Underwater Organism Color Fine-tuning via Decomposition and Guidance

Xiaofeng Cong[1], **Jie Gui**[1, 2, 3*], **Junming Hou**[4]

[1] School of Cyber Science and Engineering, Southeast University
[2] Engineering Research Center of Blockchain Application, Supervision And Management, Southeast University
[3] Ministry of Education, Purple Mountain Laboratories, Nanjing
[4] School of Information Science and Engineering, Southeast University
cxf_svip@163.com, guijie@seu.edu.cn, junming_hou@seu.edu.cn

Abstract

Due to the wavelength-dependent light attenuation and scattering, the color of the underwater organism usually appears distorted. The existing underwater image enhancement methods mainly focus on designing networks capable of generating enhanced underwater organisms with fixed color. Due to the complexity of the underwater environment, ground-truth labels are difficult to obtain, which results in the non-existence of perfect enhancement effects. Different from the existing methods, this paper proposes an algorithm with color enhancement and color fine-tuning (CECF) capabilities. The color enhancement behavior of CECF is the same as that of existing methods, aiming to restore the color of the distorted underwater organism. Beyond this general purpose, the color fine-tuning behavior of CECF can adjust the color of organisms in a controlled manner, which can generate enhanced organisms with diverse colors. To achieve this purpose, four processes are used in CECF. A supervised enhancement process learns the mapping from a distorted image to an enhanced image by the decomposition of color code. A self-reconstruction process and a cross-reconstruction process are used for content-invariant learning. A color fine-tuning process is designed based on the guidance for obtaining various enhanced results with different colors. Experimental results have proven the enhancement ability and color fine-tuning ability of the proposed CECF. The source code is provided in https://github.com/Xiaofeng-life/CECF.

Introduction

Underwater image enhancement (UIE) (Wei, Zheng, and Jia 2022; Mu, Qian, and Bai 2022; Tang et al. 2022) is an important research topic in the field of computer vision. Due to the absorption and attenuation of light, underwater images will suffer from color shift and detail distortion (Xu et al. 2023; González-Sabbagh and Robles-Kelly 2023; Jiang et al. 2022b). As the depth of water increases (Wang et al. 2023), the red, orange, yellow, green and blue lights gradually disappear, respectively. This means that blue and green lights are the last to disappear. As a result, underwater images that we typically observe tend to have a blue-green or blue effect. Correspondingly, the colors of colorful underwater organisms also become dull. To improve the visual quality of underwater images and organisms inside the images (Sun et al. 2022; Zhou et al. 2023), different UIE algorithms have been proposed. The current UIE algorithms are mainly divided into two categories (Zheng et al. 2022; Huang et al. 2022). They are non-deep learning-based UIE algorithms (NDL-UIE) and deep learning-based UIE algorithms (DL-UIE), respectively.

(a) $x \rightarrow \hat{y}_x^m$ (b) $\hat{y}_{x,g}^m$ (c) organism (d) g

Figure 1: The color enhancement process and color fine-tuning process of CECF. The color enhancement process is shown in (a), where the distorted underwater image x (in the first row) is enhanced to an enhanced image $\hat{y}_x^m$ (in the second row, where $\alpha = 0.3$). The color fine-tuning process is shown in (b) and (c). When guidances are given and the controllable parameter α is set, the color of $\hat{y}_x^m$ will change toward the guidance. The guidance g, as shown in (d), can be an underwater image or a natural image.

Statistical assumptions and physical models are widely adopted by the existing NDL-UIE methods, such as SMBL (Song et al. 2020), MLLE (Zhang et al. 2022), and ICSP (Hou et al. 2023a). In addition, digital image enhancement methods (Raveendran, Patil, and Birajdar 2021; Jian et al. 2021) that do not utilize physical model and image formation model are also used in NDL-UIE. DL-UIE methods mainly focus on learning the mapping from distorted images to clear images, such as UWNet (Naik, Swarnakar, and Mittal 2021), ADMNNet (Yan et al. 2022), CLUIE (Li et al. 2023), and SGUIE (Qi et al. 2022).

The well-designed UIE algorithms may help underwater photographers restore the color information of the underwater organism. However, existing research (Fu et al. 2022) has shown that the color information of distorted image fails to be restored perfectly. To alleviate this issue, UIESS (Chen and Pei 2022) proposes to obtain diverse enhanced results

*Corresponding author

by adjusting the enhancement level. Different from UIESS, this paper proposes a novel UIE method, namely Color Enhancement and Color Fine-tuning (CECF) algorithm, with the ability of color fine-tuning[1].

The main innovation of this paper is shown in Figure 1. To achieve the color fine-tuning capability shown in Figure 1, we propose a scheme to obtain the color code and content code during the enhancement process, where the color code and content code represent the visual color information and content information of the enhanced image. By manipulating the color code of the guide image and the distorted image, various enhanced images can be obtained. CECF is designed as four processes as shown in Figure 2, namely enhancement, self-reconstruction, cross-reconstruction and color fine-tuning process. The contributions of this paper include the following three points.

- We propose a method called CECF, which is capable of fine-tuning the color of enhanced organisms by selecting different guidances and controllable parameters. For the selection of guidance, we present the observation that the hue of long-wavelength colors is approximately invariant during the enhancement process.
- An encoder-decoder network is adopted to extract the color code and content code of the underwater image during the enhancement process. By fusing the color codes of the guidance and the distorted image during the fine-tuning process, the color effect of the enhanced image can be fine-tuned in a controllable way.
- We introduce a three-step joint training procedure since the color code and content code cannot be directly learned. A self-reconstruction process and a cross-reconstruction process are adopted to assist in the learning of content code during the enhancement process. Under this setting, the color code will be learned implicitly.

CECF can adjust the color of a specific organism when the mask of this organism is provided by well-studied segmentation algorithms (Islam et al. 2020; Nezla, Haridas, and Supriya 2021; Kim and Park 2022; Alshdaifat, Talib, and Osman 2020). Since segmentation is not the research scope of this paper, we do not discuss how to segment underwater organisms.

Related Work

NDL-UIE

NDL-UIE methods (Ding et al. 2022) mainly adopt statistical assumptions and physical models (Liu et al. 2023; Xue et al. 2023). ICSP (Hou et al. 2023a) proposes an illumination channel sparsity prior, which can be used in a guided variational framework for nonuniform illumination situations. SMBL (Song et al. 2020) designs a background light estimation method for underwater image restoration and color correction, which is based on the statistical analysis of the distribution characteristics of each channel. A locally adaptive color correction method built on the minimum color loss principle is proposed by MLLE (Zhang et al. 2022). HLRP (Zhuang et al. 2022) proposes a hyper-laplacian reflectance prior, which is based on the retinex variational model. These NDL-UIE methods are proven effective for the color correction of underwater images. However, limited by the complexity of the underwater situation, they may encounter difficulties in parameter estimation (Qi et al. 2022).

DL-UIE

Deep learning (Hou et al. 2023b) are widely used in image process tasks (Gui et al. 2023). DL-UIE methods (Anil, Sreelatha et al. 2023; Wang et al. 2021; Sharma, Bisht, and Sur 2023; Qiao, Dong, and Sun 2022; Jiang et al. 2022a; Liu et al. 2022; Kang et al. 2022) typically rely on underwater datasets with a sufficient number of images. An attention-guided dynamic multibranch neural network is designed by ADMNNet (Yan et al. 2022), which is shown to be effective in extracting the multiscale features. Aiming at improving the deployability of enhancement models, UWNet (Naik, Swarnakar, and Mittal 2021) is proposed to reduce the compute and memory resources. A fully-convolutional conditional GAN-based mode and a multimodal objective function are used by FUnIE (Islam, Xia, and Sattar 2020). Semi-UIR (Huang et al. 2023) uses semi-supervised and contrastive learning strategies for incorporating unlabeled data into network training. Dilated convolution and residual convolution are used by PhyNN (Chen et al. 2021) and URes (Liu et al. 2019) for the network design of UIE, respectively. SGUIE (Qi et al. 2022) introduces semantic information into the training process to improve feature learning performance. These DL-UIE methods have been proven to be effective in enhancing underwater images. However, as a recent study (Chen and Pei 2022) points out, diverse outputs can not be obtained by these methods.

Existing studies (Chen and Pei 2022; Li et al. 2023) have pointed out that it is difficult to obtain accurate reference (ground-truth) underwater images. This means that it is also challenging for DL-UIE models to perfectly restore the color of underwater images and organisms. Recent studies (Ye et al. 2022; Kim, Park, and Kwon 2021; Chen and Pei 2022) have proposed methods that enable the acquisition of diverse underwater images. A light field retention method on underwater images is proposed by UWNR (Ye et al. 2022) to obtain diverse rendering images. However, it is mainly used for scene rendering rather than underwater image enhancement. A pixel-wise wasserstein autoencoder (PWAE) (Kim, Park, and Kwon 2021) for image dehazing task, is verified for style transformation for UIE. However, the training process of PWAE needs to use an extra dataset for the training of style transformation. UIESS (Chen and Pei 2022) adopts the strategy of separation of content and style, but its coding space is used to manipulate the degree of image enhancement instead of color control. Different from UWNR, PWAE and UIESS, the CECF proposed in this paper is used to control the color of the enhanced organism for the UIE task.

Method

The overall training and inference process of CECF are shown in Figure 2, which contains four processes in total.

[1]Noticed, in this paper, the "color" in "color fine-tuning" refers to the tint, shade and tone.

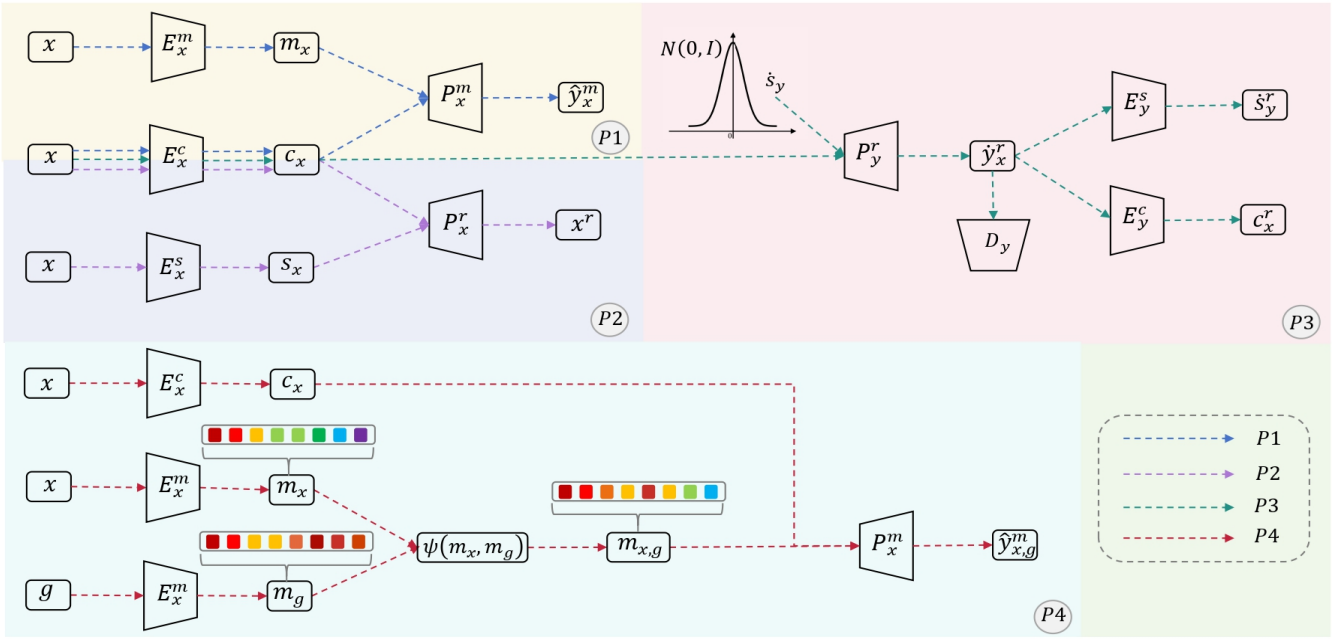


Figure 2: The overall pipeline of CECF. The $P1$, $P2$, $P3$ and $P4$ denote the enhancement, self-reconstruction, cross-reconstruction and fine-tuning process, respectively.

The distorted underwater image is denoted as $x \in \mathbb{X}$, and the clear underwater image is represented by $y \in \mathbb{Y}$. The guidance is marked as $g \in \mathbb{G}$. The enhanced image obtained by CECF is $\hat{y}_x^m$. The diversified underwater enhanced images are denoted by $\hat{y}_{x,g}^m$. There are three encoders for processing x, namely E_x^c, E_x^m and E_x^s. Two decoders are used to process the encoded information of x, namely P_x^m and P_x^r. There are two encoders and one decoder for y, which are E_y^s, E_y^c and P_y^r. Furthermore, the discriminators for the x domain and y domain are denoted as D_x and D_y, respectively. In the notation of encoder and decoder and discriminator, the abbreviations are color (m), encoder (E), decoder (P), style (s), content (c), discriminator (D), reconstruction (r), distortion domain (x), clear domain (y). Therefore, a total of 10 networks are involved in the training process. In the inference stage, only two encoders (E_x^m and E_x^c) and one decoder P_x^m are used. The role of each network and the meaning of symbols will be introduced in this section.

Enhancement Process

As shown in Figure 2, the aim of the enhancement process $P1$ is to improve the quality of distorted image x. In this process, our method can achieve the same purpose as existing single-output methods (Yan et al. 2022; Hu et al. 2022), that is to obtain the enhanced image $\hat{y}_x^m$ with fixed color. Beyond this general purpose, we also aim at fine-tuning the color of $\hat{y}_x^m$. Therefore, we split the enhancement process $P1$ into two parts, which are the encoding path and decoding path, respectively. Two codes with specific semantic meanings are learned in the encoding path. The content encoder E_x^c is responsible for learning content code c_x, which represents the content and structural information of x. The color encoder E_x^m is designed for learning the color code m_x, which denotes the color information for enhancement. The content code and color code are obtained by

$$c_x = E_x^c(x), \tag{1}$$

$$m_x = E_x^m(x). \tag{2}$$

Then, with the help of color decoder P_x^m, we are able to obtain the enhanced image $\hat{y}_x^m$ by

$$\hat{y}_x^m = P_x^m(c_x, m_x). \tag{3}$$

The enhancement network composed of E_x^c, E_x^m and P_x^m is Underwater Enhancement Network (UE-Net). Among all sub-networks of CECF, UE-Net is the only required sub-network in the inference stage. To constrain the $\hat{y}_x^m$ and the ground truth y, the enhancement process is optimized by

$$\mathcal{L}^m = \mathbb{E}_{x,y\sim p(x,y)}[||\hat{y}_x^m - y||_2 - \phi(\hat{y}_x^m, y)], \tag{4}$$

where $p(x, y)$ denotes the joint data distribution of paired data (x, y). The function $-\phi(\cdot, \cdot)$ is the Structural Loss proposed by (Zhao, Gallo, and Frosio 2016).

Color Fine-tuning Process

As depth increases, the light with longer wavelength disappears earlier. The order of disappearance is red, orange, yellow, green and blue (Wang et al. 2023). Therefore, as the depth of the water increases, the underwater images we see will have fewer colors with long wavelengths (such as pure red, orange and yellow) (González-Sabbagh and Robles-Kelly 2023). The datasets may contain images with less distortion, which approximately preserve the long-wavelength colors, such as approximately pure red, orange, and yellow.

At this point, our enhancement network should ensure that these long wavelength colors are approximately preserved. This means that the hue corresponding to these colors should not be changed. We refer to this observation as the hue invariance of the long wavelength colors.

Following this observation, we introduce the concept of guidance g. Specifically, we refer to the images with long wavelength colors as guidances. During the training process of CECF, the hue of the g is approximately invariant. Therefore, we can incorporate the color code m_g of g into the color code m_x of x to change the color of x. The color fine-tuning process is shown in $P4$ of Figure 2. For a trained UE-Net, the color controller is defined as $f(\cdot,\cdot,\cdot)$. Given $x \in \mathbb{X}$, the color controller $f(\cdot,\cdot,\cdot)$ can obtain the enhanced image with fix color as

$$\hat{y}_x^m = f(x, x, 0). \tag{5}$$

Beyond this purpose, CECF aims to obtain color controllable output $\hat{y}_{x,g}^m$ under the guide of g. Specifically, we hope to obtain color-diversified $\hat{y}_{x,g}^m$ while the content of the scene remains unchanged, as

$$\hat{y}_{x,g}^m = f(x, g, \alpha), \tag{6}$$

where g is a guidance and α denotes the degree of color change. The color code m_x can be transformed to generate color-controllable outputs. The transformation function t (Mi et al. 2021) of color code can be expressed as

$$t : \mathbb{R}^N \times \mathbb{R}^N \times [0,1] \ni (m_x, m_g, \alpha) \mapsto m_x, m_g \in \mathbb{R}^N, \tag{7}$$

where N denotes the dimensionality. Selecting a guidance g, the color code m_g corresponding to g can be obtained by

$$m_g = E_x^m(g). \tag{8}$$

Then, the m_x and m_g are fused in the encoding space by

$$m_{x,g} = \psi(m_x, m_g) = \frac{(1-\alpha)\cdot m_x + \alpha \cdot \kappa(m_g)}{\sqrt{(1-\alpha)^2 + \alpha^2}}, \tag{9}$$

where κ denotes the truncation operation, so that the value of m_g is clipped in to the specific range. Then, the color controllable enhanced image $\hat{y}_{x,g}^m$ is obtained by

$$\hat{y}_{x,g}^m = P_x^m(c_x, m_{x,g}). \tag{10}$$

When the mask of the organism is available (Laradji et al. 2021; Patil et al. 2019; Chen et al. 2019; Zhang, Wu, and Bao 2022; Li et al. 2021), the organism can be obtained by $\hat{y}_{x,g}^m \odot mask$, where $\odot$ is element-wise multiplication.

Three-step Joint Training by Self-reconstruction and Cross-reconstruction Processes

Based on the above purposes, an important question is how to ensure that the codes learned by E_x^c and E_x^m have the semantic features we want. It is difficult to optimize the E_x^c and E_x^m directly during the training of the enhancement process since we have no idea what the color code and content code look like. Therefore, we decompose the training process into a three-step solution.

Suppose the E_x^c can represent the content information in each iteration. Then, another encoder E_x^m is able to represent the color information by optimizing the enhancement loss (Equation 4) in a supervised way. It is worth pointing out that the semantics of E_x^c and E_x^m are related to a particular P_x^m. Without the decoder, the outputs of the encoders are meaningless. Inspired by existing study (Huang et al. 2018), we adopt a self-reconstruction process and a cross-reconstruction process to assist the enhancement process. These two processes can help E_x^c to learn the content information of the scene.

The self-reconstruction $P2$ and cross-reconstruction $P3$ in the $\mathbb{X}$ domain are shown in Figure 2. The pipeline for the self-reconstruction and cross-reconstruction processes in $\mathbb{Y}$ domain are symmetrical with $P2$ and $P3$. E_x^s is responsible for learning the style code (Huang et al. 2018; Jain, Matta, and Mitra 2022) of x. It is worth pointing out that the style code is different from the color code proposed in this paper. An enhancement result close to the reference image can be obtained by color code but not by style code. E_y^c and E_y^s are used to learn the content information and style information of y, respectively. Meanwhile, decoder P_x^r and P_y^r are responsible for the reconstruction of x and y, respectively.

Self-reconstruction The purpose of self-reconstruction is that x and y can be reconstructed after encoding and decoding, as follows

$$\begin{aligned}\mathcal{L}_{x,y}^r = \mathbb{E}_{x,y\sim p(x,y)}[&||P_x^r(E_x^c(x), E_x^s(x)) - x||_1 + \\ &||P_y^r(E_y^c(y), E_y^s(y)) - y||_1].\end{aligned} \tag{11}$$

Cross-reconstruction For the content code, after cross-domain encoding and decoding, it can be reconstructed as

$$\begin{aligned}\mathcal{L}_{c_x,c_y}^r = &\mathbb{E}_{c_x\sim p(c_x),\dot{s}_y\sim q(\dot{s}_y)}[||E_y^c(P_y^r(c_x,\dot{s}_y)) - c_x||_1] + \\ &\mathbb{E}_{c_y\sim p(c_y),\dot{s}_x\sim q(\dot{s}_x)}[||E_x^c(P_x^r(c_y,\dot{s}_x)) - c_y||_1],\end{aligned} \tag{12}$$

where $q(\dot{s}_x)$ and $q(\dot{s}_y)$ denote the Gaussian prior $\mathbb{N}(0, I)$. The c_x and c_y denote $E_x^c(x)$ and $E_y^c(y)$, respectively. $E_y^c(P_y^r(c_x, \dot{s}_y))$ denotes the c_x^r in Figure 2. For the style code, the reconstruction loss after cross-domain encoding and decoding is

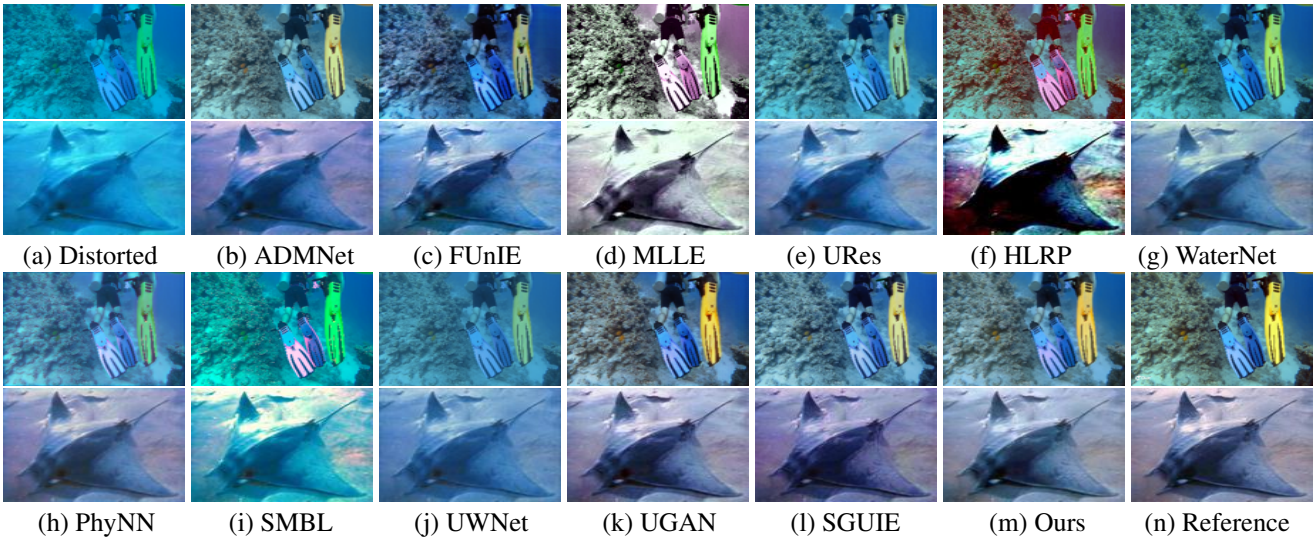
$$\begin{aligned}\mathcal{L}_{\dot{s}_x,\dot{s}_y}^r = &\mathbb{E}_{c_x\sim p(c_x),\dot{s}_y\sim q(\dot{s}_y)}[||E_y^s(P_y^r(c_x,\dot{s}_y)) - \dot{s}_y||_1] + \\ &\mathbb{E}_{c_y\sim p(c_y),\dot{s}_x\sim q(\dot{s}_x)}[||E_x^s(P_x^r(c_y,\dot{s}_x)) - \dot{s}_x||_1]\end{aligned} \tag{13}$$

where $E_y^s(P_y^r(c_x, \dot{s}_y))$ denotes the $\dot{s}_y^r$ in Figure 2. GAN loss (Goodfellow et al. 2014) is also used to match the distribution of generated images to the distribution of real images. Adversarial losses are

$$\begin{aligned}\mathcal{L}_x^D = &\mathbb{E}_{c_y\sim p(c_y),\dot{s}_x\sim q(\dot{s}_x)}[\log(1 - D_x(P_x^r(c_y,\dot{s}_x)))] + \\ &\mathbb{E}_{x\sim p(x)}[\log D_x(x)],\end{aligned} \tag{14}$$

$$\begin{aligned}\mathcal{L}_y^D = &\mathbb{E}_{c_x\sim p(c_x),\dot{s}_y\sim q(\dot{s}_y)}[\log(1 - D_y(P_y^r(c_x,\dot{s}_y)))] + \\ &\mathbb{E}_{y\sim p(y)}[\log D_y(y)],\end{aligned} \tag{15}$$

where $P_y^r(c_x, \dot{s}_y)$ denote the $\dot{y}_x^r$ in Figure 2. The purpose of all losses ($\mathcal{L}_{x,y}^r$, $\mathcal{L}_{c_x,c_y}^r$, $\mathcal{L}_{\dot{s}_x,\dot{s}_y}^r$, $\mathcal{L}_x^D$, $\mathcal{L}_y^D$) mentioned above is to assist the learning of c_x by E_x^c.



(a) Distorted (b) ADMNet (c) FUnIE (d) MLLE (e) URes (f) HLRP (g) WaterNet

(h) PhyNN (i) SMBL (j) UWNet (k) UGAN (l) SGUIE (m) Ours (n) Reference

Figure 3: Comparison of visual results. The scenes in the first and second rows are from UIEB and EUVP-I, respectively.

Loss Functions

In order to train UE-Net to obtain the color code and content code, three processes are jointly optimized, which are enhancement, self-reconstruction and cross-reconstruction. The overall loss function is

$$\begin{aligned}\min_{\mathbb{M}}\max_{\mathbb{N}}\mathcal{L}(\mathbb{M},\mathbb{N}) =&\mathcal{L}_x^D+\mathcal{L}_y^D+\lambda_1\cdot(\mathcal{L}^m+\mathcal{L}_{x,y}^r)+\\ &\lambda_2\cdot(\mathcal{L}_{c_x,c_y}^r+\mathcal{L}_{\dot{s}_x,\dot{s}_y}^r),\end{aligned} \tag{16}$$

where λ_1 and λ_2 are factors that control the weight of each loss function. The networks represented by $\mathbb{M}$ and $\mathbb{N}$ are

$$\mathbb{M}=\{E_x^m,E_x^c,E_x^s,E_y^c,E_y^s,P_x^r,P_x^m,P_y^r\}, \tag{17}$$

$$\mathbb{N}=\{D_x,D_y\}. \tag{18}$$

Experiments

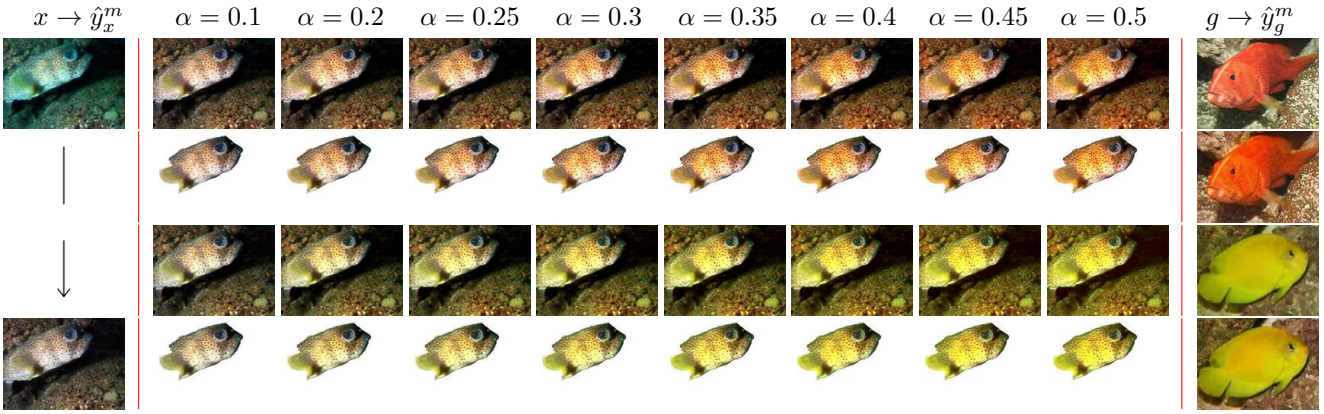
Settings

Experiments for evaluation are conducted on two datasets. In detail, the UIEB (Li et al. 2019) dataset is split into 800 training images and 90 test images. The Underwater ImageNet (EUVP-I) in the EUVP (Islam, Xia, and Sattar 2020) dataset is split into 3300 training images and 400 test images. To ensure that the color code can be learned by CECF, the dataset needs to contain a subset of images with less distortion. Therefore, in this paper, the Underwater Scenes (EUVP-S) provided by EUVP are used for analyzing the color fine-tuning ability. The EUVP-S is split into 1967 training images and 218 test images. The masks for testing are from (Islam, Luo, and Sattar 2020). Evaluation metrics include full-reference metrics Peak Signal-to-Noise Ratio (PSNR) (Shen, Zhao, and Zhang 2023; Shen et al. 2023) and Structural Similarity Measure (SSIM) (Cong et al. 2020). No-reference metrics include Underwater Image Quality Measure (UIQM) (Panetta, Gao, and Agaian 2015) and Underwater Color Image Quality Evaluation (UCIQE) (Yang and Sowmya 2015; Qi et al. 2022).

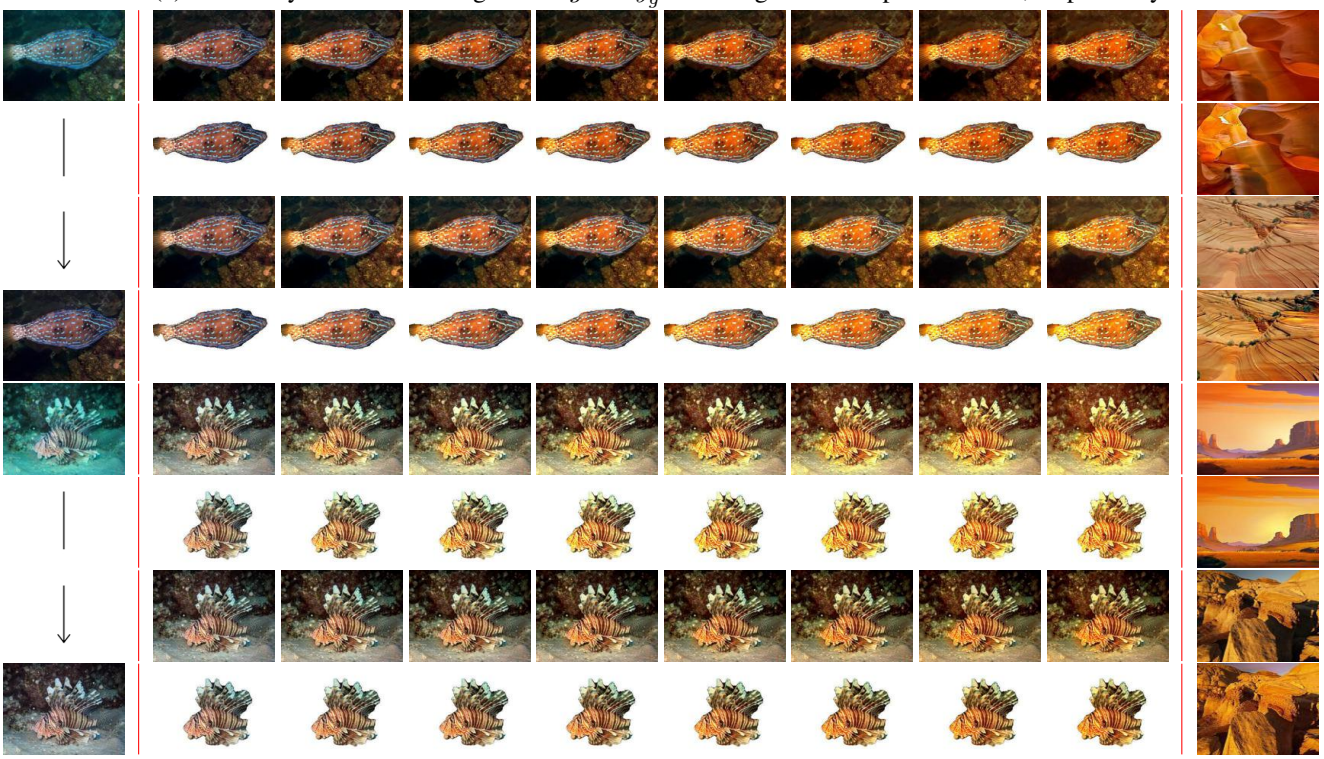
The interval of the truncation operation $\kappa(\cdot)$ of m_g is $[-3,3]$. The learning rate is set to 0.0001, the Adam optimizer is used, and the batch size is 4. The values of λ_1 and λ_2 are 10 and 1, respectively. We need to emphasize two points. First, compared with (Huang et al. 2018), color encoder E_x^m and color decoder P_x^m are newly added networks in this paper. Since the network design is not the main research purpose of this paper, our color encoder E_x^m and color decoder P_x^m simply use the similar structures to the Style Encoder and Style Decoder proposed by (Huang et al. 2018). The details about E_x^m and P_x^m are placed at Supplementary Material (see our code repository). The rest networks remain the same as (Huang et al. 2018). The method in (Huang et al. 2018) cannot provide fixed results and it cannot be used as a comparative experiment. Second, the UIE algorithms for comparison include ADMNet (Yan et al. 2022), FUnIE (Islam, Xia, and Sattar 2020), MLLE (Zhang et al. 2022), URes (Liu et al. 2019), HLRP (Zhuang et al. 2022), WaterNet (Li et al. 2019), PhyNN (Chen et al. 2021), SMBL (Song et al. 2020), UWNet (Naik, Swarnakar, and Mittal 2021), UGAN (Fabbri, Islam, and Sattar 2018) and SGUIE (Qi et al. 2022), respectively.

Quantitative and Visual Results

When comparing with existing UIE algorithms, no guiding image is adopted by CECF. The quantitative evaluation values in Table 1 show that CECF achieves relatively good results on both PSNR and SSIM. Moreover, for the no-reference evaluation metric UIQM, CECF still outperforms most algorithms. Figure 3 shows the visual effect comparison of CECF and other UIE algorithms. The visual enhancement effects obtained by various algorithms demonstrate that CECF has a relatively better effect on color restoration. Overall, quantitative and visual results demonstrate the effectiveness of CECF on the UIE task.



(a) Guided by underwater images. The g and $\hat{y}_g^m$ are images at the top and bottom, respectively.



(b) Guided by natural images. The g and $\hat{y}_g^m$ are images at the top and bottom, respectively.

Figure 4: Visual comparisons when taking different images as guidances.

Underwater and Natural Images as Guidance

Natural images with long wavelength colors are easier to acquire than underwater images with long wavelength colors. Therefore, the generalization ability of CECF in natural images is worthy of exploration. The color fine-tuning ability of CECF is verified in underwater images in Figure 4-(a) and natural images in Figure 4-(b). The experimental results provide two conclusions. First, the degree of the color transformation increases as α increases. Second, different rendering effects can be obtained using different guidances.

Ablation Study and Discussions

Length of Color Code When the length of the color code is set to 8, 16, 24 and 32, the corresponding evaluation results are shown in Table 2. The results demonstrate that color codes of different lengths do not significantly affect the quantitative evaluation results. Therefore, setting the color code to 8 is a reasonable choice.

Values of Color Code During the training process of CECF, the constraint of color code is added implicitly. Therefore, we need to explore whether the neurons of E_x^m are dead ($E_x^m(x) \approx 0$) or not. The histograms of the training outputs from the first two dimensions (the length of m_x

Table 1: Quantitative results on UIEB and EUVP-I. The best results are in bold.

Config.	UIEB				EUVP-I			
	PSNR↑	SSIM↑	UIQM↑	UCIQE↑	PSNR↑	SSIM↑	UIQM↑	UCIQE↑
ADMNet	21.138	0.882	2.887	0.608	24.668	0.844	3.000	0.557
FUnIE	18.707	0.800	3.081	0.607	21.899	0.796	3.032	0.567
MLLE	17.207	0.739	1.991	0.617	16.233	0.608	2.519	0.583
URes	18.324	0.841	2.892	0.580	23.327	0.825	3.016	0.541
HLRP	16.519	0.718	2.463	**0.667**	15.217	0.562	2.499	**0.628**
WaterNet	21.198	0.858	2.978	0.612	23.673	0.832	2.991	0.562
PhyNN	18.386	0.802	2.785	0.594	23.583	0.814	2.983	0.553
SMBL	16.647	0.764	2.050	0.631	16.083	0.651	2.195	0.617
UWNet	17.453	0.794	2.773	0.553	22.485	0.815	2.960	0.532
UGAN	21.472	0.860	**3.276**	0.622	24.586	0.833	**3.053**	0.557
SGUIE	21.757	**0.894**	3.058	0.609	24.797	0.840	3.038	0.559
Ours	**21.819**	**0.894**	3.122	0.615	**25.156**	**0.856**	2.991	0.558

Table 2: Evaluation of different length (denoted by l) of m_x. The values in the table are the variations relative to $l = 8$.

Config.	UIEB			EUVP-I		
	PSNR↑	SSIM↑	UIQM↑	PSNR↑	SSIM↑	UIQM↑
$l = 16$	**-0.1148**	0.0017	-0.0185	-0.0294	0.0002	0.0318
$l = 32$	-0.1459	0.0025	-0.0138	-0.0415	**0.0036**	0.0319
$l = 48$	-0.1159	**0.0034**	**-0.0116**	**0.0861**	0.0025	**0.0369**

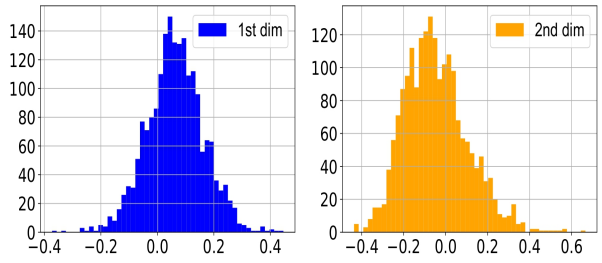


Figure 5: The histogram of m_x, where the x and y axis coordinates represent the value range and amount, respectively.

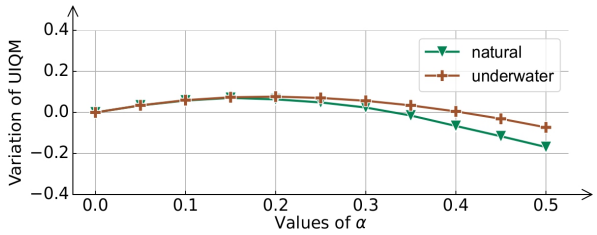


Figure 6: The variation curves of UIQM when $\alpha \in [0, 0.5]$.

is 8) are shown in Figure 5. The results prove that E_x^m is capable of providing a non-zero numerical output.

Retention of Content Information When performing color fine-tuning, we aim at the contents of image keep unchanged. It can be seen from the visual effect in Figure 1 and Figure 4 that the contents of the images are preserved.

The Values of Weight Factors The Equation 16 contains two weighting factors, which are λ_1 and λ_2, respectively. Since the values of λ_1 and λ_2 are theoretically infinite. We empirically set λ_2 to 1. Then, the ratio of $\lambda_1 : \lambda_2$ is set to five representative values, namely 10, 8, 6, 4 and 2, respectively. The results (on UIEB) in Table 4 show that the influence of the weight ratio on the enhancement performance is not obvious. Setting the λ_1 to 6 or 10 is a reasonable choice. Thus, we set λ_1 and λ_2 to 10 and 1, respectively.

Quantitative Evaluation of Color Fine-tuning To verify the quality of the diversified enhanced images obtained by the color controller $f(\cdot, \cdot, \cdot)$, the changes of no-reference evaluation metric UIQM ($UIQM(f(x, g, \alpha)) - UIQM(f(x, x, 0))$) are calculated on UIEB test dataset and presented in Figure 6. The guidances we choose are the underwater and natural images in Figure 1. As shown in Figure 6, when α increases, the value of UIQM first increases and then decreases. Overall, we can draw two conclusions. First, the results show that the value range of α is relatively reasonable between $[0, 0.3]$. Second, the quality of the enhanced images obtained with different guidances is different.

Evaluation of Structure Loss The evaluation results in Table 3 show the ablation of the Structure Loss (SL) which is used in Equation 4. The results show that the performance on the UIEB is significantly improved, especially the SSIM value is improved from 0.883 to 0.894. The improvement of SL on the EUVP-I is not obvious. The reason may be that the EUVP-I has relatively simple degradation since the distorted images are generated by a single algorithm (Islam, Xia, and Sattar 2020). Overall, the use of SL is beneficial to CECF.

Table 3: Evaluation of Structural Loss (SL).

Config.	UIEB			EUVP-I		
	PSNR↑	SSIM↑	UIQM↑	PSNR↑	SSIM↑	UIQM↑
SL	**21.819**	**0.894**	**3.122**	**25.156**	**0.856**	2.991
No-SL	21.693	0.883	3.117	25.040	**0.856**	**3.018**

Table 4: Evaluation of Weight Factors. The values in the table represent the improvements compared to $\lambda_1 : \lambda_2 = 10$.

Config.	PSNR↑	SSIM↑	UIQM↑	PSNR↑	SSIM↑	UIQM↑
	$\lambda_1 : \lambda_2 = 2$			$\lambda_1 : \lambda_2 = 4$		
Values	-0.282	-0.007	-0.024	-0.307	-0.004	0.012
	$\lambda_1 : \lambda_2 = 6$			$\lambda_1 : \lambda_2 = 8$		
Values	**-0.065**	**0.001**	**0.028**	-0.288	-0.002	-0.025

Hue Approximately-invariant to the Guidance Figures 4 show the original guidance g and the corresponding enhanced $\hat{y}_g^m$. The results demonstrate that the hues they belong to are approximately invariant. This proves that the m_g approximately preserves the color information of g.

How to Choose the Guidance There are no strict restrictions on the selection of guidance, except that the dominant colors in g should be of long wavelength. The color fine-tuning effect obtained by different guidances may be different. When users want to determine whether an image g can be used as a guidance, all that is required is to put the g in the color controller $f(g, g, 0)$ to obtain $\hat{y}_g^m$. This g can serve as suitable guidance if the hue of the $\hat{y}_g^m$ is almost unchanged.

Conclusions

This paper proposes an algorithm capable of fine-tuning the color of underwater organisms via decomposition and guidance. A three-step joint training procedure is adopted to learn the color code. The self-reconstruction and cross-reconstruction strategies are used as auxiliary training for the learning of the contents of the scene. Quantitative and qualitative experimental results show that underwater organisms with diverse colors can be obtained by manipulating the color code.

Acknowledgments

This work was supported in part by the grant of the National Science Foundation of China under Grant 62172090; Start-up Research Fund of Southeast University under Grant RF1028623097; CAAI-Huawei MindSpore Open Fund. We thank the Big Data Computing Center of Southeast University for providing the facility support on the numerical calculations in this paper.

References

Alshdaifat, N. F. F.; Talib, A. Z.; and Osman, M. A. 2020. Improved deep learning framework for fish segmentation in underwater videos. *Ecological Informatics*, 59: 101121.

Anil, A.; Sreelatha, G.; et al. 2023. A Novel Approach for Underwater Image Enhancement Using Style Transfer Technique. In *International Conference on Control, Communication and Computing*, 1–6.

Chen, X.; Zhang, P.; Quan, L.; Yi, C.; and Lu, C. 2021. Underwater image enhancement based on deep learning and image formation model. *arXiv preprint arXiv:2101.00991*.

Chen, Y.-W.; and Pei, S.-C. 2022. Domain Adaptation for Underwater Image Enhancement via Content and Style Separation. *IEEE Access*, 10: 90523–90534.

Chen, Z.; Sun, Y.; Gu, Y.; Wang, H.; Qian, H.; and Zheng, H. 2019. Underwater object segmentation integrating transmission and saliency features. *IEEE Access*, 7: 72420–72430.

Cong, X.; Gui, J.; Miao, K.-C.; Zhang, J.; Wang, B.; and Chen, P. 2020. Discrete haze level dehazing network. In *Proceedings of the 28th ACM International Conference on Multimedia*, 1828–1836.

Ding, X.; Wang, Y.; Liang, Z.; and Fu, X. 2022. A unified total variation method for underwater image enhancement. *Knowledge-Based Systems*, 255: 109751.

Fabbri, C.; Islam, M. J.; and Sattar, J. 2018. Enhancing underwater imagery using generative adversarial networks. In *IEEE International Conference on Robotics and Automation*, 7159–7165.

Fu, Z.; Wang, W.; Huang, Y.; Ding, X.; and Ma, K.-K. 2022. Uncertainty inspired underwater image enhancement. In *European Conference on Computer Vision*, 465–482.

González-Sabbagh, S. P.; and Robles-Kelly, A. 2023. A survey on underwater computer vision. *ACM Computing Surveys*.

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative adversarial nets. *Advances in neural information processing systems*, 1–9.

Gui, J.; Cong, X.; Cao, Y.; Ren, W.; Zhang, J.; Zhang, J.; Cao, J.; and Tao, D. 2023. A comprehensive survey and taxonomy on single image dehazing based on deep learning. *ACM Computing Surveys*, 55(13s): 1–37.

Hou, G.; Li, N.; Zhuang, P.; Li, K.; Sun, H.; and Li, C. 2023a. Non-uniform Illumination Underwater Image Restoration via Illumination Channel Sparsity Prior. *IEEE Transactions on Circuits and Systems for Video Technology*, 1–13.

Hou, J.; Cao, Q.; Ran, R.; Liu, C.; Li, J.; and Deng, L.-j. 2023b. Bidomain Modeling Paradigm for Pansharpening. In

Proceedings of the ACM International Conference on Multimedia, 347–357.

Hu, K.; Weng, C.; Zhang, Y.; Jin, J.; and Xia, Q. 2022. An overview of underwater vision enhancement: From traditional methods to recent deep learning. *Journal of Marine Science and Engineering*, 10(2): 241.

Huang, S.; Wang, K.; Liu, H.; Chen, J.; and Li, Y. 2023. Contrastive semi-supervised learning for underwater image restoration via reliable bank. In *IEEE Conference on Computer Vision and Pattern Recognition*, 18145–18155.

Huang, X.; Liu, M.-Y.; Belongie, S.; and Kautz, J. 2018. Multimodal unsupervised image-to-image translation. In *Proceedings of the European Conference on Computer Vision*, 172–189.

Huang, Z.; Li, J.; Hua, Z.; and Fan, L. 2022. Underwater image enhancement via adaptive group attention-based multiscale cascade transformer. *IEEE Transactions on Instrumentation and Measurement*, 71: 1–18.

Islam, M. J.; Edge, C.; Xiao, Y.; Luo, P.; Mehtaz, M.; Morse, C.; Enan, S. S.; and Sattar, J. 2020. Semantic segmentation of underwater imagery: Dataset and benchmark. In *IEEE International Conference on Intelligent Robots and Systems*, 1769–1776.

Islam, M. J.; Luo, P.; and Sattar, J. 2020. Simultaneous Enhancement and Super-Resolution of Underwater Imagery for Improved Visual Perception. In *16th Robotics: Science and Systems, RSS 2020*. MIT Press Journals.

Islam, M. J.; Xia, Y.; and Sattar, J. 2020. Fast underwater image enhancement for improved visual perception. *IEEE Robotics and Automation Letters*, 5(2): 3227–3234.

Jain, N.; Matta, G. R.; and Mitra, K. 2022. Towards Realistic Underwater Dataset Generation and Color Restoration. In *Proceedings of the Thirteenth Indian Conference on Computer Vision, Graphics and Image Processing*, 1–9.

Jian, M.; Liu, X.; Luo, H.; Lu, X.; Yu, H.; and Dong, J. 2021. Underwater image processing and analysis: A review. *Signal Processing: Image Communication*, 91: 116088.

Jiang, Q.; Zhang, Y.; Bao, F.; Zhao, X.; Zhang, C.; and Liu, P. 2022a. Two-step domain adaptation for underwater image enhancement. *Pattern Recognition*, 122: 108324.

Jiang, Z.; Li, Z.; Yang, S.; Fan, X.; and Liu, R. 2022b. Target oriented perceptual adversarial fusion network for underwater image enhancement. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(10): 6584–6598.

Kang, Y.; Jiang, Q.; Li, C.; Ren, W.; Liu, H.; and Wang, P. 2022. A perception-aware decomposition and fusion framework for underwater image enhancement. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(3): 988–1002.

Kim, G.; Park, S. W.; and Kwon, J. 2021. Pixel-wise wasserstein autoencoder for highly generative dehazing. *IEEE Transactions on Image Processing*, 30: 5452–5462.

Kim, Y. H.; and Park, K. R. 2022. PSS-net: Parallel semantic segmentation network for detecting marine animals in underwater scene. *Frontiers in Marine Science*, 9: 1003568.

Laradji, I. H.; Saleh, A.; Rodriguez, P.; Nowrouzezahrai, D.; Azghadi, M. R.; and Vazquez, D. 2021. Weakly supervised underwater fish segmentation using affinity LCFCN. *Scientific reports*, 11(1): 17379.

Li, C.; Guo, C.; Ren, W.; Cong, R.; Hou, J.; Kwong, S.; and Tao, D. 2019. An underwater image enhancement benchmark dataset and beyond. *IEEE Transactions on Image Processing*, 29: 4376–4389.

Li, K.; Wu, L.; Qi, Q.; Liu, W.; Gao, X.; Zhou, L.; and Song, D. 2023. Beyond single reference for training: underwater image enhancement via comparative learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 33: 2561–2576.

Li, L.; Dong, B.; Rigall, E.; Zhou, T.; Dong, J.; and Chen, G. 2021. Marine animal segmentation. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(4): 2303–2314.

Liu, P.; Wang, G.; Qi, H.; Zhang, C.; Zheng, H.; and Yu, Z. 2019. Underwater image enhancement with a deep residual framework. *IEEE Access*, 7: 94614–94629.

Liu, Q.; Zhang, Q.; Liu, W.; Chen, W.; Liu, X.; and Wang, X. 2023. WSDS-GAN: A Weak-Strong Dual Supervised Learning Method for Underwater Image Enhancement. *Pattern Recognition*, 143: 109774.

Liu, S.; Fan, H.; Lin, S.; Wang, Q.; Ding, N.; and Tang, Y. 2022. Adaptive learning attention network for underwater image enhancement. *IEEE Robotics and Automation Letters*, 7(2): 5326–5333.

Mi, L.; He, T.; Park, C. F.; Wang, H.; Wang, Y.; and Shavit, N. 2021. Revisiting Latent-Space Interpolation via a Quantitative Evaluation Framework. *arXiv preprint arXiv:2110.06421*.

Mu, P.; Qian, H.; and Bai, C. 2022. Structure-Inferred Bilevel Model for Underwater Image Enhancement. In *Proceedings of the 30th ACM International Conference on Multimedia*, 2286–2295.

Naik, A.; Swarnakar, A.; and Mittal, K. 2021. Shallow-uwnet: Compressed model for underwater image enhancement. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 15853–15854.

Nezla, N.; Haridas, T. M.; and Supriya, M. 2021. Semantic segmentation of underwater images using unet architecture based deep convolutional encoder decoder model. In *International Conference on Advanced Computing and Communication Systems*, volume 1, 28–33.

Panetta, K.; Gao, C.; and Agaian, S. 2015. Human-visual-system-inspired underwater image quality measures. *IEEE Journal of Oceanic Engineering*, 41(3): 541–551.

Patil, P. W.; Thawakar, O.; Dudhane, A.; and Murala, S. 2019. Motion saliency based generative adversarial network for underwater moving object segmentation. In *IEEE International Conference on Image Processing*, 1565–1569.

Qi, Q.; Li, K.; Zheng, H.; Gao, X.; Hou, G.; and Sun, K. 2022. SGUIE-Net: Semantic attention guided underwater image enhancement with multi-scale perception. *IEEE Transactions on Image Processing*, 31: 6816–6830.

Qiao, N.; Dong, L.; and Sun, C. 2022. Adaptive Deep Learning Network With Multi-Scale and Multi-Dimensional Features for Underwater Image Enhancement. *IEEE Transactions on Broadcasting*, 69: 482–494.

Raveendran, S.; Patil, M. D.; and Birajdar, G. K. 2021. Underwater image enhancement: a comprehensive review, recent trends, challenges and applications. *Artificial Intelligence Review*, 54: 5413–5467.

Sharma, P.; Bisht, I.; and Sur, A. 2023. Wavelength-based attributed deep neural network for underwater image restoration. *ACM Transactions on Multimedia Computing, Communications and Applications*, 19(1): 1–23.

Shen, H.; Zhao, Z.-Q.; and Zhang, W. 2023. Adaptive dynamic filtering network for image denoising. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2227–2235.

Shen, H.; Zhao, Z.-Q.; Zhang, Y.; and Zhang, Z. 2023. Mutual Information-driven Triple Interaction Network for Efficient Image Dehazing. In *Proceedings of the ACM International Conference on Multimedia*, 7–16.

Song, W.; Wang, Y.; Huang, D.; Liotta, A.; and Perra, C. 2020. Enhancement of underwater images with statistical model of background light and optimization of transmission map. *IEEE Transactions on Broadcasting*, 66(1): 153–169.

Sun, S.; Wang, H.; Zhang, H.; Li, M.; Xiang, M.; Luo, C.; and Ren, P. 2022. Underwater image enhancement with reinforcement learning. *IEEE Journal of Oceanic Engineering*.

Tang, Y.; Iwaguchi, T.; Kawasaki, H.; Sagawa, R.; and Furukawa, R. 2022. AutoEnhancer: Transformer on U-Net Architecture Search for Underwater Image Enhancement. In *Proceedings of the Asian Conference on Computer Vision*, 1403–1420.

Wang, K.; Shen, L.; Lin, Y.; Li, M.; and Zhao, Q. 2021. Joint iterative color correction and dehazing for underwater image enhancement. *IEEE Robotics and Automation Letters*, 6(3): 5121–5128.

Wang, N.; Chen, T.; Liu, S.; Wang, R.; Karimi, H. R.; and Lin, Y. 2023. Deep learning-based visual detection of marine organisms: A survey. *Neurocomputing*, 532: 1–32.

Wei, Y.; Zheng, Z.; and Jia, X. 2022. UHD Underwater Image Enhancement via Frequency-Spatial Domain Aware Network. In *Proceedings of the Asian Conference on Computer Vision*, 299–314.

Xu, S.; Zhang, M.; Song, W.; Mei, H.; He, Q.; and Liotta, A. 2023. A systematic review and analysis of deep learning-based underwater object detection. *Neurocomputing*, 527: 204–232.

Xue, X.; Li, Z.; Ma, L.; Jia, Q.; Liu, R.; and Fan, X. 2023. Investigating intrinsic degradation factors by multi-branch aggregation for real-world underwater image enhancement. *Pattern Recognition*, 133: 109041.

Yan, X.; Qin, W.; Wang, Y.; Wang, G.; and Fu, X. 2022. Attention-guided dynamic multi-branch neural network for underwater image enhancement. *Knowledge-Based Systems*, 258: 110041.

Yang, M.; and Sowmya, A. 2015. An underwater color image quality evaluation metric. *IEEE Transactions on Image Processing*, 24(12): 6062–6071.

Ye, T.; Chen, S.; Liu, Y.; Ye, Y.; Chen, E.; and Li, Y. 2022. Underwater light field retention: Neural rendering for underwater imaging. In *IEEE Conference on Computer Vision and Pattern Recognition*, 488–497.

Zhang, W.; Wu, C.; and Bao, Z. 2022. DPANet: Dual Pooling-aggregated Attention Network for fish segmentation. *IET computer vision*, 16(1): 67–82.

Zhang, W.; Zhuang, P.; Sun, H.-H.; Li, G.; Kwong, S.; and Li, C. 2022. Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement. *IEEE Transactions on Image Processing*, 31: 3997–4010.

Zhao, H.; Gallo, O.; and Frosio, I. 2016. Loss functions for image restoration with neural networks. *IEEE Transactions on computational imaging*, 3(1): 47–57.

Zheng, Y.; Chen, W.; Lin, R.; Zhao, T.; and Le Callet, P. 2022. UIF: An objective quality assessment for underwater image enhancement. *IEEE Transactions on Image Processing*, 31: 5456–5468.

Zhou, J.; Sun, J.; Zhang, W.; and Lin, Z. 2023. Multi-view underwater image enhancement method via embedded fusion mechanism. *Engineering Applications of Artificial Intelligence*, 121: 105946.

Zhuang, P.; Wu, J.; Porikli, F.; and Li, C. 2022. Underwater image enhancement with hyper-laplacian reflectance priors. *IEEE Transactions on Image Processing*, 31: 5442–5455.