be obtained by $2\times$ downsampling. The $\times 1/4$ output of this module is forwarded through another residual learning module, followed by $2\times$ upsampling. Then, its output is concatenated with the $\times 1/2$ output of the medium transfer module and again passed through the residual learning module with $2\times$ upsampling. We then connect the media transmission guidance result of level $\times 1$ with the previous operation result and perform a convolution to reconstruct the result to obtain a more accurate resolution and remove fog.

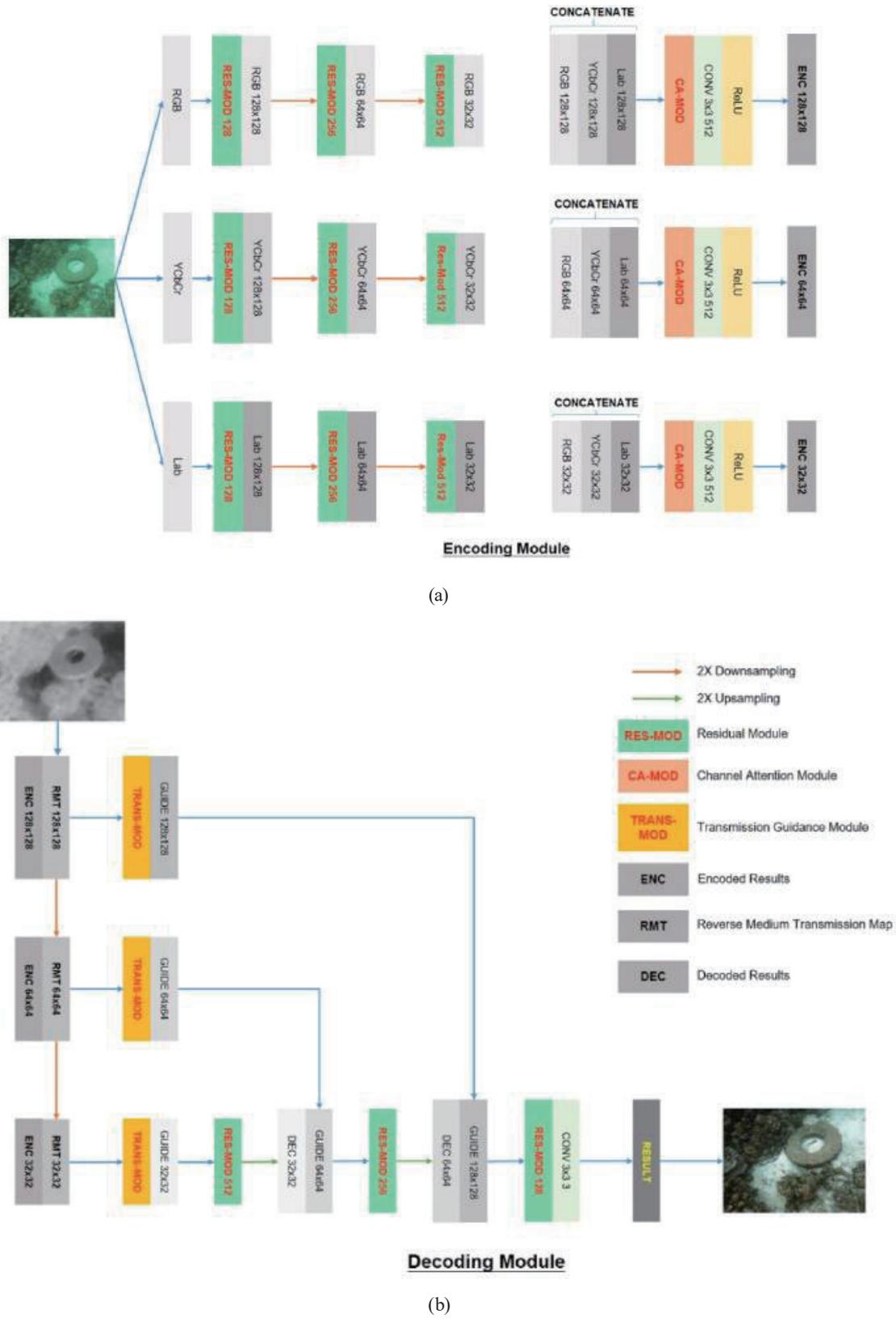


Fig. 2. (Color online) (a) Encoding module and (b) decoding module.

3.1.1. Color space encoder

The network starts with a color space transformation of the degraded underwater image. Owing to the shortcomings of the HSV color space, the multicolor space encoder module was replaced by the luminance-based color space YCbCr or Y'CbCr because human vision is more sensitive to changes in brightness than to changes in color. Y'CbCr values are normalized to the [0, 1] range before being fed into the trained network as follows:

$$Y'_{normalized} = \frac{Y' - 16}{235 - 16}, \quad (2)$$

$$C_{B_{normalized}} = \frac{C_B - 16}{240 - 16}, \quad (3)$$

$$C_{R_{normalized}} = \frac{C_R - 16}{240 - 16}. \quad (4)$$

3.1.2. Residual enhancement module

To address both increasing and decreasing network depth, a residual learning framework is used in both encoder and decoder networks. Figure 3 illustrates the building blocks of this module. This module uses the skip connection function by copying the learned layer from the lower layer and setting the additional layer as an identity map.

3.1.3. Channel attention module

To capture the interdependence between channel features from different color spaces, a squeeze-and-excitation block is integrated at the end of the encoder network. The main goal is to know which channels in different color spaces are more important and contribute more to the output. Figure 4 depicts the components of this module.

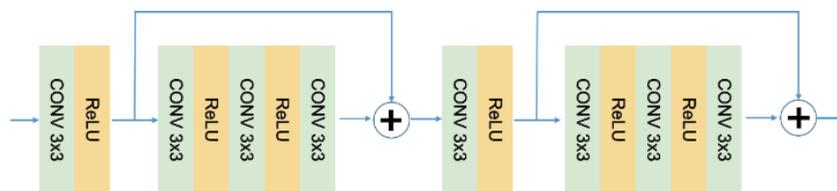


Fig. 3. (Color online) Residual augmentation module architecture. The input becomes x after passing through a set of convolutional layers with ReLU activations, then through a set of 2 CONV/ReLU layers and an additional CONV, and finally, without activation into $F(x)$. The output is $x + F(x)$. The process is repeated twice in each module.

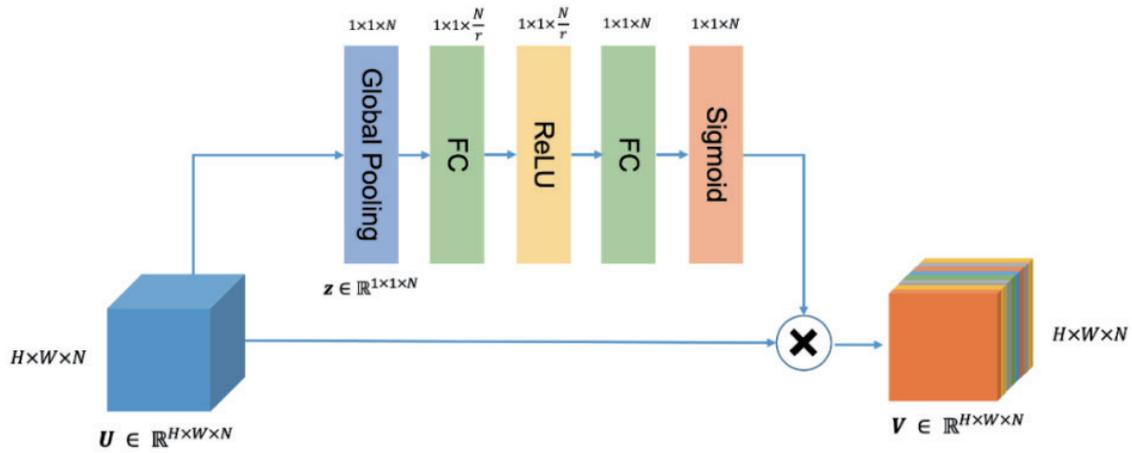


Fig. 4. (Color online) Channel attention module architecture. First, the feature volume is squeezed into a $1 \times 1 \times N$ vector by global average pooling. Through a fully connected (FC) layer, the weights are learned together with other parameters. ReLU and Sigmoid are used to select the most representative features and produce feature scores. The final result is obtained by multiplying the feature score with the original feature pixel by pixel.

3.1.4. Medium transmission guidance module

The medium transfer map at the beginning of the decoder network is calculated by computing the RMT. It is used to evaluate the importance of each position in a feature map. More specifically, the more degraded a pixel, the higher the weight assigned to it because this pixel requires more attention.

3.1.5. Loss function

Affected by the use of perceptual loss in computer vision tasks such as super-resolution and style transfer,⁽¹¹⁾ the loss function L_f is set to be the linear combination loss L_{per} :

$$L_f = \lambda_1 L_{MSE} + \lambda_2 L_{per}, \quad (5)$$

where λ_1 and λ_2 are set as hyperparameters. The MSE loss L_{MSE} is a measure of the per-pixel difference between the output feature map and the original feature map y calculated using the Euclidean distance:

$$L_{MSE}(\hat{y}, y) = \frac{1}{C \times H \times W} \|\hat{y} - y\|_2^2. \quad (6)$$

Instead of using per-pixel loss as the training target, L_{per} can perceive the loss of the output image \hat{y} to have a pattern feature representation and structure as similar as possible to the target image y that is close to the real image effect without a degraded pixel. This is carried out by processing the result \hat{y} and the ground truth y by pretraining the layer j of the network. The perceptual loss is the squared and mean Euclidean distance between these two outputs:

$$L_{per}(\hat{y}, y) = \frac{1}{C_j \times H_j \times W_j} \phi_j(y) - \phi_j(y)_2^2, \quad (7)$$

where $\phi_j(x)$ is the activation of the j th layer of the ϕ network. Here, ϕ is defined as the pretrained VGG-19 network on the ImageNet dataset,⁽¹³⁾ and the j th layer is the relu5_4 of the VGG-19 network.

3.2. Enhancement using bilateral filters and adaptive tone mapping

The algorithm uses a bilateral filter to divide the luminance channel Y' into a large-scale layer and a detailed layer. The chroma channels C_b and C_r are kept unchanged to avoid corrupting the input color information. Subsequent operations are only applied to large-scale layers to keep sharp edges, textures, and details.

We used Otsu's thresholding to divide the large-scale layer into three regions of brightness level: dark, midtone, and light. Then, we divided the large-scale layer into two thresholded regions in accordance with to the value T . The operation flow is shown in Fig. 5. Figure 6 shows the approximate values for the three regions of different brightness levels. Readers can refer to the article of Hu *et al.*⁽⁹⁾ for more details about the implementation.

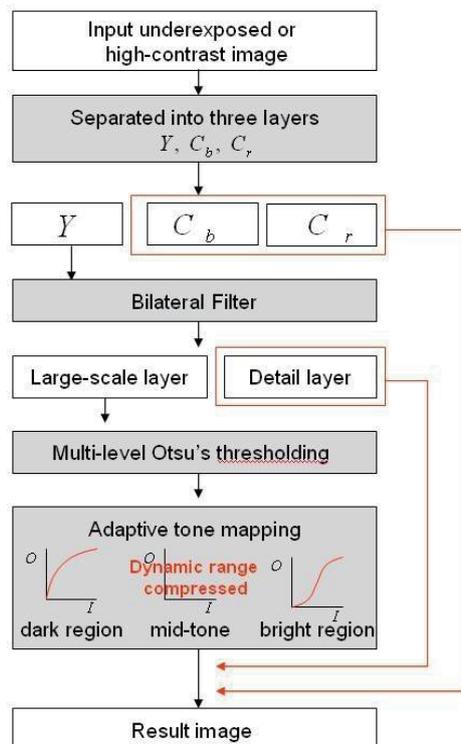


Fig. 5. (Color online) Flowchart for dividing the large-scale layer and the detailed layer with the bilateral filter algorithm.

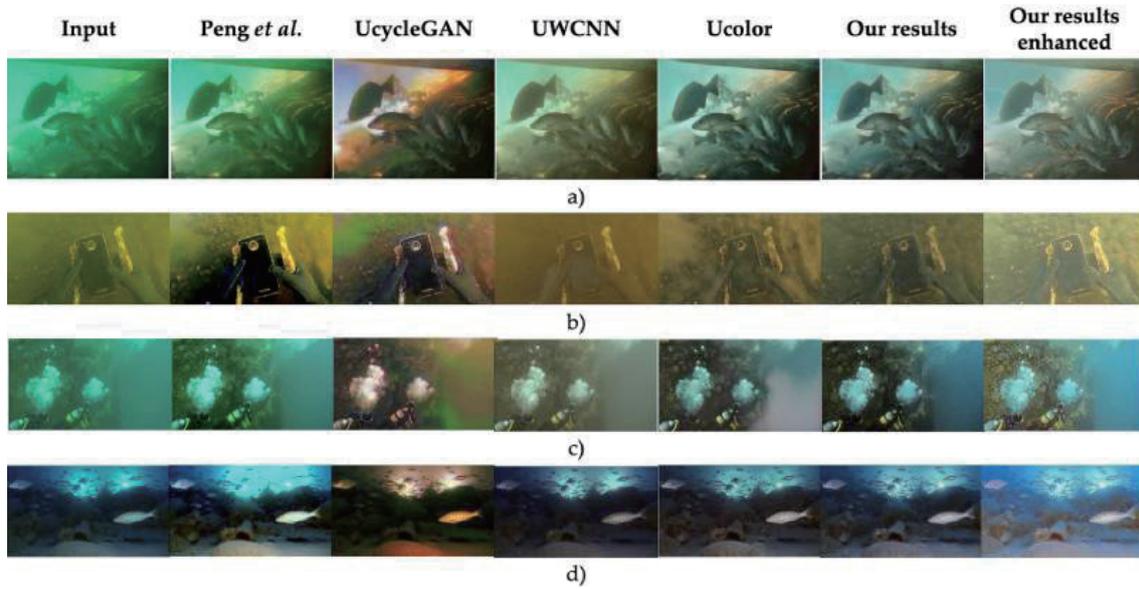


Fig. 6. (Color online) Visual results of various restoration methods for the UIEB-90 dataset consisting of 60 challenging images taken from the UIEB dataset without corresponding reference values.

4. Results and Discussion

4.1. Training and validation

We used the UIEB dataset⁽¹²⁾ as the training dataset to train the model in this study. In the selection of hyperparameters, the training data is randomly cropped to a size of 128×128 . We trained the network with minibatch gradient descent with a batch size of 16. The adaptive moment estimation optimizer (ADAM) was used for training optimization, and β_1 was set to 0.5. The learning rate was set to 1×10^{-4} .

As for the weights λ_1 and λ_2 of the per-pixel loss and the perceptual loss of the loss function, it can be seen that λ_2 considerably affects the quality of the output image as well as the training process. λ_2 in the range [0.001, 0.01] results in test images with low contrast and unnaturalness, whereas λ_2 in the range [0.01, 1.0] results in images that are more visually pleasing. On the other hand, the lower the λ_2 , the faster the training and validation errors converge to 0. Finally, we set the parameters to $\lambda_1 = 5$ and $\lambda_2 = 0.05$.

4.2. Experiment settings

We used the following datasets for experiment and comparison.

- UIEB-90: the remaining 90 pairs of images in the UIEB dataset.⁽¹²⁾
- UIEB-60: 60 underwater images from the UIEB dataset, which are deemed more challenging and do not have corresponding reference images.
- SQUID: 16 images taken from the SQUID dataset⁽¹²⁾ that contains 57 underwater image pairs

taken from various dive sites in Israel. Same as Test-C60, these data do not have corresponding ground truths or reference images.

Our network results are compared with the results from Ucolor⁽⁵⁾ and the enhanced results using adaptive tone mapping, since our goal is to see if there is any improvement over the original architecture. We run tensorflow on the same version of the UIEB dataset and keep all the best hyperparameters mentioned in the original work, and compare examples of different recovery methods, such as those by Peng *et al.*,⁽²⁾ UcycleGAN,⁽³⁾ and UWCNN.⁽¹¹⁾ Li *et al.*⁽⁵⁾ showed that compared with these methods, our network architecture through training and validation is superior to the above results ^(2,3,5,11) on the UIEB dataset.

In a comparison between two UIEB-90 datasets, the visual representation of the results is quantitatively evaluated against ground truth data using the mean square error (MSE) metric, the peak signal-to-noise ratio (PSNR) metric, and the structural similarity metric (SSIM). A low MSE score must be near 0, a higher PSNR value is better, and a higher SSIM score near 1 is better, indicating that the visual representation structure of the recovery image is closer to the results in references.^(2,3,5,11)

To evaluate the effectiveness of the adaptive tone mapping augmentation method, we compared different results by visualizing the color gamut of each image in the CIELAB color space. The goal is to further improve details in shadow areas, so the result of a color gamut with a larger volume and an even distribution on the luminance L axis is considered visually better.

We show visual results in Figs. 6–8. It can be observed that our method improves the visibility of the overall image structure and removes the color cast very effectively. Li *et al.*⁽⁴⁾ and Li *et al.*⁽⁵⁾ produced severe additional color artifacts, indicating the lack of robustness of traditional restoration methods and the ineffectiveness of GAN for underwater imaging

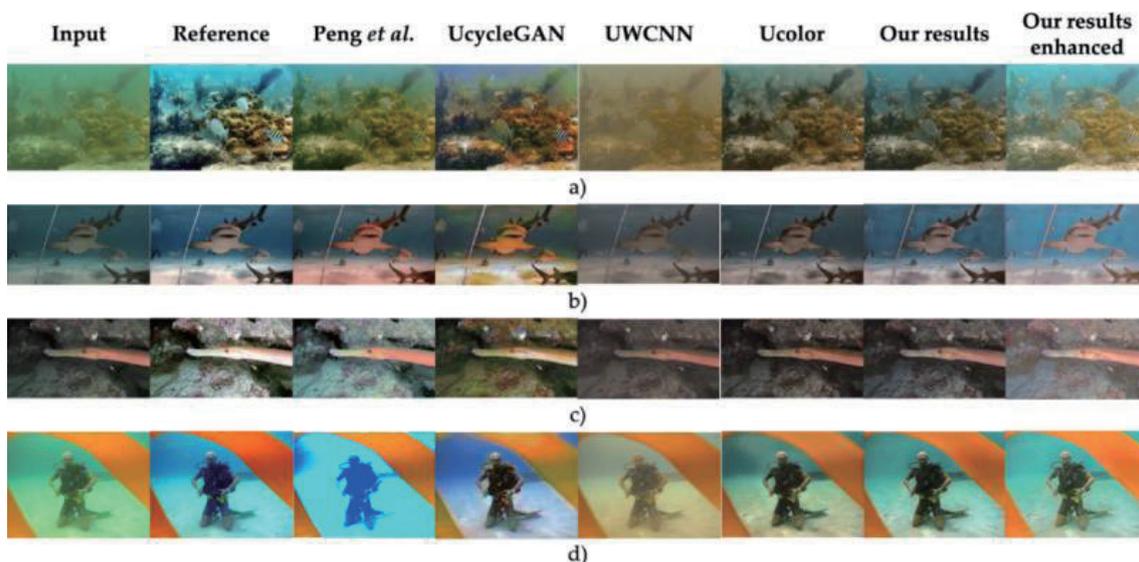


Fig. 7. (Color online) Visual results of various restoration methods for the UIEB-90 dataset consisting of 90 images taken from the UIEB dataset with corresponding reference values.

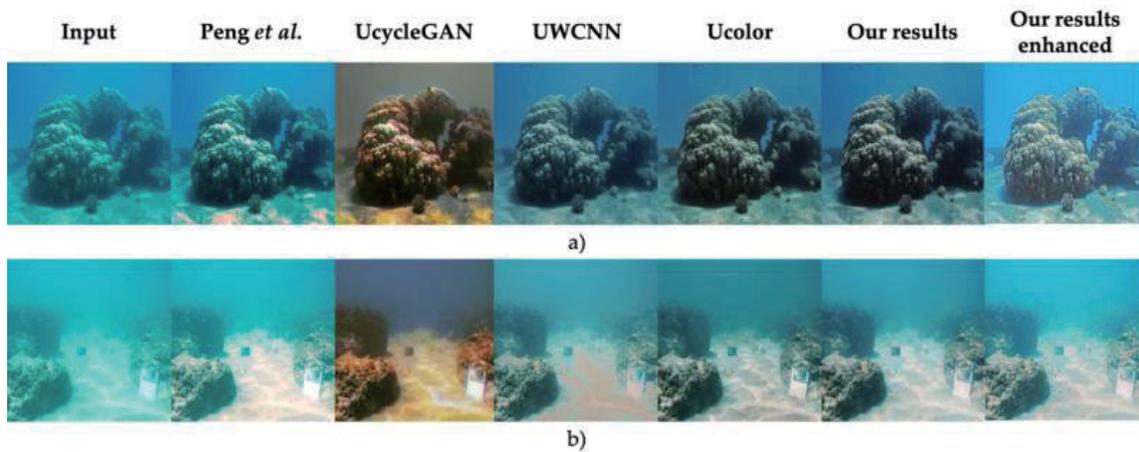


Fig. 8. (Color online) Visual results of various restoration methods for the SQUID dataset.

enhancement, which is also more obvious in the UIEB-60 dataset, as shown in Fig. 9. For UWCNN,⁽¹⁴⁾ the color cast in the input is not completely removed and has high turbidity. Ucolor⁽⁸⁾ performs relatively well in removing the green-blue cast and restoring sharpness in raw images. However, there are some parts that are not recovered uniformly in a single image, resulting in images with small halos, such as the unusual gray areas in Figs. 7(b) and 7(d) or Figs. 7(a), 7(b), and 6(c). In contrast, our network not only does not produce such artifacts, but also has visually pleasing colors and reasonable brightness and contrast. As for the adaptive tone mapping method, details in the enhanced result, especially in dark areas, are more visible. In images with limited illumination that our network fails to improve, as shown in Fig. 6(d), tone mapping also manages to reveal many details in the background. The tone mapping is still able to capture parts such as the rocky cliffs in Fig. 6(d), the background in Fig. 6(b), the coral reef in Fig. 8(c), and the rocks in Fig. 8(a). Visual comparisons highlight the effectiveness of the system, which produces satisfactory results even in degraded environments.

4.3 Quantitative evaluation

The different methods are quantitatively compared using the MSE metric, PSNR metric, and SSIM on the UIEB-90 dataset. The average scores are shown in Table 1. As can be seen from the table, our network achieves higher quantitative average scores than other methods (0.015/19.32/0.8674 in terms of MSE/PSNR/SSIM, respectively). Compared with Ucolor, our network achieves a percentage gain of 6.25/1.45/0.51% in terms of MSE/PSNR/SSIM, respectively.

Next, we analyze the effect of the tone mapping augmentation module on different images. For each comparison, we visualize the gamut of the image. The gamut of the augmented result is visualized with a white box, whereas the other gamuts are white solid lines. It can be seen that the enhanced results have the largest gamut volumes [194053 in Fig. 9(a), 223764 in Fig. 9(b), and 175768 in Fig. 9(c)]. In all three examples, the gamut of our unenhanced results manages to

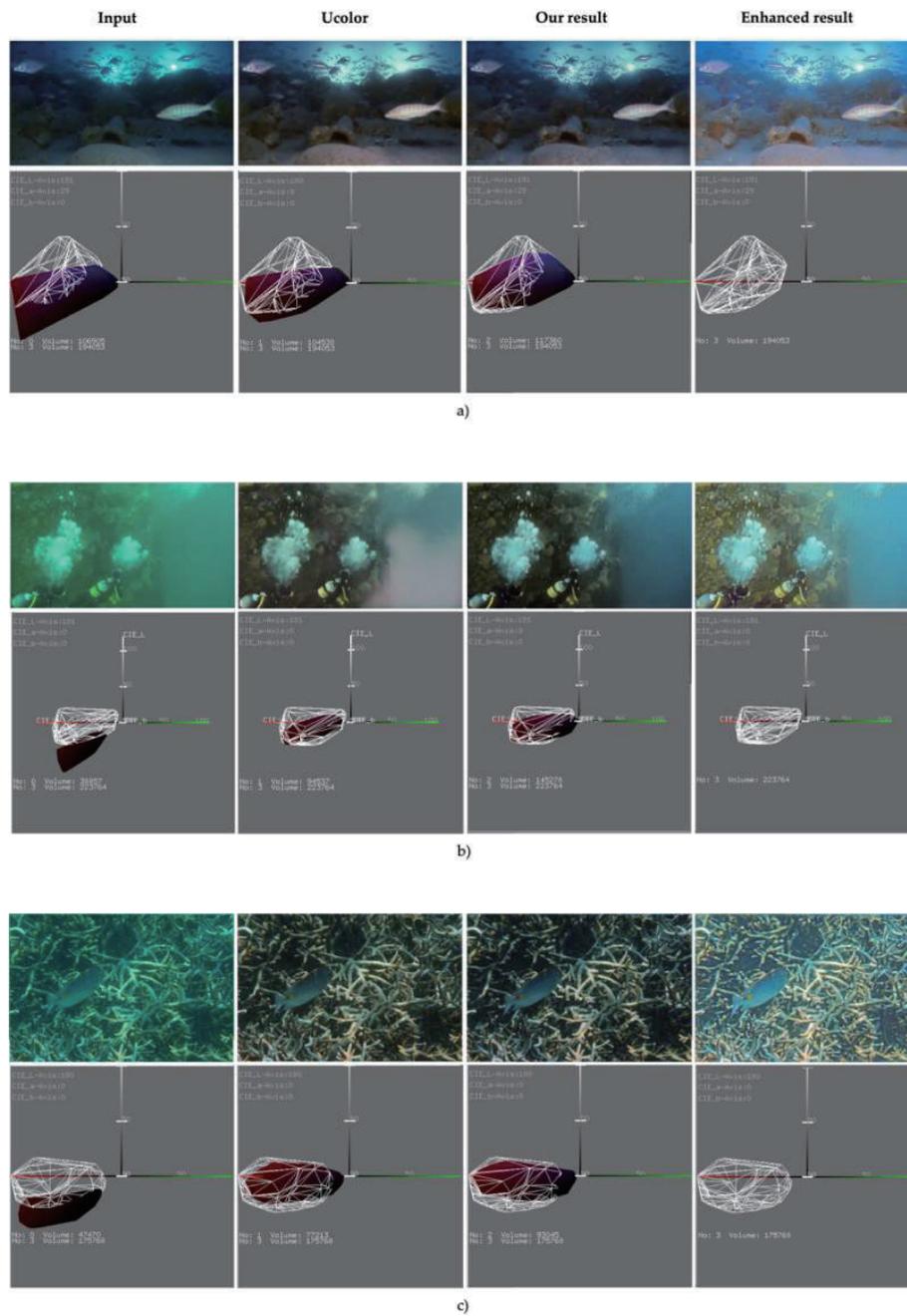


Fig. 9. (Color online) Visual demonstration of our underwater restoration method. The gamut volumes in CIELAB color space are shown and the corresponding gamuts are visualized.

Table 1
Various methods evaluated on the UIEB-R90 dataset in terms of mean MSE, PSNR (dB), and SSIM values.

Method	MSE	PSNR	SSIM
UWCNN ⁽¹⁴⁾	0.026	16.719	0.807
Peng <i>et al.</i> ⁽⁴⁾	0.026	15.777	0.791
UcycleGAN ⁽⁵⁾	0.025	16.65	0.684
Ucolor ⁽⁸⁾	0.016	19.043	0.863
Our network	0.015	19.32	0.8674

cover a larger volume while still occupying relatively the same space in the CIELAB color space. On the other hand, the gamut of the enhanced image shows that the tone mapping method stretches the graph contrast of images, showing that our method can generate images with slightly more vibrant colors and better contrast.

4.4 Analysis of the optimization of perceptual loss

During tuning, the value of λ_2 considerably affects the quality of the output image as well as the training process. As can be seen from Fig. 10, λ_2 in the range [0.0001, 0.01] results in test images with low contrast and unnaturalness. The reproduced colors are also far from the reference color checker, whereas λ_2 in the range [0.01, 1.0] makes the image more visually pleasing. On the other hand, Fig. 11 shows that the lower the λ_2 , the faster the training and validation errors converge to 0. The validation error also seems to be more stable with little fluctuation when λ_2 is low. We can also observe that the impact of λ_1 on the visual quality of output test images is less pronounced than that of λ_2 . A higher λ_1 also leads to a harder convergence of training and validation errors, although not as significant as λ_2 , as shown in Fig. 12.

For the second round of tuning, we then narrowed down the range of λ_2 to [0.01, 0.1] and $\lambda_1 \in \{1, 2, 3, 4, 5\}$ to balance visual results and training stability. We used a grid search method to find the most suitable set of hyperparameters. For each group (λ_1, λ_2) , we trained the model for 50 epochs, using the trained weights and parameters to infer the weights λ_1, λ_2 for the per-pixel loss and the perceptual loss for the loss function on the UIEB-R90 dataset that can gain better the MSE, PSNR (dB), and SSIM values.. To measure the quality of the results, we used MSE/PSNR/SSIM. As shown in Table 2, we found that $\lambda_1 = 5$ and $\lambda_2 = 0.05$ produced the best results. The hyperparameters were determined from the loss function.

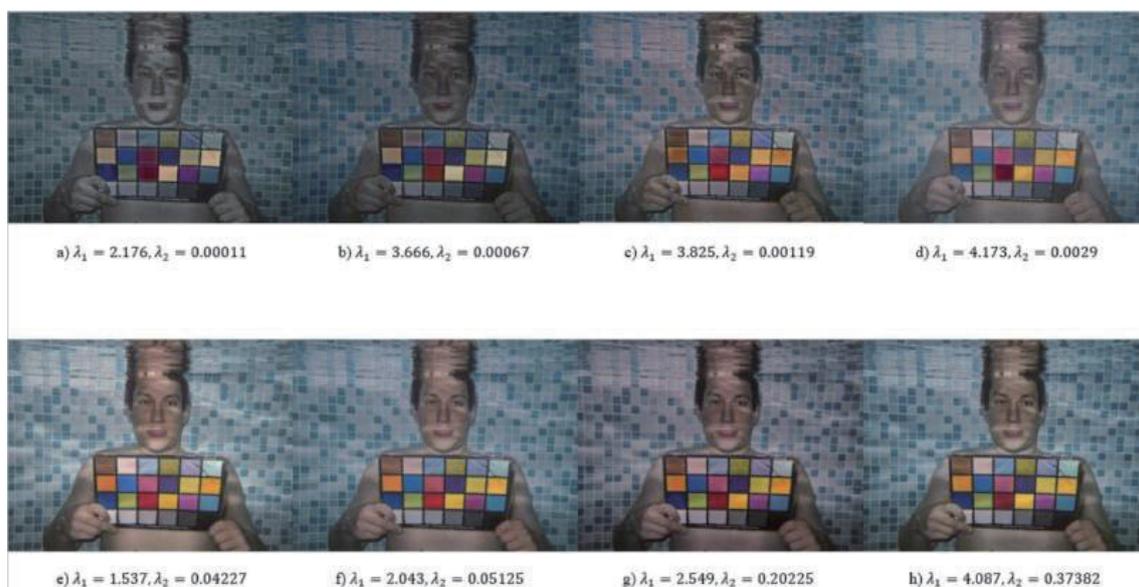


Fig. 10. (Color online) Results of first pass tuning. λ_2 values below 0.01 produce results where colors look unrealistic and distorted.

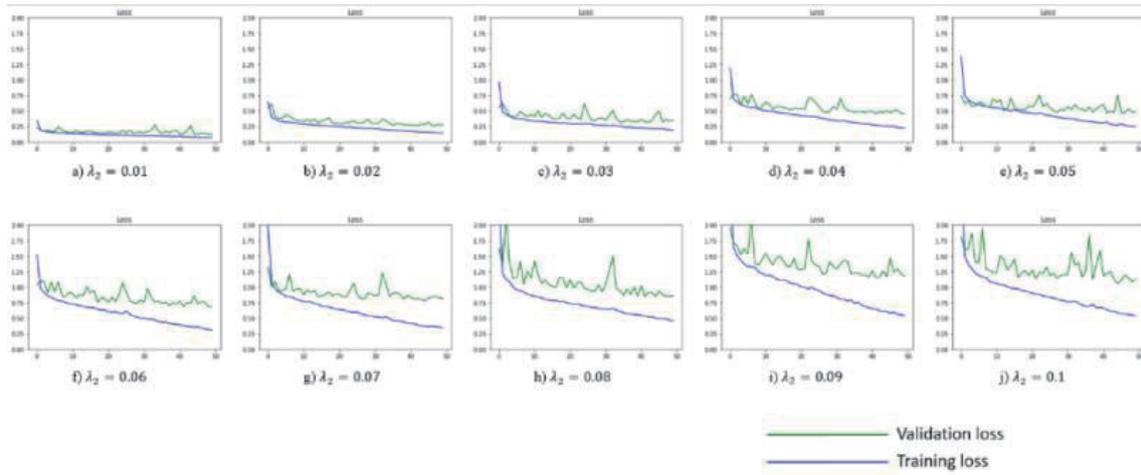


Fig. 11. (Color online) Effect of λ_2 on the training and validation processes. Here, λ_1 is set to be 1. The higher the λ_2 value, the harder for the network to converge

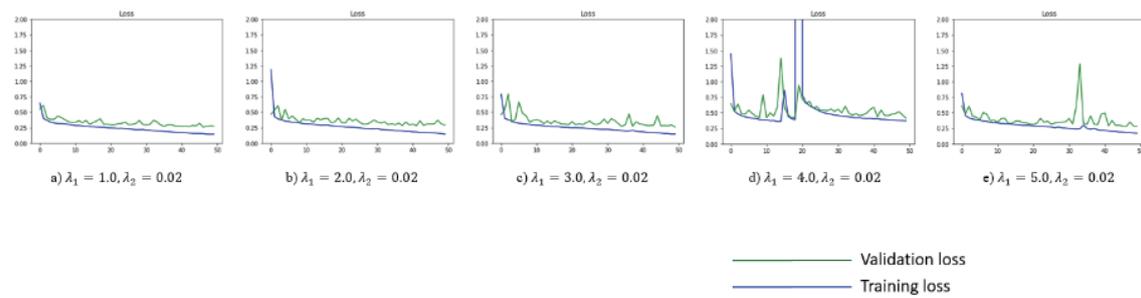


Fig. 12. (Color online) Effect of λ_1 on the training and validation processes. Here, λ_2 is set to be 0.02. The effect of λ_1 is not as significant as that of λ_2 .

Table 2.

Results of the second round of hyperparameter tuning on the UIEB-R90 dataset are shown as mean MSE, mean PSNR, and mean SSIM.

λ_1	λ_2	MSE	PSNR	SSIM
1	0.01	0.019	18.285	0.838
	0.05	0.022	17.71	0.836
2	0.01	0.016	17.888	0.831
	0.05	0.02	18.192	0.835
3	0.01	0.018	18.62	0.85
	0.05	0.02	18.06	0.837
4	0.01	0.032	16.01	0.802
	0.05	0.016	19.178	0.856
5	0.01	0.017	18.782	0.845
	0.05	0.015	19.32	0.867

5. Conclusions

Our proposed method enhances color sensing and recognition for underwater color using deep learning models and robust adaptive tone mapping technology to infer degradation models. The results of our research have significantly improved the effect of the previous work of Li *et al.*⁽⁵⁾ In our network, we tried to encode our features using Y’CbCr as well as RGB and CIELAB, and then we used an attention mechanism to highlight these features. The revised model also showed improved quantitative and qualitative results. We showed how different perceptual loss weights can considerably affect our results and training by monitoring our learning process and using quantitative metrics. As for the augmentation module, an appropriate tone map is applied to each region that was carefully segmented beforehand. The obtained images appear to have a greater dynamic range and significantly enhanced visible details compared with images inferred by the network alone and results from other methods. We avoided over-enhancement by only performing operations in large-scale layers of the image. When visualized in the CIELAB color space, the color gamut of our final result covers the largest volume compared with CIELAB color space. The volumes are also closer to the center of the color space and more evenly distributed across all axes. This demonstrates the effectiveness and robustness of our underwater recovery method, as it can be adapted to different types of underwater environment and constraint.

References

- 1 C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and P. Bekaert: IEEE Trans. Image Process **27** (2018) 1. <https://doi.org/10.1109/TIP.2017.2759252>
- 2 Y.-T. Peng and P. C. Cosman: IEEE Trans. Image Process **26** (2017) 4. <https://doi.org/10.1109/TIP.2017.2663846>
- 3 C. Li, J. Guo, and C. Guo: IEEE Signal Process Lett. **25** (2018) 323. <https://doi.org/10.1109/LSP.2018.2792050>
- 4 J. Li, K. A. Skinner, R. M. Eustice, and M. Johnson-Roberson: IEEE Rob. Autom. Lett. **3** (2018) 1. <https://doi.org/10.1109/LSP.2018.2792050>
- 5 C. Li, S. Anwar, J. Hou, R. Cong, C. Guo, and W. Ren: IEEE Trans Image Process **30** (2021) 4985. <https://doi.org/10.1109/TIP.2021.3076367>
- 6 K. He, X. Zhang, S. Ren, and J. Sun: Proc. 2016 IEEE Conf. Computer Vision and Pattern Recognition (IEEE, 2016) 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- 7 J. Hu, L. Shen, and G. Sun: Proc. 2018 IEEE Conf. Computer Vision and Pattern Recognition (IEEE, 2018) 7132–7141. <https://doi.org/10.1109/CVPR.2018.00745>
- 8 J. Johnson, A. Alahi, and L. Fei-Fei: Proc. 2016 European Conf. Computer Vision (2016) 694–711. <https://doi.org/10.48550/arXiv.1603.08155>
- 9 K.-J. Hu, M.-Y. Lu, J.-C. Wang, T.-I. Hsu, and T.-T. Chang: J. Image Video Process. (2010) 137134. <https://doi.org/10.1155/2010/137134>
- 10 Y.-T. Peng, K. Cao, and P. C. Cosman: IEEE Trans. Image Process **27** (2018) 6. <https://doi.org/10.1109/TIP.2018.2813092>
- 11 C. Li, S. Anwar, and F. Porikli: Pattern Recognit. **98** (2020) 107038. <https://doi.org/10.1016/j.patcog.2019.107038>
- 12 C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao: IEEE Trans. Image Process. **28** (2019) 4376. <https://doi.org/10.1109/TIP.2019.2955241>
- 13 J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei: Proc. 2009 IEEE Conf. Computer Vision and Pattern Recognition (IEEE, 2009) 248–255. <https://doi.org/10.1109/CVPR.2009.5206848>
- 14 C. Li, S. Anwar, and F. Porikli: Pattern Recognit. **98** (2019) 107038. <https://doi.org/10.1016/j.patcog.2019.107038>