

# Toward Individual Tone Preference in Underwater Image Enhancement

Xu Liu<sup>1</sup>, Student Member, IEEE, Yang Zhao<sup>1</sup>, Member, IEEE, Kaichen Chi<sup>1</sup>,  
Zhao Zhang<sup>1</sup>, Senior Member, IEEE, Yanxiang Chen<sup>1</sup>, and Wei Jia<sup>1</sup>, Member, IEEE

**Abstract**—Underwater images often suffer from severe color distortion due to the challenging imaging environment. Underwater image enhancement (UIE) techniques have been developed to recover clear images, laying the foundation for various underwater research. However, existing UIE methods tend to produce fixed results without considering individual preferences for different color tones. And there is no dataset with ground truth (GT) in different tones. Therefore, we came up with the possibility of using the currently popular multimodal methods to control the color tone of enhanced images. This article proposes a method for generating underwater enhanced images with cold, warm, and normal tones using multimodal information supervision (MM-UIE). First, we leverage the relationship between text prompts and images to supervise the generation of cold or warm images. In addition, we introduce a 6-D color operator, which not only enhances the tone control of underwater images but also serves as a bridge between different tone images. Finally, we also found that multimodal supervision methods can not only control the color tone of underwater images but also improve the quality of underwater image generation. Experimental results demonstrate the superior performance of our method compared to state-of-the-art (SOTA) techniques. Our codes will be publicly available at <https://github.com/perseverancelX/MM-UIE>.

**Index Terms**—6-D color operator, application of the large-scale generative model, multimodal learning, underwater image enhancement (UIE).

## I. INTRODUCTION

UNDERWATER images play an essential role in presenting oceanic information, offering a unique perspective to observe the wonders of the underwater world [1], [2], [3]. These images can provide information on organisms, terrains, and environmental conditions in the ocean, contributing to

Received 28 May 2024; revised 19 September 2024; accepted 18 October 2024. Date of publication 23 October 2024; date of current version 12 November 2024. This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant JZ2022HGPA0309; in part by the National Natural Science Foundation of China under Grant 62272142, Grant 62076086, Grant 62472137, Grant 62476077, and Grant 61972127; and in part by Anhui Provincial Natural Science Fund for the Distinguished Young Scholars under Grant 2008085J30. (Corresponding authors: Yang Zhao; Yanxiang Chen.)

Xu Liu, Zhao Zhang, Yanxiang Chen, and Wei Jia are with the School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230601, China (e-mail: dalong.xu.liu@ieee.org; cszhang@gmail.com; chenyx@hfut.edu.cn; jiawei@hfut.edu.cn).

Yang Zhao is with the School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230601, China, and also with the Peng Cheng National Laboratory, Shenzhen 518000, China (e-mail: yzhao@hfut.edu.cn).

Kaichen Chi is with the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an 710072, China (e-mail: chikaichen@mail.nwpu.edu.cn).

Digital Object Identifier 10.1109/TGRS.2024.3485030

1558-0644 © 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.  
See <https://www.ieee.org/publications/rights/index.html> for more information.

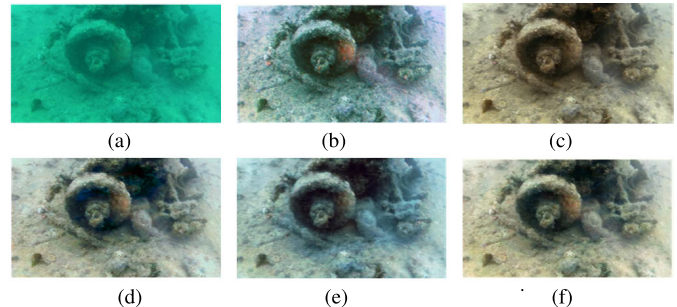


Fig. 1. Visual comparison of (a) degraded input, (b) Water-Net, (c) Ucolor, (d) ours-normal, (e) ours-cold, and (f) ours-warm in the UIEB dataset.

the protection and management of marine resources [4], [5], [6], [7], [8]. In addition, underwater images can also be used in fields such as scientific research, military reconnaissance, and entertainment, bringing enormous value and significance to humanity. However, underwater images often suffer from degradation issues such as color distortion, low contrast, and blurry details due to the complex and challenging imaging environments.

Underwater image enhancement (UIE) and restoration aim to improve the visual quality of degraded underwater images. The current algorithms generate different styles of enhancement results as shown in Fig. 1: (b) Water-Net [9]; (c) Ucolor [10]. For example, the results of Water-Net tend to be cooler, while those of Ucolor tend to be warmer. People who prefer cooler colors are more likely to prefer the results provided by Water-Net. On the other hand, people who are drawn to warmer colors tend to favor the results that Ucolor produces. However, current UIE algorithms [9], [10], [11] often overlook the subjective preferences and emotional responses of people toward the enhanced images.

Indeed, people have different color preferences due to personality variations, cultural backgrounds, and emotional experiences, which influence their feelings toward colors. For example, some people may prefer bright and vibrant colors like red, orange, and yellow as they evoke a sense of pleasure and excitement. Others may lean toward soft and elegant colors such as blue, green, and purple, which lead to a feeling of calmness and relaxation. With the advancements in large models and multimodal algorithms, learning-driven image-enhancement algorithms can also leverage useful information from other modalities. Therefore, we propose a UIE algorithm based on multimodal information supervision, which breaks the dilemma of requiring complete paired dataset supervision [as there is currently no dataset with different tone ground truth (GT)].

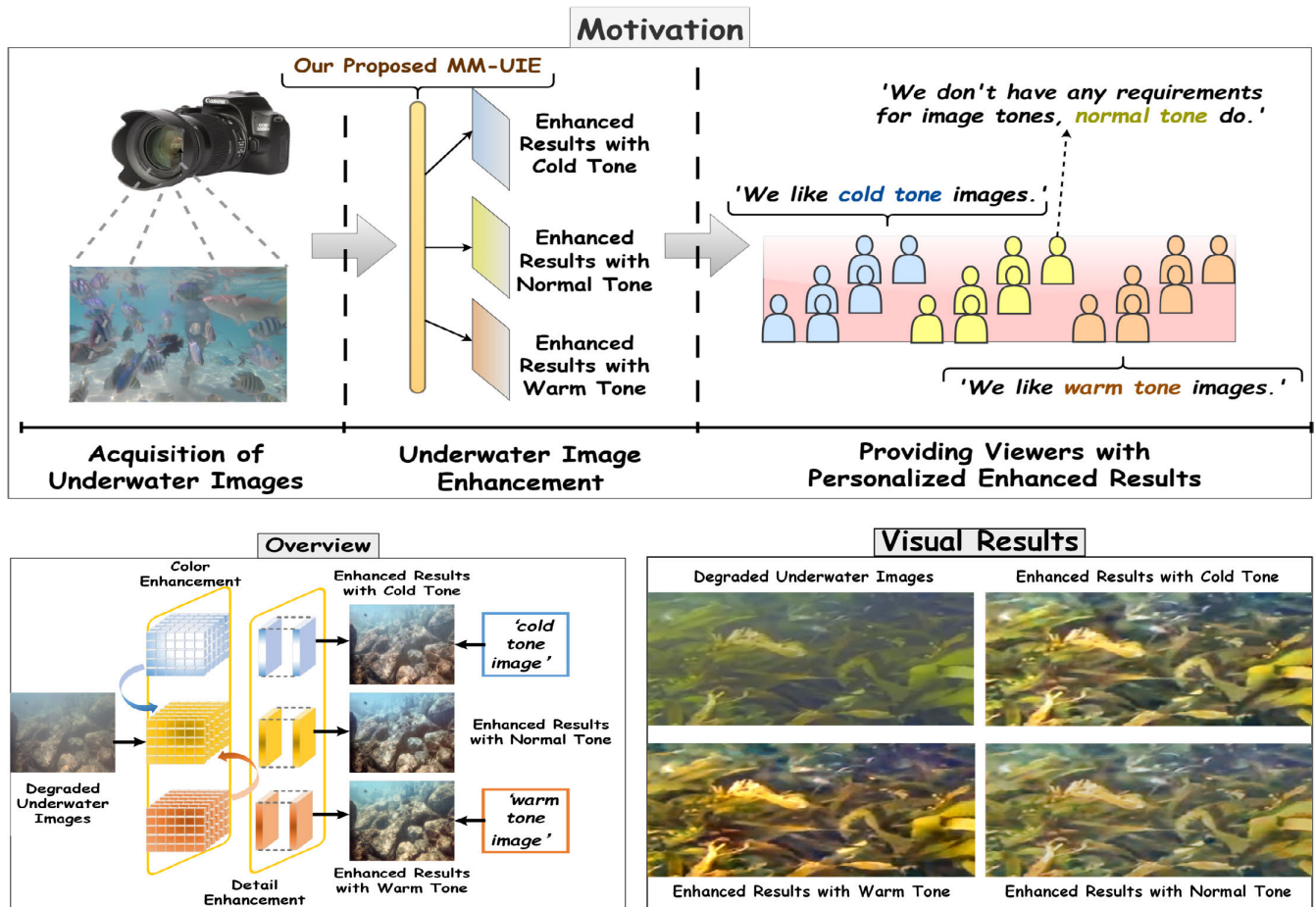


Fig. 2. Demonstration of the motivation, overview, and visual results of our proposed MM-UIE. Humans possess a unique ability to perceive color. We can observe in our daily lives that some people prefer cool tones, while others prefer warm tones, and some are unconcerned with tones at all. Thus, as a task of UIE, is it possible to generate enhanced images with different tones based on individual preferences?.

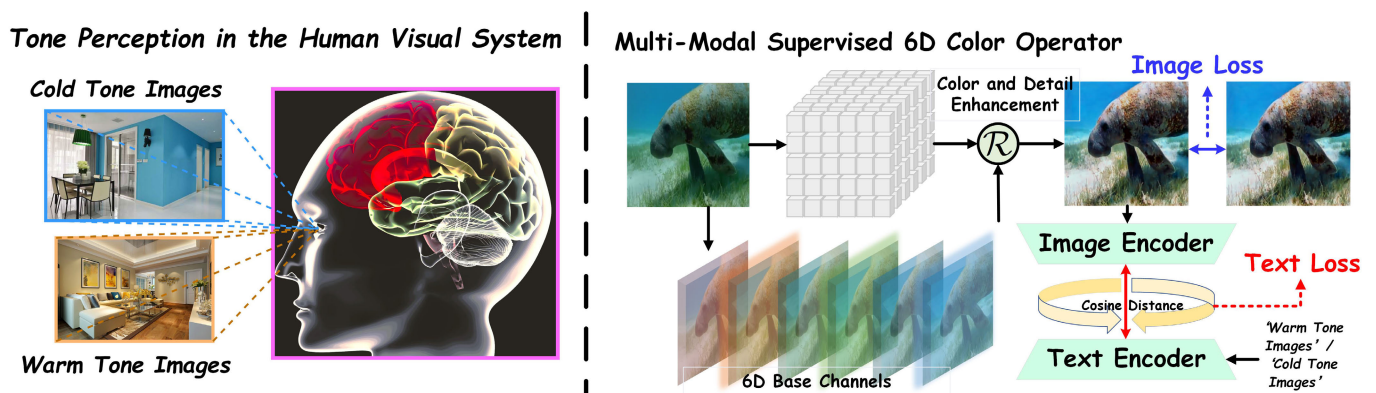


Fig. 3. Tone perception in the human visual system versus multimodal supervised 6-D color operator.

As shown in Fig. 2, our proposed method consists of two parts: color enhancement, which corrects the color of degraded underwater images, and detail enhancement, which restores the texture details of degraded images. Both text and image information are utilized to supervise the generation of enhanced underwater images with cool and warm tones. In addition, It also illustrates the results of different tones generated by the proposed method. The contributions of this article are summarized as follows.

1) We propose a method called the multimodal supervised 6-D color operator for UIE (MM-UIE) in Fig. 3, which is

the first method to achieve personalized UIE. By incorporating multimodal information supervision and a 6-D color operator, our method is pleasing to the human visual system and generates enhanced images with three different tones according to the preferences of different populations.

2) To control the color tone of the generated image, this article proposes a novel cross-modal interaction method. To the best of our knowledge, this method is also the first multimodal approach in the UIE field. We design a loss function for supervised learning by

calculating the cosine distance between two modal features.

- 3) Color cast is a significant problem in underwater images. To address this issue and enhance the hue of the generated images, we propose a 6-D color operator based on bilateral learning. This operator enables full-size color adjustment on the input image through six channel dimensions.
- 4) Comprehensive subjective and objective experiments demonstrate that the proposed method effectively enhances underwater images, obtains better colors and fewer hazing artifacts, and exhibits clear advantages over other state-of-the-art (SOTA) algorithms.

## II. RELATED WORK

This section will introduce works related to the algorithm proposed in this article. First, we will introduce the application scenario of UIE and restoration and demonstrate some representative methods. Then, we will introduce the core of this article multimodal information supervision and summarize how current enhancement algorithms apply multimodal means.

### A. Underwater Image Enhancement

Existing UIE algorithms can be roughly divided into three types, that is, prior-based, physical-model-based, and learning-based methods.

Prior-based and physical models-based methods are traditional approaches. Prior-based methods tend to directly adjust the pixel values of the input image. For example, Ancuti et al. [12] designed a novel white balance and fusion model, incorporating red channel priors and grayscale world assumptions for white balance. Zhang et al. [13] performed color correction based on histogram priors and combined it with contrast enhancement to emphasize details. Physical-model-based methods are often performed by solving imaging models. Song et al. [14] recovered underwater images based on underwater optical imaging models and proposed a dark channel prior model suitable for underwater scenes. Zhuang et al. [15] used hyper-Laplacian reflection priors for the retinex variational model, combining priors and physical models. To enhance underwater polarization images, Shen et al. [16] presented a polarization-driven method, which improves the contrast of underwater images, and they created a comprehensive benchmark for underwater polarization images. Although these traditional methods work well in some specific degraded images, for a large number of uncertainly degraded underwater images, these traditional algorithms are less robust, the enhancement effect is unstable, and some images still have color casts, artifacts, and so on.

In recent years, learning-based image enhancement algorithms have become mainstream [17], [18], [19], [20], [21]. Learning-based UIE methods can be mainly divided into two types: convolutional neural networks (CNNs)-based [9], [10], [22] and generative adversarial networks (GANs)-based [11], [23], [24], [25] methods. Learning-based methods learn the mapping from degraded images to clear images end-to-end, without the need for prior or physical models. UWCNN [26] is the first data-driven model, but its results exhibit severe color cast due to complete training based on synthetic datasets. Li et al. [9] fused the three enhanced results of traditional algorithms (i.e., white balance, gamma correction, and histogram equalization) through deep neural networks, resulting

in fine visual quality. They also proposed Ucolor [10], which utilizes different color spaces for collaborative enhancement. To generate visually pleasing results, FUnIE-GAN [23] is based on conditional GAN and considers global similarity and content consistency through loss functions. As for polarization images, the U2PNet [27] proposes an unsupervised restoration method that analyzes the relationship between the transmission map and the degree of polarization. By incorporating intensity ratio constraints, the loss function preserves details and colors in the image. However, most current models rely on GT for learning, and their enhanced images neglect people's preferences, such as some people liking cold-tone images, while others like warm-tone images. Furthermore, most algorithms utilize a deep unrolling-based architecture that results in a greater number of parameters and a slower testing speed caused by multiple iterations.

### B. Application of Multimodal Information

With the advent of large models, cross-modal learning has gained widespread attention, particularly between vision and text. The CLIP model [28] is a multimodal model based on contrastive learning, which can learn the matching relationship between text and image pairs. Recently, many works have applied CLIP models for supervised learning of images and videos. For example, Yang et al. [29] used text prompts as priors to enhance low-light images, resulting in visually pleasing results. Ju et al. [30] unified multiple tasks by adding a set of learnable vectors to the image-based visual multimodal model. Liang et al. [31] proposed a backlight image enhancement training scheme combined with the CLIP model: Image and text encoders using pretrained CLIPs encode backlit and well-illuminated images, as well as learnable cue pairs (positive and negative samples) into the latent space.

Tone perception is an advanced function of the human visual system. To achieve enhanced results with different tones, we have come up with the idea of using text information to supervise learning through multimodal models. Moreover, due to the U-Net baseline for generating the 6-D color operators, our network is lightweight when compared with deep rolling-based models.

## III. METHODOLOGY

The proposed MM-UIE consists of two stages, that is, color enhancement and detail enhancement. In this section, we will introduce these two stages and the proposed loss function for cross-modality interactive constraint.

### A. Multimodal Information for Tone Control

UIE is a challenging task, and most of the enhancement results are unidirectional. It is important to determine whether our enhanced images meet the preferences of our users. Based on this motivation, we have designed a novel UIE algorithm that can generate different tones. Enhanced images with warm and normal tones to meet user preferences.

The current UIE dataset lacks GT with multiple tones. Therefore, we utilize a pretrained CLIP [28] model for cross-modal interaction, which can establish a connection between texts and images in the feature domain.

Given different tone enhanced results as  $\mathcal{Y}_{\text{col}}$  and  $\mathcal{Y}_{\text{war}}$ , we use the image encoder  $\alpha_{\text{img}}$  to generate image features

$$\mathcal{W}_{\text{col}}^{\text{img}} = \alpha_{\text{img}}(\mathcal{Y}_{\text{col}})\mathcal{W}_{\text{war}}^{\text{img}} = \alpha_{\text{img}}(\mathcal{Y}_{\text{war}}). \quad (1)$$



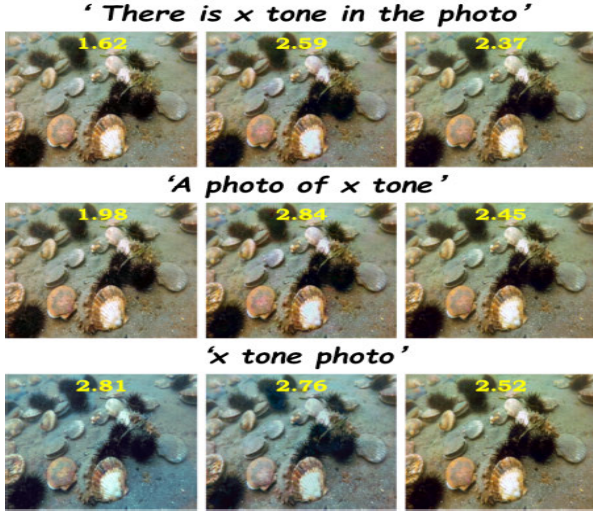


Fig. 4. Visual results comparison in different prompts. Scores range from 1 to 3, and the higher the score, the more appealing it will be to the viewer. (Left to right)  $x$  are “cold,” “normal,” and “warm,” respectively.

And text encoder  $\alpha_{\text{img}}$  for generate text features:

$$\mathcal{W}_{\text{col}}^{\text{tex}} = \alpha_{\text{tex}}(\mathcal{Z}_{\text{col}}) \mathcal{W}_{\text{war}}^{\text{tex}} = \alpha_{\text{tex}}(\mathcal{Z}_{\text{war}}) \quad (2)$$

where we use the CLIP-RN50 baseline as the image and text encoder according to [29].

Hence, the loss function for tone control can be represented as follows:

$$\begin{aligned} \mathcal{L}_{\text{mm}}^{\text{col}}(\mathcal{W}_{\text{col}}^{\text{img}}, \mathcal{W}_{\text{col}}^{\text{tex}}) &= \frac{\langle \mathcal{W}_{\text{col}}^{\text{img}}, \mathcal{W}_{\text{col}}^{\text{tex}} \rangle}{\|\mathcal{W}_{\text{col}}^{\text{img}} \cdot \mathcal{W}_{\text{col}}^{\text{tex}}\|} \\ \mathcal{L}_{\text{mm}}^{\text{war}}(\mathcal{W}_{\text{war}}^{\text{img}}, \mathcal{W}_{\text{war}}^{\text{tex}}) &= \frac{\langle \mathcal{W}_{\text{war}}^{\text{img}}, \mathcal{W}_{\text{war}}^{\text{tex}} \rangle}{\|\mathcal{W}_{\text{war}}^{\text{img}} \cdot \mathcal{W}_{\text{war}}^{\text{tex}}\|} \end{aligned} \quad (3)$$

where  $\langle \cdot, \cdot \rangle$  represents the cosine distance between two features, and  $\|\cdot\|$  means the L1 norm.

Influenced by [32], we use different text prompts for text tensor generation in Fig. 4. To find the feelings of different groups of people, we searched for 60 volunteers, including 20 volunteers who prefer cold tones and 20 volunteers who prefer warm tones. Then, the remaining 20 volunteers did not have a clear preference for color tones. The average score of each image is annotated in Fig. 4. Obviously, when the prompt is “ $x$ -tone photo,” the scores for each color tone are relatively balanced. Therefore, we choose this form as the prompt. To enhance image quality, we also add the image loss. It includes  $\mathcal{L}_1$  and  $\mathcal{L}_{\text{VGG}}$  [10]

$$\begin{aligned} \mathcal{L}_{\text{img}}^1 &= \mathcal{L}_1(\mathcal{Y}_{\text{col}}, \mathcal{Y}_{\text{gt}}) + \mathcal{L}_1(\mathcal{Y}_{\text{nor}}, \mathcal{Y}_{\text{gt}}) \\ &\quad + \mathcal{L}_1(\mathcal{Y}_{\text{war}}, \mathcal{Y}_{\text{gt}}) \\ \mathcal{L}_{\text{img}}^{\text{VGG}} &= \mathcal{L}_{\text{VGG}}(\mathcal{Y}_{\text{col}}, \mathcal{Y}_{\text{gt}}) + \mathcal{L}_{\text{VGG}}(\mathcal{Y}_{\text{nor}}, \mathcal{Y}_{\text{gt}}) \\ &\quad + \mathcal{L}_{\text{VGG}}(\mathcal{Y}_{\text{war}}, \mathcal{Y}_{\text{gt}}). \end{aligned} \quad (4)$$

Hence, the total losses are as follows:

$$\begin{aligned} \mathcal{L}_{\text{tex}} &= \mathcal{L}_{\text{mm}}^{\text{col}}(\mathcal{W}_{\text{col}}^{\text{img}}, \mathcal{W}_{\text{col}}^{\text{tex}}) + \mathcal{L}_{\text{mm}}^{\text{war}}(\mathcal{W}_{\text{war}}^{\text{img}}, \mathcal{W}_{\text{war}}^{\text{tex}}) \\ \mathcal{L}_{\text{total}} &= \lambda_a \cdot \mathcal{L}_{\text{img}}^1 + \lambda_b \cdot \mathcal{L}_{\text{img}}^{\text{VGG}} + \lambda_c \cdot \mathcal{L}_{\text{tex}} \end{aligned} \quad (5)$$

where  $\lambda_a$ ,  $\lambda_b$ , and  $\lambda_c$  are three weights set to 3, 0.15, and 0.01, respectively.

## B. Color Enhancement

In this article, a 6-D color operator is designed to achieve color enhancement. Efficient bilateral learning model, which has been successfully applied for related tasks such as image dehazing [33]. However, due to the severe degradation of underwater images and the scattered distribution of degradation at different pixel points, traditional bilateral learning cannot handle the various and complex color distortions. Therefore, we propose a full-size 6-D color operator, which can capture the color adjustment operator of the entire underwater image.

Given a degraded underwater image  $\mathcal{X}$  composed of  $R$ ,  $G$ , and  $B$  three channels. To express the information of the original image more completely and enhance color effectively, we have modified the original RGB three channels as follows:

$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = \mathcal{S} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix}, \quad \mathcal{S} = \begin{bmatrix} M & 0 & 0 \\ 0 & N & 0 \\ 0 & 0 & P \end{bmatrix} \quad (6)$$

where  $R'$ ,  $G'$ ,  $B'$  denote transformed RGB channels, and  $\mathcal{S}$  represents a mapping diagonal matrix with three rows and three columns. Then, (6) can be rewritten as follows:

$$\begin{aligned} \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} &= \begin{bmatrix} M & 0 & 0 \\ 0 & N & 0 \\ 0 & 0 & P \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \\ &= \begin{bmatrix} M-1 & 0 & 0 \\ 0 & N-1 & 0 \\ 0 & 0 & P-1 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} R \\ G \\ B \end{bmatrix} \\ &= \mathcal{S}' \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} R \\ G \\ B \end{bmatrix} \end{aligned} \quad (7)$$

where  $\mathcal{S}'$  denotes generalized matrix from  $\mathcal{S}$ . The result is similar to  $y = f(x) + x$ , so we use a convolutional layer  $\mathcal{U}$  and a residual connection for simulating the color transformation process as

$$\mathcal{X}' = \mathcal{U}(\mathcal{X}) + \mathcal{X}, \quad \mathcal{X} \in \begin{bmatrix} R \\ G \\ B \end{bmatrix}, \quad \mathcal{X}' \in \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix}. \quad (8)$$

The original  $R$ ,  $G$ ,  $B$  and transformed  $R'$ ,  $G'$ ,  $B'$  are combined to six dimensions  $\mathcal{T}_c$  for color adjustment, as follows:

$$\mathcal{T}_c \in [R \ R' \ G \ G' \ B \ B']^T. \quad (9)$$

In Fig. 5, we first use three U-Net to generate the 6-D color operators for normal tone, cold tone, and warm tone, respectively,

$$\mathcal{C}_{\text{col}} = \mathcal{G}_{\text{col}}(\mathcal{X}), \quad \mathcal{C}_{\text{nor}}^0 = \mathcal{G}_{\text{nor}}(\mathcal{X}), \quad \mathcal{C}_{\text{war}} = \mathcal{G}_{\text{war}}(\mathcal{X}) \quad (10)$$

where  $\mathcal{G}_{\text{col}}$ ,  $\mathcal{G}_{\text{nor}}$ , and  $\mathcal{G}_{\text{war}}$  denote cold-tone U-Net, normal-tone U-Net, and warm-tone U-Net, respectively, and  $\mathcal{C}_{\text{col}}$ ,  $\mathcal{C}_{\text{nor}}^0$ , and  $\mathcal{C}_{\text{war}}$  are generated 6-D color operators. The detailed diagram of these three U-Net is shown in Fig. 6. Moreover, Table I indicates the memory usage consumed by different blocks.

Fig. 7 shows the generation process of different tone images. The details of the color and detail enhancement are displayed in Fig. 8. The normal-tone operator should well balance the warm and cold tones, we thus fuse three types of 6-D operators to further improve the normal-tone operator, as follows:

$$\mathcal{C}_{\text{nor}} = \mathcal{F}_{\text{ope}}(\mathcal{C}_{\text{col}}, \mathcal{C}_{\text{nor}}^0, \mathcal{C}_{\text{war}}) \quad (11)$$

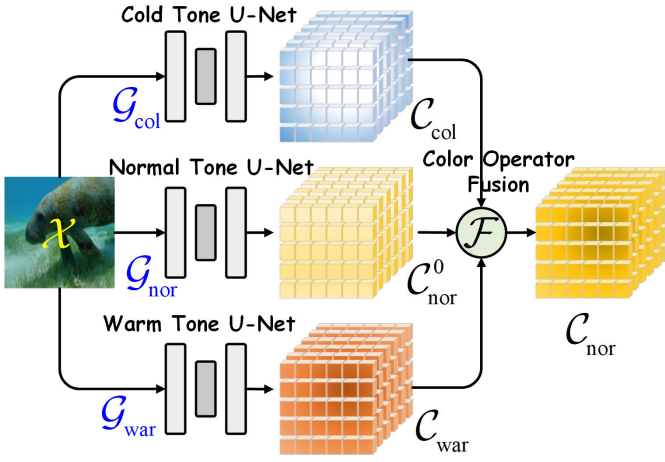


Fig. 5. Generation process of the 6-D color operator.

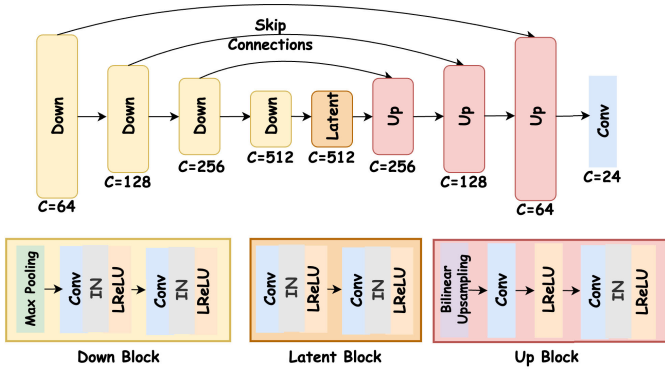
Fig. 6. Detailed diagram of the cold tone U-Net, normal tone U-Net, and warm tone U-Net, and  $C$  represents the number of channels. Conv denotes the convolutional layer. LReLU represents the leaky ReLU function, and IN is the instance normalization.

TABLE I

MEMORY USAGE COMPARISON OF DIFFERENT BLOCKS (THE IMAGE SIZE IS  $256 \times 256$ ). UP, LATENT, AND DOWN REPRESENT THE UP BLOCK, LATENT BLOCK, AND DOWN BLOCK, RESPECTIVELY

U-Net Model	Memory Usage	
	Training↓	Testing↓
3 Down + 1 Latent + 2 Up	2161M	1929M
4 Down + 1 Latent + 3 Up	3432M	3282M
5 Down + 1 Latent + 4 Up	4338M	4017M

which also confirms that the proposed 6-D color operator can serve as a bridge connecting three different tones.

The main function of the operator fusion network  $\mathcal{F}_{\text{ope}}$  is to fuse 6-D color operators with three different tones ( $C_{\text{col}}$ : the operator for the cold tone,  $C_{\text{nor}}^0$ : the first operator for the normal tone, and  $C_{\text{war}}$ : the operator for the warm tone). Hence, we design a lightweight network  $\mathcal{F}_{\text{ope}}$  to generate the final normal-tone operator  $C_{\text{nor}}$ . Through the operators for warm and cold tones and the first operator for the normal tone, the lightweight network  $\mathcal{F}_{\text{ope}}$  can combine their respective excellent features and refine them in Fig. 9.

We use the affine function  $\mathcal{A}$  to combine the operators and six channels for color enhancement

$$\begin{aligned} \mathcal{V}_{\text{col}} &= \mathcal{A} \cdot (C_{\text{col}}, T_c), & \mathcal{V}_{\text{nor}} &= \mathcal{A} \cdot (C_{\text{nor}}, T_c) \\ \mathcal{V}_{\text{war}} &= \mathcal{A} \cdot (C_{\text{war}}, T_c) \end{aligned} \quad (12)$$

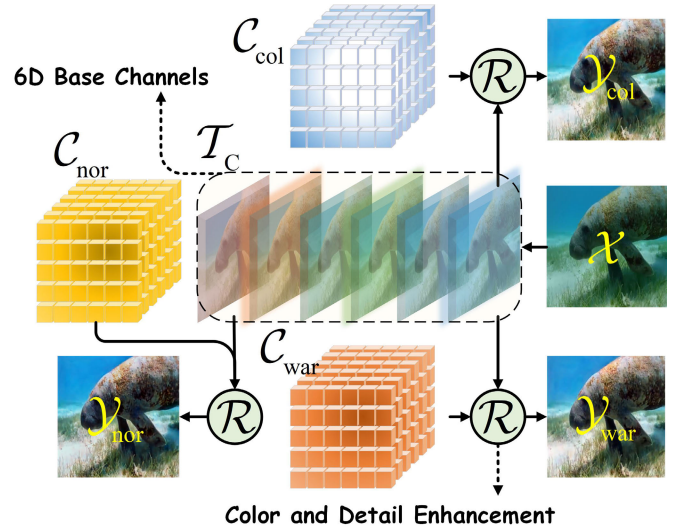


Fig. 7. Demonstration of different tone image generation.

TABLE II

EVALUATIONS OF DIFFERENT METHODS ON THE UIEB-100 TEST SET IN TERMS OF FULL-REFERENCE (PSNR AND SSIM) AND NO-REFERENCE (UCIQE AND UIQM) METRICS. RED, BLUE, AND UNDERSCORED BOLD FONTS INDICATE THE BEST THREE RESULTS

Method	Full-Reference		No-Reference	
	PSNR↑	SSIM↑	UCIQE↑	UIQM↑
GDCP	12.6871	0.5256	0.5889	<u><b>1.5023</b></u>
HLRP	13.4891	0.2741	0.5629	<b>1.5218</b>
NUDCP	15.8837	0.6915	0.5891	1.4419
ACDC	18.1247	0.7016	0.5635	<b>1.5391</b>
ERH	19.2439	0.6929	0.5529	1.4781
Fusion	20.3664	0.8180	0.5925	1.4971
UWCNN	14.0549	0.4947	0.4930	1.3961
URSCT	14.1659	0.4623	0.5789	1.4200
L2UWE	14.1710	0.6892	0.5518	1.4865
FUnIE-GAN	18.1488	0.6509	0.5531	1.4324
LCNet	18.8990	0.7154	0.5588	1.4316
CLUIE-Net	20.5229	0.8021	0.5864	1.4548
TOPAL	21.1428	0.8142	0.5710	1.4282
Water-Net	21.5394	0.8223	0.5904	1.4319
UICoE-Net	21.6395	0.8472	0.5880	1.4607
Ucolor	21.6671	0.8411	0.5635	1.4168
<b>Ours-Normal</b>	<u><b>24.9131</b></u>	<u><b>0.8886</b></u>	<u><b>0.6043</b></u>	1.4968
<b>Ours-Warm</b>	<b>25.5133</b>	<b>0.9065</b>	<b>0.6074</b>	1.4876
<b>Ours-Cold</b>	<b>25.7043</b>	<b>0.9069</b>	<b>0.6057</b>	1.4913

where  $\mathcal{V}_{\text{col}}$ ,  $\mathcal{V}_{\text{nor}}$ , and  $\mathcal{V}_{\text{war}}$  denote three features after color enhancement. Then, the details of these operations:  $\mathcal{A} \cdot (C_{\text{col}}, T_c)$ ;  $\mathcal{A} \cdot (C_{\text{war}}, T_c)$ ;  $\mathcal{A} \cdot (C_{\text{nor}}, T_c)$  are as follows:

$$\begin{aligned} \mathcal{A} \cdot (C_{\text{col}}, T_c) &= \text{Cat} \left( \sum_{\eta_1 \in \{1, 2, 3, 4, 5, 6\}} C_{\text{col}}^{\eta_1} T_c + C_{\text{col}}^7, \right. \\ &\quad \left. \sum_{\eta_2 \in \{8, 9, 10, 11, 12, 13\}} C_{\text{col}}^{\eta_2} T_c + C_{\text{col}}^{14}, \sum_{\eta_3 \in \{15, 16, 17, 18, 19, 20\}} C_{\text{col}}^{\eta_3} T_c + C_{\text{col}}^{21} \right) \\ \mathcal{A} \cdot (C_{\text{nor}}, T_c) &= \text{Cat} \left( \sum_{\eta_1 \in \{1, 2, 3, 4, 5, 6\}} C_{\text{nor}}^{\eta_1} T_c + C_{\text{nor}}^7, \right. \end{aligned} \quad (13)$$

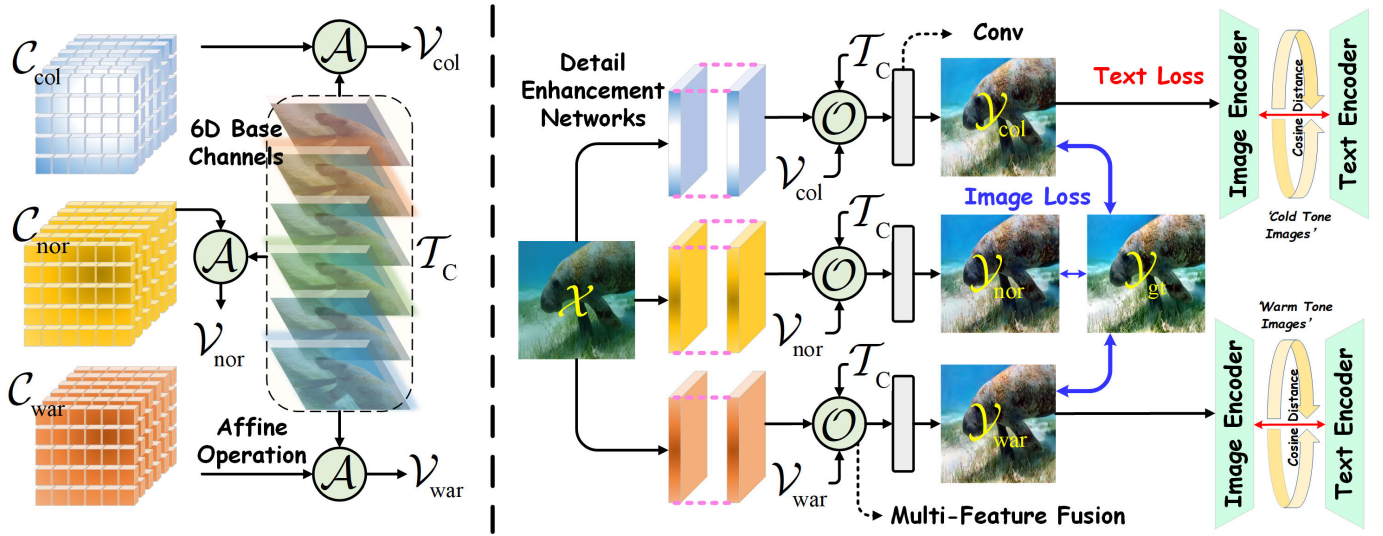


Fig. 8. Indication of (left) color and (right) detail enhancement.

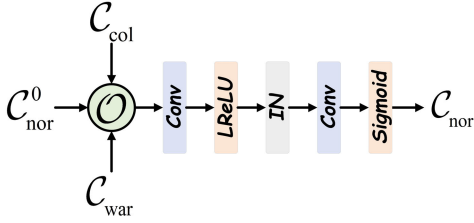


Fig. 9. Demonstration of the operator fusion network. Conv denotes the convolutional layer, LReLU and Sigmoid represent the leaky ReLU and sigmoid functions, respectively, IN is the instance normalization, and  $\mathcal{O}$  means the concatenation.

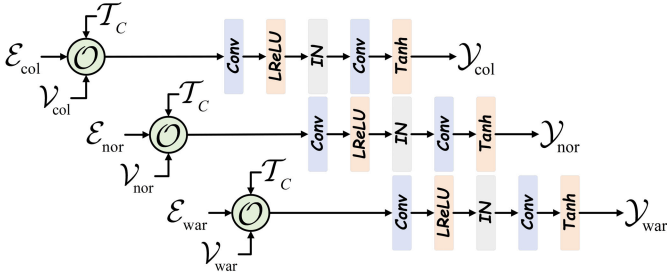


Fig. 10. Indication of three CNN-based multifeature fusion networks. Conv denotes the convolutional layer. LReLU, Sigmoid, and Tanh represent the leaky ReLU, sigmoid, and tanh functions, respectively, IN is the instance normalization, and  $\mathcal{O}$  means the concatenation.

$$\left( \sum_{\eta_2 \in \{8,9,10,11,12,13\}} C_{\text{nor}}^{\eta_2} \mathcal{T}_c + C_{\text{nor}}^{14}, \quad \sum_{\eta_3 \in \{15,16,17,18,19,20\}} C_{\text{nor}}^{\eta_3} \mathcal{T}_c + C_{\text{nor}}^{21} \right) \quad (14)$$

$$\mathcal{A} \cdot (C_{\text{war}}, \mathcal{T}_c) = \text{Cat} \left( \sum_{\eta_1 \in \{1,2,3,4,5,6\}} C_{\text{war}}^{\eta_1} \mathcal{T}_c + C_{\text{war}}^7, \quad \sum_{\eta_1 \in \{8,9,10,11,12,13\}} C_{\text{war}}^{\eta_2} \mathcal{T}_c + C_{\text{war}}^{14}, \quad \sum_{\eta_1 \in \{15,16,17,18,19,20\}} C_{\text{war}}^{\eta_3} \mathcal{T}_c + C_{\text{war}}^{21} \right) \quad (15)$$

where  $C_{\text{col/war/nor}} \in \mathbb{R}^{21 \times h \times w}$ ,  $\mathcal{T}_c \in \mathbb{R}^{6 \times h \times w}$ .

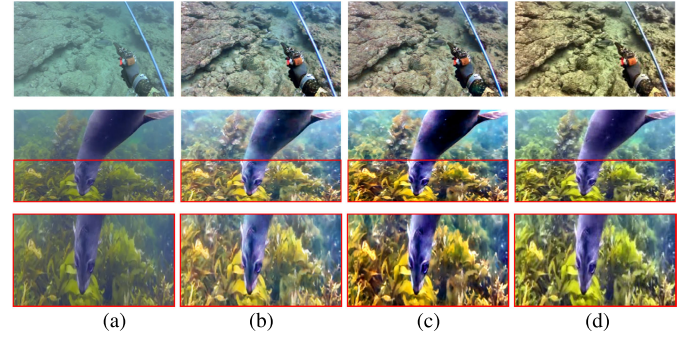


Fig. 11. Visual comparison of the cold-tone results, the warm-tone results, and the reference results in the UIEB dataset. (a) Input image. (b) Ours-cold. (c) Ours-warm. (d) Reference image.

### C. Detail Enhancement

The goal of detail enhancement is to restore information such as texture that has been degraded in the image, thereby improving the visual quality. We use the UIE model Boths [34] as the baseline to preliminarily enhance the details as

$$\mathcal{E}_{\text{col}} = \mathcal{D}_{\text{col}}(\mathcal{X}), \quad \mathcal{E}_{\text{nor}} = \mathcal{D}_{\text{nor}}(\mathcal{X}), \quad \mathcal{E}_{\text{war}} = \mathcal{D}_{\text{war}}(\mathcal{X}) \quad (16)$$

where  $\mathcal{E}_{\text{col}}$ ,  $\mathcal{E}_{\text{nor}}$ , and  $\mathcal{E}_{\text{war}}$  denote results of three same detail enhancement baseline modules  $\mathcal{D}_{\text{col}}$ ,  $\mathcal{D}_{\text{nor}}$ , and  $\mathcal{D}_{\text{war}}$ , respectively, as shown in Fig. 8. Then, we concatenate the color enhancement feature  $\mathcal{V}$ , the detail enhancement feature  $\mathcal{E}$ , and six channels  $\mathcal{T}_c$  through three multifeature fusion networks, as follows:

$$\begin{aligned} \mathcal{Y}_{\text{col}} &= \mathcal{F}_{\text{fea}}^{\text{col}}(\mathcal{O}(\mathcal{V}_{\text{col}}, \mathcal{E}_{\text{col}}, \mathcal{T}_c)) \\ \mathcal{Y}_{\text{nor}} &= \mathcal{F}_{\text{fea}}^{\text{nor}}(\mathcal{O}(\mathcal{V}_{\text{nor}}, \mathcal{E}_{\text{nor}}, \mathcal{T}_c)) \\ \mathcal{Y}_{\text{war}} &= \mathcal{F}_{\text{fea}}^{\text{war}}(\mathcal{O}(\mathcal{V}_{\text{war}}, \mathcal{E}_{\text{war}}, \mathcal{T}_c)) \end{aligned} \quad (17)$$

where  $\mathcal{O}$  denotes the concat operation, and  $\mathcal{Y}_{\text{col}}$ ,  $\mathcal{Y}_{\text{nor}}$ , and  $\mathcal{Y}_{\text{war}}$  are three different tone-enhanced results. We use three CNN-based multifeature fusion networks ( $\mathcal{F}_{\text{fea}}^{\text{col}}$ ,  $\mathcal{F}_{\text{fea}}^{\text{nor}}$ , and  $\mathcal{F}_{\text{fea}}^{\text{war}}$ ) to fuse the features and generate the final enhanced results of three different tones ( $\mathcal{Y}_{\text{col}}$ : the final results for the cold tone,  $\mathcal{Y}_{\text{nor}}$ : the final results for the normal tone, and  $\mathcal{Y}_{\text{war}}$ : the final results for the warm tone). In Fig. 10, these three CNN-based multifeature fusion networks can adaptively



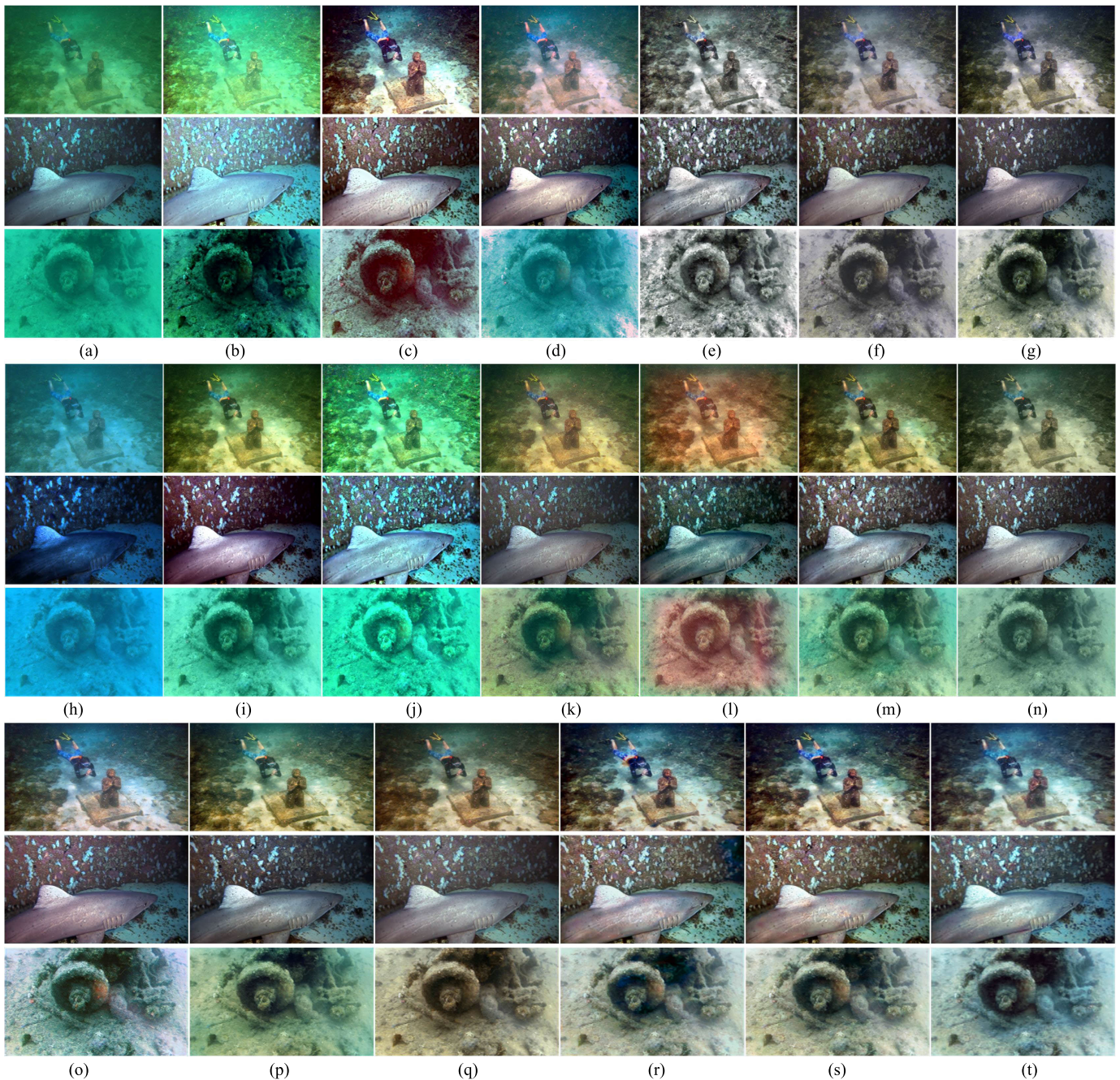


Fig. 12. Visual comparison of the UIEB-100 testset. (a) Input image. (b) GDCP [35]. (c) HLRP [15]. (d) NUDCP [14]. (e) ACDC [36]. (f) ERH [37]. (g) Fusion [12]. (h) UWCNN [26]. (i) URSCT [38]. (j) L2UWE [39]. (k) FUnIE-GAN [23]. (l) LCNet [40]. (m) CLUIE-Net [41]. (n) TOPAL [42]. (o) Water-Net [9]. (p) UICoE-Net [43]. (q) Ucolor [10]. (r) Ours-normal. (s) Ours-warm. (t) Ours-cold.

combine the detail results ( $\mathcal{E}_{\text{col}}$ ,  $\mathcal{E}_{\text{nor}}$ , and  $\mathcal{E}_{\text{war}}$ ), color correction results ( $\mathcal{V}_{\text{col}}$ ,  $\mathcal{V}_{\text{nor}}$ , and  $\mathcal{V}_{\text{war}}$ ), of three tones with 6-D base channels  $\mathcal{T}_c$ . Fig. 11 shows the final results with different tones in the UIEB dataset [9].

#### IV. EXPERIMENTS

*Implementation Details:* The proposed MM-UIE is trained with an AdamW optimizer for 180 epochs. The initial learning rate is set to 0.0001 and then halved after every 30 epochs. The batch size is set as 4. All training images have been cropped to  $224 \times 224$  patches and then normalized to range [0, 1].

The experiments are implemented with the PyTorch platform on two RTX3090 GPUs.

*Datasets:* The UIEB [9] and SQUID [44] datasets are two commonly used UIE benchmark sets. There are 950 real underwater images included in the UIEB dataset, which has been collected from the Internet. As GTs, the author manually selects images with better visual results following the processing of different algorithms. To obtain GT, different enhancement algorithms are implemented and then images with better visual quality are manually selected as GTs. As for the SQUID dataset, 57 pairs of stereo images are included in the database, two of which are located in the Red Sea



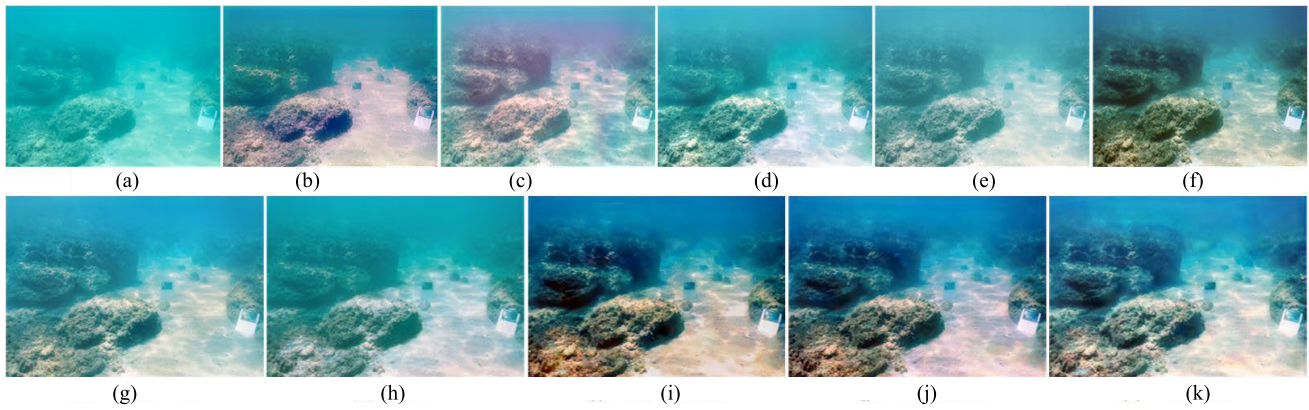


Fig. 13. Visual comparison of the SQUID-100 test set. (a) Input image. (b) FUnIE-GAN [23]. (c) LCNet [40]. (d) CLUIE-Net [41]. (e) TOPAL [42]. (f) Water-Net [9]. (g) UICoE-Net [43]. (h) Ucolor [10]. (i) Ours-normal. (j) Ours-warm. (k) Ours-cold.

TABLE III

EVALUATIONS OF DIFFERENT METHODS ON THE UIEB-890 TEST SET IN TERMS OF FULL-REFERENCE (MSE, RMSE, PSNR, SSIM, AND LPIPS) AND NO-REFERENCE (PI, MA, NIQE, UCIQE, AND UIQM) METRICS. RED, BLUE, AND UNDERSCORED BOLD FONTS INDICATE THE BEST THREE RESULTS

Method	Full-Reference					No-Reference				
	MSE↓	RMSE↓	PSNR↑	SSIM↑	LPIPS↓	PI↓	MA↑	NIQE↓	UCIQE↑	UIQM↑
FUnIE-GAN	242.9328	15.5863	23.7391	0.8751	0.2838	3.7993	7.6478	5.2463	0.5620	1.4063
LCNet	344.4753	18.5600	23.6221	0.9069	0.2207	3.4458	7.7556	4.6472	0.5738	1.4319
CLUIE-Net	389.2661	19.7298	24.4310	<b>0.9436</b>	0.1453	<b>3.1038</b>	<b>8.2809</b>	4.4885	0.5974	1.4355
TOPAL	387.3276	19.6806	24.2523	0.9064	0.1716	<b>3.1667</b>	<b>8.1648</b>	4.4983	0.5794	1.4036
Water-Net	548.4220	23.4184	22.5334	0.8581	0.3074	3.5524	7.7097	4.8144	0.6029	1.4164
UICoE-Net	294.1896	17.1520	24.1768	0.9188	0.1662	3.6122	7.6844	4.9088	0.5917	1.4263
Ucolor	225.4618	15.0154	25.8397	0.9174	0.1456	3.4851	7.5607	4.5310	0.5726	1.3558
<b>Ours-Normal</b>	<b>75.9174</b>	<b>8.7130</b>	<b>29.9073</b>	0.9169	<b>0.1310</b>	3.2671	7.9354	<b>4.4697</b>	<b>0.6109</b>	<b>1.4556</b>
<b>Ours-Warm</b>	<b>64.9680</b>	<b>8.0602</b>	<b>30.6512</b>	<b>0.9197</b>	<b>0.1215</b>	3.2045	<b>7.9646</b>	<b>4.3736</b>	<b>0.6114</b>	<b>1.4499</b>
<b>Ours-Cold</b>	<b>66.4712</b>	<b>8.1529</b>	<b>30.5979</b>	<b>0.9202</b>	<b>0.1218</b>	<b>3.2016</b>	7.9598	<b>4.3630</b>	<b>0.6117</b>	<b>1.4448</b>

(representing tropical waters) and two in the Mediterranean Sea (representing temperate waters). This dataset does not contain GTs.

We first perform data augmentation on the UIEB dataset by using horizontal and vertical flipping and rotation at angles of  $A \in [0, \pi/2, \pi, 3\pi/2]$ . A total of 12 680 training samples are obtained after augmentation. We select 990 images for testing, and the remaining samples are used for training. Among the 990 test images, 100 images with severe degradation are specifically selected to conduct a difficult UIEB-100 test set, and the other 890 images make up a UIEB-890 test set. In addition, 100 images are extracted from the SQUID dataset for the cross-dataset testing, namely the SQUID-100 test set.

*Metrics:* For full-reference evaluation, we use five commonly used metrics, that is, MSE, RMSE, PSNR, SSIM, and LPIPS [45]. These full reference indicators evaluate the distance between the image and the GT, which can measure the distortions. For no-reference indicators, we used another five assessments, that is, PI [46], MA [47], NIQE [48], UCIQE [49], and UIQM [50]. Compared with distortion measurements, these indicators focus on the visual quality, contrast, and color level of images. Among them, UCIQE and UIQM are two unique indicators for underwater images, which can comprehensively evaluate enhanced underwater images.

*Comparison Methods:* To verify the effectiveness of the proposed method, 16 UIE methods are selected for comparisons, including six traditional algorithms of GDCP [35], HLRP [15],

NUDCP [14], ACDC [36], ERH [37], and Fusion [12], and ten SOTA learning-based models of UWCNN [26], URSCT [38], L2UWE [39], FUnIE-GAN [23], LCNet [40], CLUIE-Net [41], TOPAL [42], Water-Net [9], UICoE-Net [43], and Ucolor [10].

#### A. Qualitative Evaluation

The subjective results of this method on the UIEB-100 and SQUID-100 test sets are illustrated in Figs. 12 and 13, respectively. From these figures, we can obtain the following observations. First, traditional algorithms can improve color degradation, resulting in enhanced results without obvious greenish or bluish colors. However, these processed images often suffer from monotonous colors and severe noise in the details. The GDCP and NUDCP algorithms have made some improvements in brightness, but they tend to produce uneven fogging phenomena and color distortion. The ACDC algorithm produces grayish-white images with poor visual effects. The HLRP algorithm tends to overexpose the images, while the ERH and Fusion algorithms produce darker results. In terms of learning-based methods, most of them exhibit better color and well-processed details compared to traditional methods. However, their results still contain some limitations. URSCT and L2UWE struggle to handle severe color cast. FUnIE-GAN and LCNet tend to introduce a red color cast to the enhanced image, with LCNet exhibiting severe red artifacts. CLUIE-Net, TOPAL, and UICoE-Net still exhibit some color degradation in certain results. In contrast, our proposed MM-UIE not only



TABLE IV

EVALUATIONS OF DIFFERENT METHODS ON THE SQUID-100 TEST SET IN TERMS OF NO-REFERENCE (UCIQE AND UIQM) METRICS. RED, BLUE, AND UNDERScoreD BOLD FONTS INDICATE THE BEST THREE RESULTS

Method	No-Reference	
	UCIQE $\uparrow$	UIQM $\uparrow$
FUnIE-GAN	0.5359	<b>1.0584</b>
LCNet	0.5254	1.0034
CLUIE-Net	0.5099	0.9689
TOPAL	0.5131	0.9346
Water-Net	0.5621	0.9747
UICoE-Net	0.5267	1.0014
Ucolor	0.5035	0.8974
<b>Ours-Normal</b>	<b>0.5845</b>	<b>1.1315</b>
<b>Ours-Warm</b>	<b>0.5654</b>	<b>1.0762</b>
<b>Ours-Cold</b>	<b>0.5672</b>	1.0574

TABLE V

ABLATION STUDY ON THE UIEB-100 TEST SET IN TERMS OF FULL-REFERENCE (PSNR AND SSIM) AND NO-REFERENCE (UCIQE AND UIQM) METRICS. BOLD FONTS INDICATE THE BEST RESULTS. THREE PIECES OF DATA CORRESPOND TO NORMAL-, WARM-, AND COLD-TONE RESULTS

Method	Full-Reference		No-Reference	
	PSNR $\uparrow$	SSIM $\uparrow$	UCIQE $\uparrow$	UIQM $\uparrow$
3D Col. Ope.	23.5425	0.8511	0.5998	1.4872
w/o 6D Col. Ope.	21.8979	0.7989	0.5902	1.4734
w/o Warm Sup.	20.7406	0.7570	0.5759	1.4483
w/o Cold Sup.	20.9705	0.7795	0.5763	1.4443
<b>Ours-Normal</b>	<b>24.9131</b>	<b>0.8886</b>	<b>0.6043</b>	<b>1.4968</b>
3D Col. Ope.	24.2124	0.8680	0.6045	1.4788
w/o 6D Col. Ope.	22.8625	0.8275	0.6041	1.4767
w/o Warm Sup.	22.2889	0.8115	0.6005	1.4596
w/o Cold Sup.	21.6017	0.7898	0.5974	1.4768
<b>Ours-Warm</b>	<b>25.5133</b>	<b>0.9065</b>	<b>0.6074</b>	<b>1.4876</b>
3D Col. Ope.	23.5425	0.8511	0.5998	1.4872
w/o 6D Col. Ope.	21.8979	0.7989	0.5902	1.4734
w/o Warm Sup.	20.7406	0.7570	0.5759	1.4483
w/o Cold Sup.	20.9705	0.7795	0.5763	1.4443
<b>Ours-Cold</b>	<b>25.7043</b>	<b>0.9069</b>	<b>0.6057</b>	<b>1.4913</b>

achieves good results on the UIEB-100 and SQUID-100 test sets but also provides users with enhanced results in three different tones.

### B. Quantitative Evaluation

In this section, we will compare our MM-UIE and SOTA methods quantitatively. Table II shows that we evaluate UIEB-100 using four metrics: PSNR, SSIM, UCIQE, and UIQM. Accordingly, the method proposed in this article has advantages in terms of PSNR, SSIM, and UCIQE metrics. In addition, Table III shows some excellent algorithms tested on UIEB-890. This test is more comprehensive and specific, showing clearly the MM-UIE characteristics. In Table IV, eight of the ten metrics of MM-UIE exceed other algorithms. According to the SQUID-100 test set, our proposed method still has significant advantages over excellent algorithms in terms of both UCIQE and UIQM metrics. Overall, the MM-UIE proposed in this article performs well in nearly ten metrics across three datasets, with improved color performance and enhanced image quality.

### C. Ablation Study

An ablation experiment was designed to demonstrate the efficacy of the multimodal information supervision and 6-D

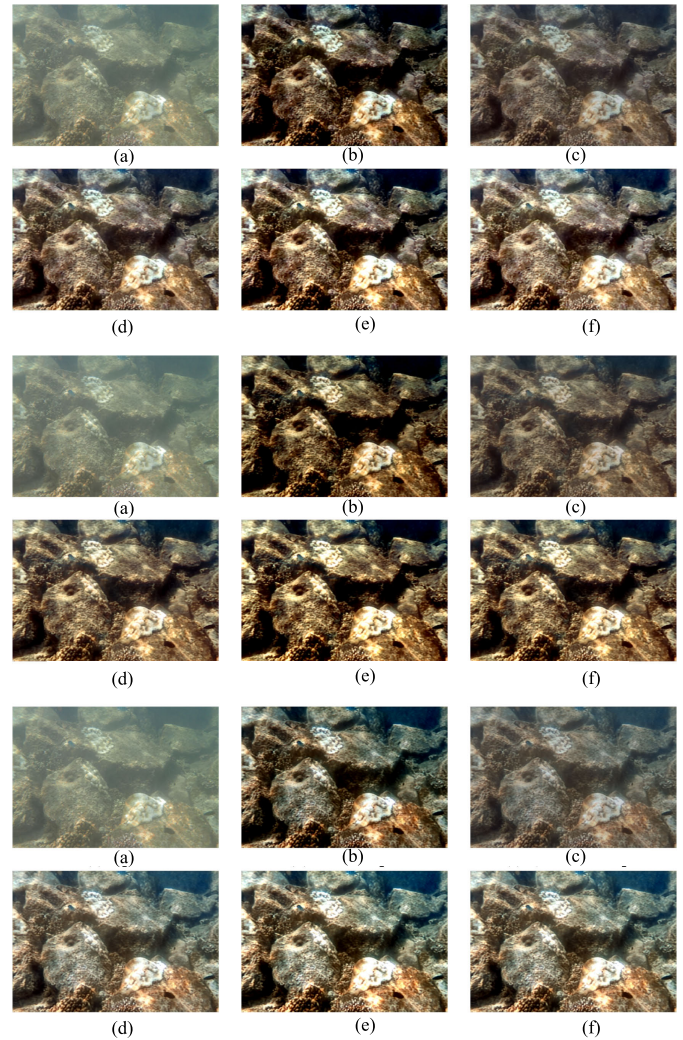


Fig. 14. Visual comparison of the ablation study. Three pieces of the figure correspond to normal-, warm-, and cold-tone ablation study results. (a) Input image. (b) With 3-D color operator. (c) Without a 6-D color operator. (d) Without warm information supervision. (e) Without cold information supervision. (f) Normal/warm/cold tone results.

color operator proposed in this article. Table V and Fig. 14 show ablation results, w/o means without. First, we can observe through w/o warm supervision and w/o cold supervision that there is a significant improvement in warm or cold-tone results directly supervised by multimodal information compared to none, and it can affect normal tone image quality. The 6-D color operators of normal-tone results also contain information about warm and cold tones. An indirect explanation of multimodal information supervision can control image tone and improve image quality. In addition, we can observe that the use of w/o 6-D color operator results in a certain decrease in the effect. In contrast, the use of the 3-D color operator is not as effective as a 6-D color operator, indicating that the 6-D color operator has a significant effect on image color adjustment.

## V. CONCLUSION

This article proposed a novel framework for personalized UIE with different tones. The proposed model consists of two stages, that is, color enhancement and detail enhancement. For color enhancement, a multimodal supervised 6-D color operator is presented. By using text information and multimodal

interactive learning, the proposed method can control the tone of enhanced images and can learn to generate different tones through the same GT without tone information. Experimental results demonstrate that the proposed method can outperform SOTA methods in both subjective and objective evaluation.

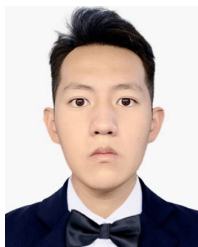
In summary, it has been shown that the method we propose can generate enhancement results of different tones in accordance with individual preferences, which can satisfy the needs of groups of people who like cold tones and warm tones and have no preference for tones. However, the perception of color is also influenced by their emotions. People tend to prefer warm colors when they are passionate, and cool colors when they are calm. As our method cannot detect the emotions of the viewer, we are unable to change the enhancement results of different tones as necessary. A future project aims to design an interactive system that adapts to the emotional changes of the viewer and provides more accurate enhancement results in response.

#### REFERENCES

- [1] R. Schettini and S. Corchs, "Underwater image processing: State of the art of restoration and image enhancement methods," *EURASIP J. Adv. Signal Process.*, vol. 2010, no. 1, pp. 1–14, Dec. 2010.
- [2] M. Jian, X. Liu, H. Luo, X. Lu, H. Yu, and J. Dong, "Underwater image processing and analysis: A review," *Signal Process., Image Commun.*, vol. 91, Feb. 2021, Art. no. 116088.
- [3] S. Raveendran, M. D. Patil, and G. K. Birajdar, "Underwater image enhancement: A comprehensive review, recent trends, challenges and applications," *Artif. Intell. Rev.*, vol. 54, no. 7, pp. 5413–5467, Oct. 2021.
- [4] S. M. Luria and J. A. S. Kinney, "Underwater vision: The physical and psychological bases of the visual distortions that occur underwater are discussed.," *Science*, vol. 167, no. 3924, pp. 1454–1461, Mar. 1970.
- [5] C. Zhang, J. Su, Y. Ju, K.-M. Lam, and Q. Wang, "Efficient inductive vision transformer for oriented object detection in remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5616320.
- [6] C. Zhang, K.-M. Lam, T. Liu, Y.-L. Chan, and Q. Wang, "Structured adversarial self-supervised learning for robust object detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5613720.
- [7] Z. Zhang, H. Zheng, R. Hong, M. Xu, S. Yan, and M. Wang, "Deep color consistent network for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 1899–1908.
- [8] K. Chi, Y. Yuan, and Q. Wang, "Trinity-Net: Gradient-guided Swin transformer-based remote sensing image dehazing and beyond," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4702914.
- [9] C. Li et al., "An underwater image enhancement benchmark dataset and beyond," *IEEE Trans. Image Process.*, vol. 29, pp. 4376–4389, 2019.
- [10] C. Li, S. Anwar, J. Hou, R. Cong, C. Guo, and W. Ren, "Underwater image enhancement via medium transmission-guided multi-color space embedding," *IEEE Trans. Image Process.*, vol. 30, pp. 4985–5000, 2021.
- [11] R. Liu, Z. Jiang, S. Yang, and X. Fan, "Twin adversarial contrastive learning for underwater image enhancement and beyond," *IEEE Trans. Image Process.*, vol. 31, pp. 4922–4936, 2022.
- [12] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and P. Bekaert, "Color balance and fusion for underwater image enhancement," *IEEE Trans. Image Process.*, vol. 27, pp. 379–393, 2017.
- [13] W. Zhang, P. Zhuang, H.-H. Sun, G. Li, S. Kwong, and C. Li, "Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement," *IEEE Trans. Image Process.*, vol. 31, pp. 3997–4010, 2022.
- [14] W. Song, Y. Wang, D. Huang, A. Liotta, and C. Perla, "Enhancement of underwater images with statistical model of background light and optimization of transmission map," *IEEE Trans. Broadcast.*, vol. 66, no. 1, pp. 153–169, Mar. 2020.
- [15] P. Zhuang, J. Wu, F. Porikli, and C. Li, "Underwater image enhancement with hyper-Laplacian reflectance priors," *IEEE Trans. Image Process.*, vol. 31, pp. 5442–5455, 2022.
- [16] L. Shen, M. Reda, X. Zhang, Y. Zhao, and S. G. Kong, "Polarization-driven solution for mitigating scattering and uneven illumination in underwater imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 4202615.
- [17] H. Wang, W. Zhang, L. Bai, and P. Ren, "Metalantis: A comprehensive underwater image enhancement framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5618319.
- [18] D. Liang, J. Chu, Y. Cui, Z. Zhai, and D. Wang, "NPT-UL: An underwater image enhancement framework based on nonphysical transformation and unsupervised learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5608819.
- [19] J. Yin, Y. Wang, B. Guan, X. Zeng, and L. Guo, "Unsupervised underwater image enhancement based on disentangled representations via double-order contrastive loss," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 4201815.
- [20] J. Zhou, Q. Gai, D. Zhang, K.-M. Lam, W. Zhang, and X. Fu, "IACC: Cross-illumination awareness and color correction for underwater images under mixed natural and artificial lighting," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 4201115.
- [21] M. Yu, L. Shen, Z. Wang, and X. Hua, "Task-friendly underwater image enhancement for machine vision applications," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5601014.
- [22] Q. Jiang, Y. Zhang, F. Bao, X. Zhao, C. Zhang, and P. Liu, "Two-step domain adaptation for underwater image enhancement," *Pattern Recognit.*, vol. 122, Feb. 2022, Art. no. 108324.
- [23] M. J. Islam, Y. Xia, and J. Sattar, "Fast underwater image enhancement for improved visual perception," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 3227–3234, Apr. 2020.
- [24] X. Liu, S. Lin, and Z. Tao, "Learning multiscale pipeline gated fusion for underwater image enhancement," *Multimedia Tools Appl.*, vol. 82, no. 21, pp. 32281–32304, Sep. 2023.
- [25] L. Peng, C. Zhu, and L. Bian, "U-shape transformer for underwater image enhancement," *IEEE Trans. Image Process.*, vol. 32, pp. 3066–3079, 2023.
- [26] C. Li, S. Anwar, and F. Porikli, "Underwater scene prior inspired deep underwater image and video enhancement," *Pattern Recognit.*, vol. 98, Feb. 2020, Art. no. 107038.
- [27] L. Shen et al., "U<sup>2</sup>PNet: An unsupervised underwater image-restoration network using polarization," *IEEE Trans. Cybern.*, vol. 54, no. 9, pp. 5164–5177, Sep. 2024.
- [28] A. Radford et al., "Learning transferable visual models from natural language supervision," in *Proc. Int. Conf. Mach. Learn.*, vol. 139, 2021, pp. 8748–8763.
- [29] S. Yang, M. Ding, Y. Wu, Z. Li, and J. Zhang, "Implicit neural representation for cooperative low-light image enhancement," *arXiv:2303.11722*.
- [30] C. Ju, T. Han, K. Zheng, Y. Zhang, and W. Xie, "Prompting visual-language models for efficient video understanding," in *Proc. Eur. Conf. Comput. Vis. Cham, Switzerland: Springer*, 2022, pp. 105–124.
- [31] Z. Liang, C. Li, S. Zhou, R. Feng, and C. C. Loy, "Iterative prompt learning for unsupervised backlit image enhancement," in *Proc. IEEE Int. Conf. Comput. Vis.*, Jun. 2023, pp. 8094–8103.
- [32] J. Wang, K. C. Chan, and C. C. Loy, "Exploring clip for assessing the look and feel of images," in *Proc. AAAI Conf. Artif. Intell. (AAAI)*, vol. 37, 2023, pp. 2555–2563.
- [33] Z. Zheng et al., "Ultra-high-definition image dehazing via multi-guided bilateral learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Sep. 2021, pp. 16180–16189.
- [34] X. Liu, S. Lin, K. Chi, Z. Tao, and Y. Zhao, "Boths: Super lightweight network-enabled underwater image enhancement," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.
- [35] Y. Peng, K. Cao, and P. C. Cosman, "Generalization of the dark channel prior for single image restoration," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2856–2868, Jun. 2018.
- [36] W. Zhang, Y. Wang, and C. Li, "Underwater image enhancement by attenuated color channel correction and detail preserved contrast enhancement," *IEEE J. Ocean. Eng.*, vol. 47, no. 3, pp. 718–735, Jul. 2022.
- [37] H. Song, L. Chang, Z. Chen, and P. Ren, "Enhancement-registration-homogenization (ERH): A comprehensive underwater visual reconstruction paradigm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 6953–6967, Oct. 2022.
- [38] T. Ren et al., "Reinforced Swin-ConvS transformer for simultaneous underwater sensing scene image enhancement and super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4209616.
- [39] T. P. Marques and A. Branzan Albu, "L2UWE: A framework for the efficient enhancement of low-light underwater images using local contrast and multi-scale fusion," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 2286–2295.



- [40] N. Jiang, W. Chen, Y. Lin, T. Zhao, and C.-W. Lin, "Underwater image enhancement with lightweight cascaded network," *IEEE Trans. Multimedia*, vol. 24, pp. 4301–4313, 2022.
- [41] K. Li et al., "Beyond single reference for training: Underwater image enhancement via comparative learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 6, pp. 2561–2576, Jun. 2023.
- [42] Z. Jiang, Z. Li, S. Yang, X. Fan, and R. Liu, "Target oriented perceptual adversarial fusion network for underwater image enhancement," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 10, pp. 6584–6598, Oct. 2022.
- [43] Q. Qi et al., "Underwater image co-enhancement with correlation feature matching and joint learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 3, pp. 1133–1147, Mar. 2022.
- [44] D. Berman, D. Levy, S. Avidan, and T. Treibitz, "Underwater single image color restoration using haze-lines and a new quantitative dataset," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 8, pp. 2822–2837, Aug. 2020.
- [45] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 586–595.
- [46] Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor, "The PIRM challenge on perceptual super resolution," 2018, *arXiv:1809.07517*.
- [47] C. Ma, C.-Y. Yang, X. Yang, and M.-H. Yang, "Learning a no-reference quality metric for single-image super-resolution," *Comput. Vis. Image Understand.*, vol. 158, pp. 1–16, May 2017.
- [48] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Apr. 2012.
- [49] M. Yang and A. Sowmya, "An underwater color image quality evaluation metric," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 6062–6071, Dec. 2015.
- [50] K. Panetta, C. Gao, and S. Agaian, "Human-visual-system-inspired underwater image quality measures," *IEEE J. Ocean. Eng.*, vol. 41, no. 3, pp. 541–551, Jul. 2016.



**Xu Liu** (Student Member, IEEE) received the B.E. degree in electronic and information engineering from Liaoning Technical University, Huludao, China, with a minor in data science and big data technology (Tencent Premier Class) from Tencent Cloud Computing (Beijing) Company Ltd, Shanghai Motong Huakai Education Technology Company Ltd., and Liaoning Technical University, Huludao, China, jointly, in 2022. He is currently pursuing the M.E. degree in information and communication engineering with the School of Computer Science and Information Engineering, Hefei University of Technology, Hefei, China.

His research interests include deep learning and computer vision.



**Yang Zhao** (Member, IEEE) received the B.E. and Ph.D. degrees from the Department of Automation, University of Science and Technology of China, Hefei, China, in 2008 and 2013, respectively.

From September 2013 to October 2015, he was a Post-Doctoral Fellow with the School of Electronic and Computer Engineering, Peking University Shenzhen Graduate School, Shenzhen, China. He is currently a Research Associate Professor with the School of Computer and Information, Hefei University of Technology, Hefei. His research interests include image processing and pattern recognition.



**Kaichen Chi** received the B.E. degree in electronic and information engineering and the M.E. degree in communication and information systems from Liaoning Technical University, Huludao, China, in 2019 and 2022, respectively. He is currently pursuing the Ph.D. degree with the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an, China.

His research interests include image processing and deep learning.



**Zhao Zhang** (Senior Member, IEEE) received the Ph.D. degree from the City University of Hong Kong, Hong Kong, in 2013.

During Ph.D., he visited the National University of Singapore, Singapore, where he worked with Prof. Shuicheng Yan, from February to May 2012. He also visited the Chinese Academy of Sciences, Beijing, China, where he worked with Prof. Cheng-Lin Liu, from September to December 2012. He is currently a Full Professor with the School of Computer and Information, Hefei University of

Technology, Hefei, China. He has authored or co-authored more than 130 technical papers published at prestigious journals and conferences, including 49 IJCV or IEEE/ACM Transactions, and 30 top-tier conference papers, such as CVPR, NeurIPS, ACM MM, ICLR, AAAI, and IJCAI, with Google Scholar citations more than 5900 times and an H-index of 44. His research interests include machine learning, computer vision, and pattern recognition.

Dr. Zhang is a Distinguished Member of the CCF. He is/has been an SPC Member/Area Chair of ACM MM, AAAI, IJCAI, SDM, and BMVC. He was/is an Associate Editor of IEEE TRANSACTIONS ON IMAGE PROCESSING, *Pattern Recognition*, and *Neural Networks*.



**Yanxiang Chen** received the B.Sc. and M.Sc. degrees in electronic information engineering from Hefei University of Technology, Hefei, China, in 1993 and 1996, respectively, and the Ph.D. degree in signal and information processing from the University of Science and Technology of China, Hefei, in 2004.

She was a Visiting Scholar with the University of Illinois at Urbana-Champaign, Champaign, IL, USA, from 2006 to 2008, and the National University of Singapore, Singapore, from 2012 to 2013. She

is currently a Professor with the School of Computer Science and Information Engineering, Hefei University of Technology. Her research interests include multimodal signal processing, pattern recognition, and machine learning.



**Wei Jia** (Member, IEEE) received the B.Sc. degree in informatics from Central China Normal University, Wuhan, China, in 1998, the M.Sc. degree in computer science from Hefei University of Technology, Hefei, China, in 2004, and the Ph.D. degree in pattern recognition and intelligence systems from the University of Science and Technology of China, Hefei, in 2008.

He was a Research Assistant Professor and an Associate Professor with Hefei Institutes of Physical Science, Chinese Academy of Sciences, Beijing, China, from 2008 to 2016. He is currently a Professor with the School of Computer and Information, Hefei University of Technology. His research interests include computer vision, biometrics, pattern recognition, and image processing.