Joint Iterative Color Correction and Dehazing for Underwater Image Enhancement

Kun Wang, Liquan Shen, Yufei Lin, Mengyao Li, and Qijie Zhao

***Abstract*—The captured underwater images suffer from color cast and haze effect caused by absorption and scattering. These interdependent phenomena jointly degrade images, resulting in failure of autonomous machines to recognize image contents. Most existing learning-based methods for underwater image enhancement (UIE) treat the degraded process as a whole and ignore the interaction between color correction and dehazing. Thus, they often obtain unnatural results. To this end, we propose a novel joint network to optimize the results of color correction and dehazing in multiple iterations. Firstly, a novel triplet-based color correction module is proposed to obtain color-balanced images with identical distribution of color channels. By means of inherent constraints of the triplet structure, the information of channel with less distortion is utilized to recover the information of other channels. Secondly, a recurrent dehazing module is designed to alleviate haze effect in images, where the Gated Recurrent Unit (GRU) as the memory module optimizes the results in multiple cycles to deal with severe underwater distortions. Finally, an iterative mechanism is proposed to jointly optimize the color correction and dehazing. By learning transform coefficients from dehazing features, color features and basic features of raw images are progressively refined, which maintains color balanced during the dehazing process and further improves clarity of images. Experimental results show that our network is superior to the existing state-of-the-art approaches for UIE and provides improved performance for underwater object detection.**

***Index Terms*—Marine robotics, underwater image enhancement, deep learning in robotics, recurrent network, computer vision for automation.**

I. Introduction

In RECENT years, marine engineering and research have increasingly relied on underwater images captured by autonomous underwater vehicles (AUVs) and remotely operated vehicles (ROVs). However, one major operational challenge for these underwater robots is that underwater images usually suffer from degradation, such as color cast and haze effect due to wavelength-dependent light absorption and scattering [1]–[3]. The degradation severely affects the performance of vision-based tasks such as tracking, detection, classification, segmentation, and visual serving [4]. Thus, there is an urgent need to improve visibility in underwater images and underwater machine vision.

The degradation of underwater images including color cast and haze effect is caused by light absorption and scattering in water. Eq. 1 shows the simplified degradation process of underwater images derived from [5]:

$$I_\lambda(x) = J_\lambda(x)C_{\alpha_\lambda}(x)R_{\beta_\lambda}(x) + B_{\beta_\lambda}(x) \tag{1}$$

where $I_\lambda(x)$ is the distorted image. $J_\lambda(x)$ is the clear image. $C_{\alpha_\lambda}(x)$ is the term of absorption attenuation affected by the absorption coefficient α_λ. $R_{\beta_\lambda}(x)$ and $B_{\beta_\lambda}(x)$ are scattering attenuation terms of foreground and background, affected by total attenuation coefficient β_λ. The existing UIE methods are divided into traditional methods and learning-based methods. However most of these methods ignore changes of color during dehazing process, which leads to obtain unnatural results.

The existing traditional UIE algorithms are mainly divided into model-based approaches [6], [7] and model-free approaches [8]–[12]. The model-free approaches improve quality of underwater images by raising sharpness and contrast. The model-based methods estimate parameters of physical models (background light or transmission map) to obtain clear images. Even though these methods achieve relatively good results, they may perform poorly when their models are invalid on some extremely distorted images.

Many learning-based approaches for UIE [13]–[18] are employed to estimate physical parameters or directly predict clear images. However, these methods are data-driven and their networks are designed without considering underwater absorption and scattering. Besides, they handle the color correction and dehazing of underwater images separately, causing the destruction of balanced color during the dehazing process.

To address above problems, we propose a novel iterative network to jointly correct color and dehaze. The framework of our network is shown in Fig. 1. Contributions of the work are listed as follows: 1) a joint iterative network for UIE is proposed, which divides the enhancement into two sub-tasks including color correction and dehazing corresponding to light

Manuscript received November 24, 2020; accepted March 8, 2021. Date of publication March 31, 2021; date of current version April 22, 2021. This letter was recommended for publication by Associate Editor M. Ghaffari and Editor C. Cadena Lerma upon evaluation of the reviewers' comments. This work was supported in part by the National Natural Science Foundation of China under Grants 61931022 and 61671282, in part by the Open Fund of Key Laboratory of Advanced Display and System Applications of Ministry of Education (Shanghai University), in part by the Shanghai Science and Technology Innovation Plan under Grant 18010500200, and in part by the Shanghai Shuguang Program under Grant 17SG37. *(Corresponding author: Liquan Shen.)*

Kun Wang, Yufei Lin, and Mengyao Li are with the Shanghai Institute for Advanced Communication and Data Science, Shanghai University, Shanghai 200444, , China (e-mail: fangfang_tu@163.com; anzhonglyf@163.com; sdlmy@shu.edu.cn).

Liquan Shen is with the Key laboratory of Specialty Fiber Optics and Optical Access Networks, Joint International Research Laboratory of Specialty Fiber Optics and Advanced Communication, Shanghai University, Shanghai 200444, China (e-mail: jsslq@163.com).

Qijie Zhao is with the School of Mechatronics Engineering and Automation, Shanghai University, Shanghai 200444 (e-mail: zqj@shu.edu.cn).

Digital Object Identifier 10.1109/LRA.2021.3070253

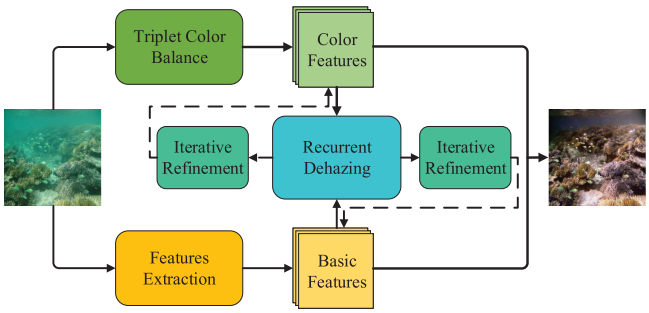

Fig. 1. The framework of our network for underwater image enhancement (UIE). The modules of Triplet Color Balance (TCBM) and Features Extraction derives color features and basic features from input image. Employing the results of Recurrent Dehazing Module (RDM), the features are refined by Iterative Refine Module (JIRM) in multiple iterations. The enhanced image is obtained by refined color features and basic features finally.

absorption and scattering; 2) a novel color correction module is designed to correct color of underwater images. By means of shared weights of the triplet structure, the information of channel with less distortion is utilized to recover the information of other channels; 3) a novel iterative algorithm is proposed to optimize sub-tasks of color correction and dehazing. The algorithm learns transform coefficients to progressively update results of color correction and dehazing tasks. The joint refinement of sub-tasks reduces losses caused by decoupling between color correction and dehazing to obtain more realistic results. Experimental results reveal that our method can obtain color-balanced and clear underwater images and improve object detection performance.

II. Related Work

A. Traditional UIE Methods

Model-free methods: This type of algorithms adjust pixel values of given images without explicitly modeling the image formation process. The adjustments are mainly performed in the spatial domain including histogram equalization [19], the gray world algorithm [20], contrast limited adaptive histogram equalization (CLAHE) [21], multiscale retinex with color restoration [22], automatic white balance [23], and color constancy [24], [25]. The spatial domain methods can improve the visual quality to some extent, but may accentuate noise, introduce artifacts, and cause color distortions. In current work, Ancuti *et al.* [6] propose an UIE method by blending a contrast enhanced image and a color-corrected image using a multi-scale fusion strategy. These enhancement methods just utilize statistical characteristics of images and rely on the observed information. However, they ignore the mutual influence between color distortion and haze effects. Thus, their generalization performance is poor.

Model-based methods: Traditional model-based methods introduce prior knowledge of underwater scene or estimate physical model parameters e.g. background light or transmission map. Chi *et al.* [5] estimate model parameters of attenuation process to obtain clean images. In [26], the Dark Channel Prior (DCP) [27] is modified as Underwater Dark Channel Prior (UDCP) to handle the serious attenuation of red light in water. Researchers also propose various physical priors designated to underwater images, other than those priors derived from DCP. Li *et al.* [8] propose an underwater image enhancement method based on the minimum information loss principle and histogram distribution prior to estimate transmission map. Peng *et al.* [11] propose a depth estimation method based on image blurriness and light absorption, which is employed in an underwater image formation model to enhance underwater images. Even though these methods achieve relatively good results, they depend on specific scenes and lighting conditions, and are even related to temperature and turbidity. When their models are invalid on some extremely distorted images, they may perform poorly. For example, it is well known that DCP is inapplicable to white objects or regions. Further, errors in the estimation of separated physical model parameters will be gradually accumulated, eventually making the results deviate from expectations.

B. Learning-Based UIE Methods

In the past decades, the community witnesses the great success of deep neural networks in many low-level and high-level computer vision tasks such as image colorization [28], color/contrast adjustment [29], dehazing [30], etc. Different from other low-level vision problems where the ground truth can be easily obtained, it is challenging to achieve a large amount of paired degraded underwater images and the corresponding ground truth. Therefore, most learning-based methods for UIE usually use synthetic datasets for training. Li *et al.* [13] fuse pictures prepossessed by white balance, gamma correction and histogram equalization to enhance images. In [31], the enhancement process is divided into two isolated stages followed by color correction and dehazing. Through assigning different weights to characteristics of each color channel, Dudhane *et al.* [17] propose an enhancement method based on CNN to obtain clear images.

In recent works, GAN-based models show great success for generating realistic images in an unsupervised pipeline that only relies on an unlabeled set of images with a desired representation. Li *et al.* [15] propose a deep learning-based underwater image enhancement model, WaterGAN, which comprises of an end-to-end network consisting of depth estimation module followed by color correction module. Guo *et al.* [32] present a multi-scale dense GAN which combines GAN loss with L1 loss and gradient loss to learn the distribution of features for UIE.

It can be noted that most currently published learning-based methods ignore correlation between color and haze. Besides, few of them consider the design of unique network structures to deal with color cast but rely more on datasets or loss to solve. Thus, we propose a joint network to refine sub-tasks of color correction and dehazing, which can solve these problems and obtain better results.

III. Proposed Method

This letter presents a joint iterative approach for UIE. Fig. 2 shows an overview of our network. Triplet Color Balance Module (TCBM) is the first component of out network, performing

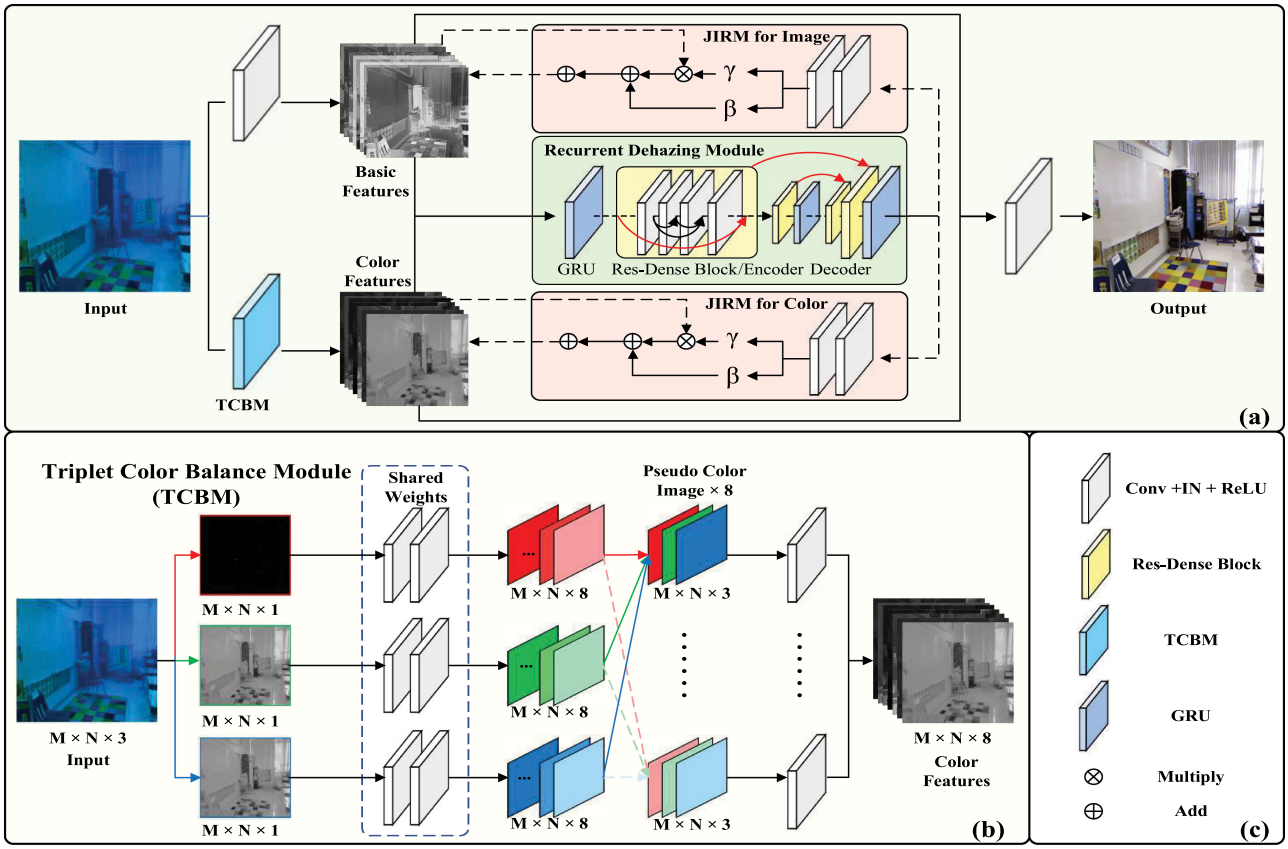

Fig. 2. The specific details of our network. (a) The specific details of Recurrent Dehazing Module and Joint Iterative Refine Module. (b) The detail of Triplet Color Balance Module (TCBM) module. (c) The legend of (a) and (b).

elimination of color cast via correlation between color channels. Recurrent Dehazing Module (RDM) as the second component of our method can obtain dehazing features through multiple iterations. Jointly Iterative Refine Module (JIRM) as the core part of the entire network, jointly updates color features from TCBM and basic features from raw images via dehazing features in each iterations.

A. *Triplet Color Balance Module*

Inspired by the gray world assumption [20], color of undistorted images is balanced that means distribution of each RGB component is same. However, the distribution of each RGB component in underwater images is different due to the partial absorption of light. Fig. 3(a) shows a sample underwater image and its corresponding histogram. It can be observed that the pixel values of G and R channels are lower than pixel values of B channel. To this end, Triplet Color Balance Module (TCBM) is designed to handle color cast caused by differences between distributions of RGB components. It utilizes channel correlation to generate the pseudo color images ($\hat{I}_i$) whose distributions of RGB components are same. The overall details of TCBM can be seen in Fig. 2(b).

Basically, each RGB component contributes some information towards the color cast. Thus, they are separately processed as input. However, complete separation of channels loses channel correlation, and it is very difficult to restore color information with serious distortion. Inspired by the Siamese network, each color channel is regarded as the similar sample of the identical scene and thus a convolutional layer with shared weights is used to extract features corresponding to each channel. The shared-weight convolution layers in the training process try to extract the same features from each RGB channel to restore color. The channel with less attenuation dominates the training optimization, so that its information can be used to compensate the channel with more attenuation. Moreover, to further improve the flow of information between channels, the features from RGB components are concatenated by channel-wise concatenation (CwC) layer to obtain new color images. The generation process can be presented mathematically as:

$$\hat{I}_i = CwC\left\{TC_i\left(I^R\right), TC_i\left(I^G\right), TC_i\left(I^B\right)\right\} \tag{2}$$

where TC represents triple shared-weight convolution layers containing two blocks of Conv+IN (instance normalization)+ReLU, and $i \in \{1:8\}$ is the index of color features.

Ideally, the generated images are color balanced since color features from shared-weight layer are same. Considering different information from features, the generated images are called as pseudo color images with different color balance information. Fig. 3(c)-Fig. 3(f) demonstrate pseudo color images and corresponding histograms. Compared with Fig. 3(a), pseudo color images $\hat{I}_i$ own distributions of the pixel values in different color

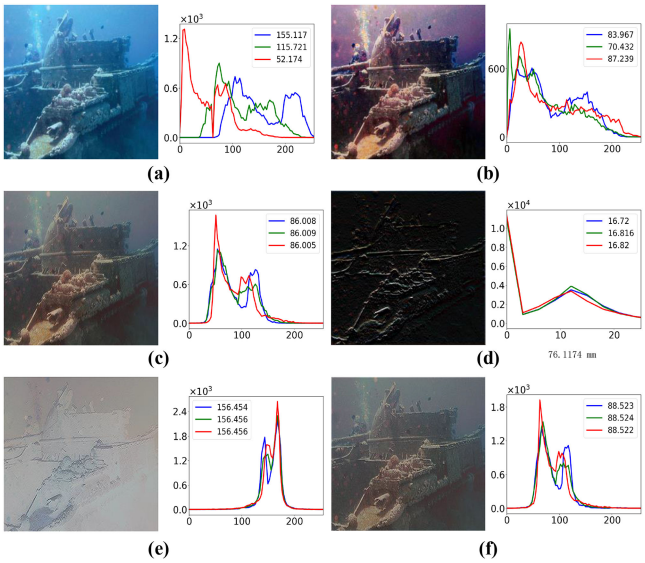

Fig. 3. (a) The original image and its corresponding histogram. (b) The enhanced image and its corresponding histogram. (c)–(f) The generated pseudo color images and theirs corresponding histograms. The values in the legend of images represent the mean values of RGB channels.

channels consistent with the same mean value, which are in line with nature images. Besides, it can be found that the content of each $\hat{I}_i$ is different. Hence, a parallel convolution layer is utilized to convert multiple $\hat{I}_i$ into color features (CF) as follows:

$$CF = \underset{i\in\{1:8\}}{concat}\left(f_i^m * \hat{I}_i + b_i\right) \tag{3}$$

where f_i^m and b_i represent the filters and biases, respectively. m is the filter size and i is the index of the pseudo color images.

CF represents correct color information. However, after CF undergoes multiple convolution operations, the texture structure information will be partially discarded. Thus, the basic features called BF are extracted from raw image through the layer of Feature Extraction to retain original texture structure information. Both BF and CF are advanced features and can better guide the next stage of enhancement.

B. Recurrent Dehazing Module

The haze effect in underwater images is caused by light scattering of suspended particles. Moreover, the complexity and diversity of underwater environment makes simple network structures unable to restore excessively distorted colors and extremely blurred details, resulting in unsatisfactory enhancement results. Inspired by deraining method, we apply recurrent blocks to construct a dehazing module (RDM) and progressively remove haze from CF and BF as shown in Fig. 2(a).

The RDM is designed based on Unet with Residual-Dense blocks. Basically, one underwater image can not be used as a sequence. Thus, the last output of the RDM is used as the input of next cycle, so that RDM removes haze stage by stage by incorporating information from previous cycle. The Gated Recurrent Unit (GRU) [33] is utilized as memory units to keep useful information and guide the next cycle.

We add three memory modules between the encoder and decoder to retain information of input features, intermediate high-dimensional features, and output features. Unlike conventional GRU [33] used in the entire model, we add GRU at key nodes in the structure in each step which significantly reduces the required model size. The whole block can be described mathematically as:

$$h_1(n) = f_{GRU}\left(h_1(n-1), x(n)\right) \tag{4}$$

$$y_{encoder} = f_e\left(h_1(n)\right) \tag{5}$$

$$h_2(n) = f_{GRU}\left(h_2(n-1), y_{encoder}\right) \tag{6}$$

$$y_{decoder} = f_d\left(h_2(n)\right) \tag{7}$$

$$y_{dehaze}(t) = f_{GRU}\left(y_{dehaze}(n-1), y_{\text{decoder}}\right) \tag{8}$$

$$1 \le n \le N \tag{9}$$

where $x(n) = \text{concat}(CF, BF)$ is input of RDM. $h(n-1)$ represents the output of the previous state of GRU module. $h(n)$ represents the output of current state and will participate in the next operation. n is the index of cycle. N is the max value of cycle numbers.

C. Jointly Iterative Refine Module

As the analysis in section I, color cast occurs throughout the whole degradation process, which can not be isolated from the haze effect. Although color correction and dehazing are executed successively in images, balanced color will be destroyed during the dehazing process in [31]. To this end, we propose a Jointly Iterative Refine Module (JIRM) which utilizes dehazing features to refine basic features and color features based on iterative algorithm as shown in Fig. 2(a).

Equation 1 shows that the degradation process of haze images is mainly formed by scale transformation R_{β_λ} and translation B_{β_λ}. Thus, a refine module (Ψ) is designed to learn the parameters of scale transformation and translation to reverse the degradation process to eliminate the influence of haze effect. It utilizes three convolution layers to extract conditional maps from dehazing features. The conditional maps are fed to two other convolution layers to predict the modulation parameters, γ and β respectively. Based on the refine module, the iterative update process can be expressed as follows:

$$f_{dehaze}(n) = f_{RDM}(BF_{n-1}, CF_{n-1}) \tag{10}$$

$$\beta_\theta, \gamma_\theta = \Psi_\theta(f_{dehaze}(n)), \theta \in (b, c) \tag{11}$$

$$BF_n = (1+\gamma_b) * BF_{n-1} + \beta_b \tag{12}$$

$$CF_n = (1+\gamma_c) * CF_{n-1} + \beta_c \tag{13}$$

where f_{dehaze} indicates the dehazing features from RDM.

Based on this update algorithm, color features and basic features are updated in each iteration, as the input of the next loop. After finishing all iterations, color features and basic features can accurately represent the characters of no-water images. Finally, features are used to reconstruct enhanced images through reconstruction layer, which comprises of a convolutional layer and an activation layer.

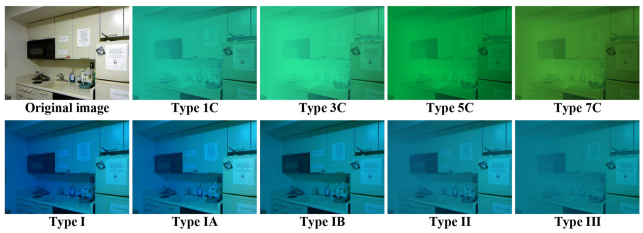

Fig. 4. Samples of synthetic underwater images with different types based on NYU-depth2 dataset.

D. Loss Function

Two loss functions are utilized to train our network. L1 Loss as the content loss is used to maintain consistency between pixels. It is defined as follows,

$$L_c = \left\| I^{out} - I^{label} \right\|_1 \tag{14}$$

where I^{out} and I^{label} denote the enhanced and label image from our synthetic dataset, respectively.

The other loss is the perceptual loss [34], which compares the features obtained from enhanced images with the features from the original images, making the high-level information close. The perceptual loss defined on the VGG-19 network pre-trained on ImageNet can be represented as:

$$L_p = \frac{1}{H * W} \sum_{i=1}^{H} \sum_{j=1}^{W} \left[\varphi \left(I^{out} \right)_{i,j} - \varphi \left(I^{label} \right)_{i,j} \right]^2 \tag{15}$$

where $\varphi(\cdot)$ represents the feature map of the pool-3 layer of the VGG-19 network pre-trained on ImageNet. H, W is height and width of feature maps.

The total loss function is defined as:

$$L_t = \lambda_1 L_c + \lambda_2 L_p \tag{16}$$

where λ_1, λ_2 is the hyper-parameter which is empirically set to 0.8 and 0.2, respectively.

IV. Experimental Results

A. Implementation Details

Dataset: A synthetic underwater image dataset is proposed for paired training of our method. The synthetic dataset is generated from Eq. 1, which contains 10 water types defined in [35], each of which has 1449 images from NYU dataset. Fig. 4 shows the samples of our synthetic dataset. Furthermore, the constructed dataset is divided into two subsets: a training set of 9900 (1100×9) samples, and a validation set of 3141 (349×9) samples. For the quantitative evaluation on real underwater images, we use UIEBD [13], EUVP [18] and UFO-120 [4] datasets. Both in training and testing phase, we randomly crop all the images to 256×256 and normalize the pixel values to $[-1, 1]$. The synthetic dataset is available at *https://github.com/zyWang-Power/ANA-SYN*.

Training details: The proposed network is trained in an end-to-end manner, which is implemented based on the PyTorch framework and trained for 200 epochs on NVIDIA Titan V. Adam is used as the optimization algorithm with batch size of 2. Default values of β_1 and β_2 are 0.5 and 0.999, respectively. The weight decay is set as 0.00 005. The learning rate starts at 0.0002, which is divided by 10 after 100 epochs. In addition, the training examples are augmented by randomly rotating and horizontal flipping.

Comparison methods: We evaluate our method against the following state-of-the-art approaches: Fusion [6], UIBLA [11], Two-Step [7], HP [8], UGAN [16], FUIE-GAN [18] and Water-Net [13]. The peak signal to noise ratio (PSNR) and structural similarity (SSIM) are used as the full reference evaluation in synthetic images. Besides, the UIQM [36] is employed as the no reference objective evaluation indicator to measure the performances of algorithms on real underwater images.

TABLE I
Quantitative Comparison (Average PSNR/SSIM) of the UIE Results on Synthetic Dataset

Methods	PSNR(dB)	SSIM
HP [8]	12.796	0.643
Fusion [6]	11.439	0.639
UIBLA [11]	9.445	0.458
Two-Step [7]	12.513	0.632
UGAN [16]	15.722	0.709
Water-Net [13]	13.388	0.627
FUIE-GAN [18]	14.241	0.615
Ours	**29.945**	**0.946**

B. Experiments on Synthetic Dataset

Table I shows the quantitative comparison of UIE results. As shown, our method achieves the highest PSNR and SSIM value on the synthetic dataset. Compared to the state-of-the-art method UGAN [16], our approach achieves the gain with 14.22 dB in terms of PSNR and 0.24 in SSIM on the synthetic dataset, respectively. Through estimating blurriness degrees as scene information to UIE, UIBLA [11] performs slightly worse than other methods since blurriness cannot accurately represent haze effect and color cast of images. Both UGAN [16] and FUIE-GAN [18] are trained via real-world datasets and thus can achieve relatively better results. The substantial improvements in performance show that our network has strong learning capabilities and can better handle underwater distortions.

C. Experiments on Real Dataset

Evaluate color accuracy: To evaluate the color accuracy of our method, we test the average angular reproduction error on SQUID dataset [12]. We test 10 images containing 36 color boards from SQUID dataset. Fig. 5 shows a sample image and the statistical graph about color error of all boards on corresponding enhanced image of different methods. It can be seen that our method achieves low errors in all color boards. The second column of Table II lists the average error values for all evaluated images. We can see from Table II that our method achieves the lowest average error compared to other methods.

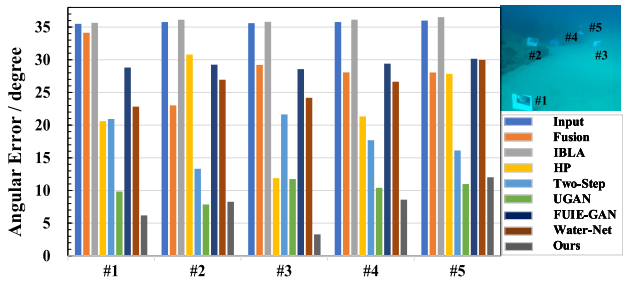

Fig. 5. The statistical graph about angular reproduction errors which are measured for different underwater image enhancement methods on the sample image from SQUID dataset. The board' order is by increasing distance from the camera.

TABLE II
QUANTITATIVE COMPARISON RESULTS (UIQM/COLOR ERROR) OF DIFFERENT METHODS ON REAL DATASETS (UIEBD, EUVP, AND UFO-120)

Methods	Color Error↓	UIQM↑		
	SQUID	UIEBD	EUVP	UFO-120
HP [8]	18.32	2.787	3.109	2.676
Fusion [6]	30.74	3.015	2.905	2.893
UIBLA [11]	34.94	2.393	2.133	2.150
Two-Step [7]	22.36	3.144	2.943	2.909
UGAN [16]	12.12	2.834	2.885	2.893
Water-Net [13]	23.12	3.089	3.109	2.958
FUIE-GAN [18]	28.00	3.121	3.223	3.090
Ours	**7.87**	**3.144**	**3.245**	**3.137**

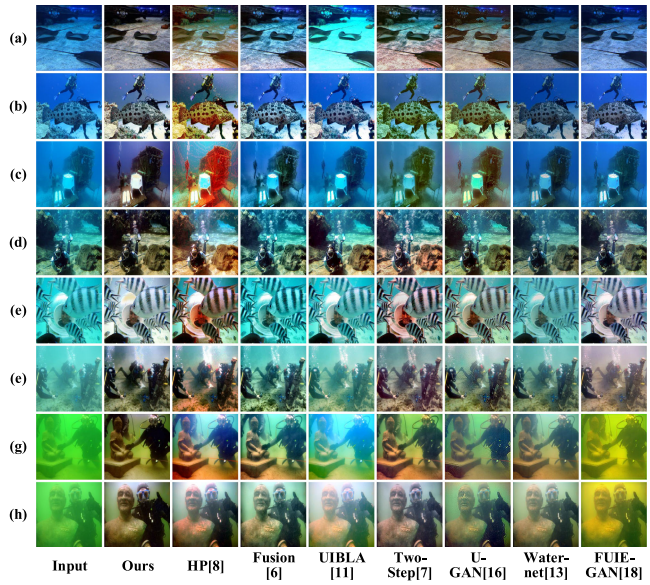

Fig. 6. Visual comparison results of different methods including Fusion [6], UIBLA [11], Two-Step [7], HP [8], UGAN [16], FUIE-GAN [18], Water-Net [13] and ours on 8 sample images from real underwater datasets.

TABLE III
AVERAGE RECOGNITION ACCURACY BETWEEN REAL IMAGES AND ENHANCED IMAGES

	Input	Ours
Recognition Accuracy	50.3%	**56.9%**

Quantitative result: To further show the robustness of our network, we compare the visual and quantitative results of different approaches on real-world datasets. Table II shows the evaluation of state-of-the-art methods and our method on real-world datasets. It is clearly observed that our approach achieves the highest UIQM values in all datasets.

Qualitative result: Fig. 6 shows the visual results of different approaches on real datasets. The traditional model-free methods including Fusion [6] and Two-step [7] cannot eliminate color distortion well due to the inability to adapt to the more complex underwater environment as shown in Fig. 6(c) and Fig. 6(e). The range-dependency methods including UIBLA [11] and HP [8] achieve good UIQM values, which show serious overexposure or over-enhancement on some real underwater images as shown in Fig. 6(a) and Fig. 6(c).

Although they consider the depth information, the inaccurate depth misleads their enhancement. The learning-based methods including Water-Net [13], UGAN [16] and FUIE-GAN [18] perform relatively well. However, they cannot completely eliminate color cast especially in green images and details of enhanced images are still blurry as shown in Fig. 6(g) and Fig. 6(h), since they ignore correlation between color correction and dehazing. Compared to these methods, our method can both remove haze and restore color cast effectively. For extremely distorted images as shown in Fig. 6(e)-Fig. 6(h), our network restores the color of images well. Besides, our network attempts to enhance images to no-water images without color cast and haze effect like Fig. 6(a)-Fig. 6(d). The pleasant results show that our network outperforms other existing methods for UIE.

Application: To better present application of our method on object recognition, we analyze the performance of for underwater object detection in the state-of-the-art method [4] on SUIM dataset [37]. Since this method can only recognize robots and divers, we use accuracy rate in the classification problem to measure performance, which is calculated by following formula:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{17}$$

where TP represents the numbers of true positives. TN represents the numbers of true negatives. FP represents the numbers of false positives. FN represents the numbers of false negatives. Table III shows quantitative results on recognition accuracy between original underwater images and enhanced images of our method. It can be seen that the accuracy of object recognition increases by 13% based on the enhanced image of our method.

D. Ablation Study

1) Contributions of Quantity of Pseudo Images: We present an ablation experiment to analyze the influences of pseudo image numbers. We train our network on different pseudo image numbers (2, 4, 8, 16). Table IV shows comparison results on PSNR (dB) and parameters of different pseudo image numbers. It can be seen that with increase of pseudo image number, the performance is gradually increased, which eventually tends to

TABLE IV
ABLATION STUDY: QUANTITATIVE RESULTS(PSNR/PARAMETERS) ON DIFFERENT NUMBERS OF PSEUDO IMAGES

Number	2	4	8	16
PSNR/dB	29.42	29.79	29.95	29.99
Parameters	1703661	1709599	1722339	1751275

TABLE V
ABLATION STUDY: QUANTITATIVE RESULTS OF NETWORK WITH DIFFERENT FEATURES

	BF only	CF only	$BF + CF$
PSNR/dB	29.68	29.22	**29.945**

TABLE VI
ABLATION STUDY: QUANTITATIVE RESULTS OF SHARED-WEIGHT LAYERS

	Shared weight	Not shared-weight
PSNR/dB	**29.95**	29.87

be flat. However, increase in quantity of pseudo images leads to a rapid increase in network parameters. Overall, we finally decide 8 as the optimal number in our network considering the trade-off of performance and parameters.

2) Contributions of BF and CF: To verify the differences between basic features and color features, we add an ablation experiment in case of the network with BF only, the network with CF only and the network with $BF + CF$. We can see from Table V that the performance on network with BF only is 0.46 dB on PSNR higher than the network with CF only, and the network with $BF + CF$ achieves the best performance. The main reason is that a lot of texture details are lost due to multiple convolutions in the process of obtaining color features. Thus, the basic features extracted from the original image can supplement the part of the missing information.

3) Contributions of Shared-Weight Layers: To evaluate the effectiveness of shared weights, we add an ablation experiment with shared-weight convolution layers or not. Table VI shows that the network with shared weights improves the enhanced image quality by 0.08 dB on PSNR compared with the network without shared weights. Thus, it can verify that although the shared-weight convolution layers restrict learning degrees of freedom, this relation among channels can better provide information for color restoration.

4) Contributions of TCBM, RDM and JIRM: To demonstrate the effectiveness of each component in our network, we conduct ablation experiments comprised of five structure: 1) BL: a Unet based on Res-Dense Block without TCBM and JIRM. 2) BL+TCBM: TCBM module based on BL. 3) BL+TCBM+RDM: BL+TCBM along with the GRU module. 4) BL+TCBM+RDM+JIRM (+): BL+TCBM +RDM along with iterative update of residuals. 5) BL+TCBM+RDM+JIRM (β, γ): BL+TCBM+RDM along with our proposed iterative update network.

Table VII shows the performances of the above five structures on the synthetic dataset. As a conclusion, these ablation experiments demonstrate that TCBM, RDM and JIRM are useful to reduce noise caused by color cast and haze effect and improve underwater image quality.

TABLE VII
ABLATION STUDY: QUANTITATIVE RESULTS OF DIFFERENT MODELS ON SYNTHETIC DATASET DOMAIN

Methods	PSNR (dB)	SSIM
BL	27.837	0.936
BL+TCBM	28.990	0.941
BL+TCBM+RDM	29.445	0.944
BL+TCBM+RDM+JIRM (+)	29.808	0.946
BL+TCBM+RDM+JIRM (β, γ)	29.945	0.949

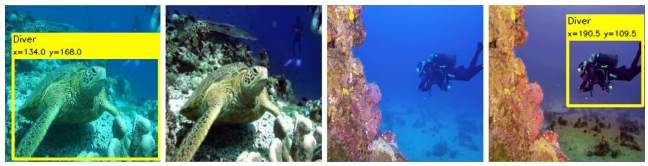

Fig. 7. Comparison results on object recognition between the original underwater images and enhanced images of our method.

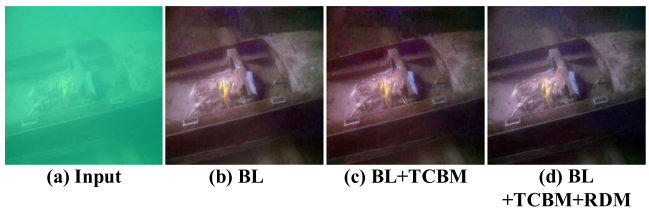

Fig. 8. Ablation study of our model using different blocks. (a) Input; (b) BL; (c) BL+TCBM; (d) BL+TCBM+RDM.

TABLE VIII
PARAMETERS AND FLOPS OF LEARNING-BASED METHODS

	Ours	UGAN	FUIE-GAN	Water-Net
Parameters	2.42M	57.20M	17M	1.09M
Flops	221G	2664G	0.08G	1229G

Fig. 8 shows the visual comparisons of our network with different blocks. As shown in Fig. 8(c), when the process of color correction is executed alone in images, color in images appeals deviant (BL+TCBM). The joint learning of TCBM and RDM makes the details of image more explicit, which can correct color accurately as shown in Fig. 8(d).

E. Discussion of Computational Aspect

We calculate parameters and flops of different learning-based methods. TABLE VIII shows the comparison results on parameters and flops. It can be seen that our method has relatively few parameters and flops. Although the flops of FUIE-GAN are 0.08 G, its parameters are 14.58 M, which are larger than ours. The parameters of Water-Net are 1.33 M less than ours. However, its flops are 1008 G, which are larger than ours.

F. Limitations and Failure Cases

Our work is supervised trained on synthetic dataset, which needs a set of pre-labeled images to compute the losses. Thus, it is important to synthesize the training dataset which can simulate real underwater images well. Sometimes, real images do not match the distribution of our synthetic images leading to poor results on some special cases such as images with artificial light or shallow water images. In the future, we will improve the way to synthesize dataset and solve these failure cases.

V. Conclusion

In this letter, we propose a novel joint iterative network for UIE. The network contains three modules including color correction, dehazing and joint optimization. Firstly, a novel triplet-based color correction module is proposed to correct distorted color. Secondly, a recurrent dehazing module is designed to alleviate haze effect in images, where the Gated Recurrent Unit as the memory module gradually optimizes the dehazing results in multiple cycles to deal with severe underwater distortions. Finally, the iterative mechanism guarantees that the features of haze and color appropriately representing the underwater condition can be jointly learned. Extensive experimental evaluations demonstrate that our method outperforms the state-of-the-art methods. In the future, we plan to investigate its feasibility in other underwater robot vision applications.

References

[1] M. J. Islam, M. Ho, and J. Sattar, "Understanding human motion and gestures for underwater human-robot collaboration," *J. Field Robot.*, vol. 36, pp. 851–873, Nov. 2019.

[2] H. Lu, Y. Li, and S. Serikawa, "Underwater image enhancement using guided trigonometric bilateral filter and fast automatic color correction," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 3412–3416.

[3] S. Zhang, T. Wang, J. Dong, and H. Yu, "Underwater image enhancement via extended multi-scale retinex," *Neurocomputing*, vol. 245, pp. 1–9, Mar. 2017.

[4] M. J. Islam, M. Fulton, and J. Sattar, "Toward a generic diver-following algorithm: Balancing robustness and efficiency in deep visual detection," *IEEE Robot. Autom. Lett.*, vol. 4, no. 1, pp. 113–120, Jan. 2019.

[5] J. Y. Chiang and Y. Chen, "Underwater image enhancement by wavelength compensation and dehazing," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1756–1769, Apr. 2012.

[6] C. Ancuti, C. O. Ancuti, T. Haber, and P. Bekaert, "Enhancing underwater images and videos by fusion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2012, pp. 81–88.

[7] X. Fu, Z. Fan, M. Ling, Y. Huang, and X. Ding, "Two-step approach for single underwater image enhancement," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst.*, Jan. 2017, pp. 789–794.

[8] C. Li, J. Guo, R. Cong, Y. Pang, and B. Wang, "Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior," *IEEE Trans. Image Process.*, vol. 25, no. 12, pp. 5664–5677, Dec. 2016.

[9] P. L. J. Drews, E. R. Nascimento, S. S. C. Botelho, and M. F. Montenegro Campos, "Underwater depth estimation and image restoration based on single images," *IEEE Comput. Graph Appl.*, vol. 36, no. 2, pp. 24–35, Mar./Apr. 2016.

[10] A. Galdran, D. Pardo, A. Picón, and A. Alvarez-Gila, "Automatic red-channel underwater image restoration," *J. Vis. Commun. Image Representation*, vol. 26, pp. 132–145, Jan. 2015.

[11] Y. Peng and P. C. Cosman, "Underwater image restoration based on image blurriness and light absorption," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1579–1594, Apr. 2017.

[12] 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.*, to be published, doi: 10.1109/TPAMI.2020.2977624.

[13] C. Li *et al.*, "An underwater image enhancement benchmark dataset and beyond," *IEEE Trans. Image Process.*, vol. 29, pp. 4376–4389, Nov. 2020.

[14] 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.

[15] J. Li, K. A. Skinner, R. M. Eustice, and M. Johnson-Roberson, "WaterGAN: Unsupervised generative network to enable real-time color correction of monocular underwater images," *IEEE Robot. Autom. Lett.*, vol. 3, no. 1, pp. 387–394, Jan. 2018.

[16] C. Fabbri, M. J. Islam, and J. Sattar, "Enhancing underwater imagery using generative adversarial networks," in *Proc. IEEE Int. Conf. Robot. Automat.*, May 2018, pp. 7159–7165.

[17] A. Dudhane, P. Hambarde, P. Patil, and S. Murala, "Deep underwater image restoration and beyond," *IEEE Signal Process. Lett.*, vol. 27, pp. 675–679, Apr. 2020.

[18] 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.

[19] M. Abdullah-Al-Wadud, M. H. Kabir, M. A. A. Dewan, and O. Chae, "A dynamic histogram equalization for image contrast enhancement," *IEEE Trans. Cons. Elect.*, vol. 53, pp. 593–600, Jul. 2007.

[20] E. Y. Lam, "Combining gray world and retinex theory for automatic white balance in digital photography," in *Proc. 9th Int. Symp. Consum. Electron.*, Sep. 2005, pp. 134–139.

[21] S. M. Pizer, R. E. Johnston, J. P. Ericksen, B. C. Yankaskas, and K. E. Muller, "Contrast-limited adaptive histogram equalization: Speed and effectiveness," in *Proc. 1st Conf. Visualizat. Biomed. Comput.*, May 1990, vol. 1, pp. 337–345.

[22] D. J. Jobson, Z. Rahman, and G. A. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Trans. Image Process.*, vol. 6, no. 7, pp. 965–976, Jul. 1997.

[23] Y.-C. Liu, W.-H. Chan, and Y.-Q. Chen, "Automatic white balance for digital still camera," *IEEE Trans. Consum. Electron.*, vol. 41, no. 3, pp. 460–466, Aug. 1995.

[24] J. van de Weijer, T. Gevers, and A. Gijsenij, "Edge-based color constancy," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2207–2214, Sep. 2007.

[25] D. H. Foster, "Color constancy," *Vis. Res.*, vol. 51, pp. 674–700, Aug. 2011.

[26] H. Yang, P. Chen, C. Huang, Y. Zhuang, and Y. Shiau, "Low complexity underwater image enhancement based on dark channel prior," in *Proc. 2nd Int. Conf. Innov. Bio-inspired Comput. Appl.*, Jan. 2011, pp. 17–20.

[27] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.

[28] R. Zhang, P. Isola, and A. A. Efros, "Colorful image colorization," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2016, pp. 649–666.

[29] Z. Cheng, Q. Yang, and B. Sheng, "Deep colorization," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 415–423.

[30] H. Zhang and V. M. Patel, "Densely connected pyramid dehazing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Dec. 2018, pp. 3194–3203.

[31] X. Ye, H. Xu, X. Ji, and R. Xu, "Underwater image enhancement using stacked generative adversarial networks," in *Proc. Pacific Rim Conf. Multimedia*, Sep. 2018, pp. 514–524.

[32] Y. Guo, H. Li, and P. Zhuang, "Underwater image enhancement using a multiscale dense generative adversarial network," *IEEE J. Ocean Eng.*, vol. 45, no. 3, pp. 862–870, Jul. 2020.

[33] R. Dey and F. M. Salem, "Gate-variants of gated recurrent unit (gru) neural networks," in *Proc. IEEE 60th Int. Midwest Symp. Circuits Syst.*, Aug. 2017, pp. 1597–1600.

[34] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2016, pp. 694–711.

[35] J. S. Jaffe, "Computer modeling and the design of optimal underwater imaging systems," *IEEE J. Ocean Eng.*, vol. 15, no. 2, pp. 101–111, Apr. 1990.

[36] 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.

[37] M. J. Islam *et al.*, "Semantic segmentation of underwater imagery: Dataset and benchmark," 2020, *arXiv:2004.01241*.