USLN: A statistically guided lightweight network for underwater image enhancement via dual-statistic white balance and multi-color space stretch

Ziyuan Xiao, Yina Han*, Susanto Rahardja, and Yuanliang Ma

Abstract—Underwater images are inevitably affected by color distortion and reduced contrast. Traditional statistic-based methods such as white balance and histogram stretching attempted to adjust the imbalance of color channels and narrow distribution of intensities a priori thus with limited performance. Recently, deep-learning-based methods have achieved encouraging results. However, the involved complicate architecture and high computational costs may hinder their deployment in practical constrained platforms. Inspired by above works, we propose a statistically guided lightweight underwater image enhancement network (USLN). Concretely, we first develop a dual-statistic white balance module which can learn to use both average and maximum of images to compensate the color distortion for each specific pixel. Then this is followed by a multi-color space stretch module to adjust the histogram distribution in RGB, HSI, and Lab color spaces adaptively. Extensive experiments show that, with the guidance of statistics, USLN significantly reduces the required network capacity (from existing $10^6$ to no more than $10^3$) and achieves state-of-the-art performance. The code and relevant resources are available at https://github.com/deepxzy/USLN.

Index Terms—underwater image enhancement, white balance, histogram stretching, neural network, deep learning.

I. INTRODUCTION

Underwater images are of great significance for the utilization and research of underwater resources [1], [2]. However, underwater images have always been suffering from quality degradation. Specifically, since the red light travels the shortest in the water, the underwater images are always dominated by blue and green. Besides, the light is scattered by underwater micro suspending particles, which causes low contrast in underwater images. These problems seriously limit the ability to obtain the meaningful information that is not conducive to further processing of underwater images [1]. Therefore, many algorithms [3]–[9] are proposed to solve these issues.

Researchers early focused on the degradation causes and pixel values of underwater images, proposing two targeted traditional methods, physical-model-based methods [5], [10]–[12] and statistic-based methods [3], [13]–[15]. Physical-model-based methods try to enhance images by reversing the underwater imaging physical model. Since the physical model is ill-posed, additional constraints must be introduced to estimate the relevant optical parameters. However, these artificial constraints do not always hold in changeable underwater environment, which affects the robustness of these methods. [16]. In contrast, statistic-based methods mainly aim to improve the image quality directly by adjusting image pixel values a priori. For instance, the white balance-based methods [13], [17], [18] balance the image by multiplying the color channels by different handcrafted factors, while the histogram stretching-based methods [14], [15], [19] improve the contrast by stretching the color histogram of image from a narrower range to a wider range. These methods can improve images quality in color balance and contrast correction to a great extent. However, the white balance-based methods multiply all pixel values in each color channel by the same factor, which ignores the different distortion degree of each pixel. For histogram stretching-based methods, because each color channel is stretched to a predefined range separately, these methods ignore the correlation between channels and introduce artificial enhancement effect inevitably.

Recently, deep-learning-based methods [6]–[9], [20] have been successfully used in this field. Early deep-learning-based methods reference existing noted backbones, which can
directly learn the mapping relationship between raw image and clear image. However, as pointed by [8], these models ignore the domain knowledge of underwater image enhancement, which results in limited performance. To address this issue, Li et al. [6], [8] took the results of traditional algorithms as additional inputs to enhance the response of networks towards specific regions. These methods exploit the advantages of traditional and deep-learning-based methods, which improve enhancement performance successfully. Nevertheless, existing deep-learning-based models [6]–[9], [20] take numerous convolutional layers to extract high-dimensional image features. Thousands of parameters are involved to fit the underwater image enhancement mapping relationship, which makes these models complicate and time-consuming.

Based on above methods, we take a further step in this paper. We proposes a statistically guided lightweight network via dual-statistic white balance and multi-color space stretch. Specifically, dual-statistic white balance module embeds average and maximum of images to help model learn to compensate the color distortion for each specific pixel, while multi-color space stretch module adaptively increases the contrast of images by adjusting histogram distribution in different color spaces (RGB, HSI, Lab). As Fig. 1 presented, compared with existing deep-learning-based methods, our model takes the least parameters (894) but achieves the best performance. The main contributions of this paper are highlighted as follows.

- We propose two trainable modules, dual-statistic white balance module and multi-color space stretch module to remove color casts and improve contrast, respectively.
- Our model follows the idea of statistic-based methods, that is modifying the pixel values in three-dimensional color space directly, rather than processing high-dimensional image features as current deep-learning-based methods do.
- Our model takes the least parameters, but exceeds the SOTA models on several datasets (UIEB, U imaginet, UFO).

II. RELATED WORK

Underwater enhancement methods can be roughly classified into three categories: physical-model-based methods, statistic-based methods and deep-learning-based methods.

Physical-model-based methods. Physical-model-based methods mainly focus on introducing visual priors to estimate the underwater optical parameters. Then, the clean images can be obtained by reversing the underwater imaging physical model. These priors include red channel prior [10], underwater dark channel prior [11], minimum information prior [12], underwater light attenuation prior [5], and so on. Among them, underwater dark channel prior method [11] and red channel prior method [10] both make improvement on the DCP algorithm [21] for the underwater environment. UDCP only considers the blue and green channels, while ARUIR considers reverse red channel instead of red channel.

Although these physical-model-based methods can achieve promising performance in some cases, these methods are sensitive to the types of underwater images, which means these priors are not always plausible in changeable and complicated underwater environment [16]. Compared with these methods, our model can easily adapt to various underwater environments, due to the powerful learning capability of neural networks.

Statistic-based methods. Statistic-based methods improve the images quality by adjusting pixel values. For example, Iqbal et al. proposed an unsupervised color correction method (UCM) [3]. UCM firstly uses white balance to equalize the RGB channels to make image balanced. Then the histogram equalization is applied in the RGB and HSI color spaces respectively to increase the contrast of the images. On this basis, Ahmad et al. [14] [15] proposed an adaptive histogram enhancement method base on limited Rayleigh distribution stretching, which can enhance details and reduce over-enhanced. These methods have obvious advantages for increasing the brightness and contrast of underwater images.

These statistic-based methods need set many hyperparameters manually in advance. This always introduces artificial color. In contrast, our method integrates statistic-based methods into deep-learning-based methods, which automatically learns these hyperparameters according to different underwater environment.

Deep-learning-based methods. Recently, deep-learning-based methods do impressive performance on underwater image enhancement task. These methods can be divided into convolutional neural networks (CNN) [22] and generative adversarial networks (GAN) [23]. For CNN-based models, Sun et al. [20] first proposed a pixel-to-pixel deep learning model based on encode-decoder framework, which has significant effect on underwater image denoising. Furthermore, Li et al. [7] proposed a light-weight CNNs model (UWCNN) based on underwater scene prior, which can synthesize image datasets with different degrees of degradation. Most recently, inspired by physical-model-based methods, Li et al. [8] proposed a model via medium transmission-guided multi-color space embedding (Ucolor), which uses reverse medium transmission map to emphasize quality-degraded regions. With regard to GAN-based model, Islam et al. [9] proposed FUnIE-GAN, which is inspired by U-net [24], using five encoder-decoder pairs with skip-connection to build generator. The processing speed of this model is fast and results have a good effect on color recovery and sharpening.

These deep-learning-based models need too much parameters to fit the mapping relationship between clear image and raw image. Even lightweight networks like UWCNN [7] and Shallow-UWnet [25] still contain thousands of parameters. In contrast, instead of extracting high-dimensional image features, USLN follows the idea of statistic-based algorithms, that is, to improve the quality of images by adjusting the pixel values in three color channels directly, thus greatly reducing the amount of parameters (merely takes 894 parameters).

III. PROPOSED METHOD

We present the overview architecture of USLN in Fig. 2. USLN consists of two main modules, dual-statistic white balance module and multi-color space stretch module. In dual-statistic white balance module, an underwater image is first
Fed into two trainable balance modules which are based on average and maximum of image, respectively. After getting the balanced images, there are also two residual-enhancement modules to recover details of images. Then, we merge the two enhanced images together and feed the final result to multi-color space stretch module. In multi-color space stretch module, the image first goes through color space transformation where the RGB image is converted into HSI and Lab color spaces. We then get stretched images by feeding these multi-color spaces images into trainable stretch module, respectively. Simultaneously, in each color space, the original image is also forwarded to residual-enhancement module and pixel-wise add up to the stretched image. Finally, these enhanced images are all converted back to RGB color space and merge together. Note that in HSI color space, the H channel is split out and keep unchanged.

In what follows, we detail all the components of our method, including dual-statistic white balance module (Sec. III-A), multi-color space stretch module (Sec. III-B), residual-enhancement module (Sec. III-C), and the loss function (Sec. III-D). In addition, we also discuss the superiority of our model from view of visual attention at the end of this section (Sec. III-E).

### A. Dual-statistic Balance Module

In the underwater environment, images are rarely color balanced. In general, there will be a phenomenon of higher blue and green light intensity and lower red light intensity. Hence, the first step in our model is to make image balanced. Inspired by white balance algorithm, we propose dual-statistic white balance module.

---

![Fig. 2. Overview of the architecture of USLN. Our model consists of dual-statistic white balance module, multi-color space stretch module and several residual-enhancement modules. The input of USLN is three-dimensional underwater image in which the pixel values is between 0 and 1. 'convolutional layer' has the kernel of size $3 \times 3$ and stride 1, which is used to merge enhanced images together. A detailed network module structure is shown in follow subsections.](image)

| Module | Dual-statistic White Balance Module | Multi-color Space Stretching Module | Residual-Enhancement Module | Total |
|--------|-------------------------------------|-------------------------------------|-----------------------------|-------|
| Parameters | 192 | 282 | 420 | 894 |

![Fig. 3. The schematic illustration of dual-statistic white balance module, which consists of two trainable balance modules, the right part is based on Gray World theory, while the left part is based on White Patch algorithm. GMP and GAP are abbreviations of global max pooling and global average pooling, respectively.](image)

![Fig. 4. The schematic illustration of pointwise convolution. We use pointwise convolution to generalize Von Kries hypothesis, which indicates not only the relationship between color channels but also correction of each channel.](image)
The structure of dual-statistic white balance module is shown in Fig. 3. This module consists of two trainable balance modules, one is based on Gray World theory, while the other is based on White Patch algorithm. Next, we will explain the detailed derivation process of this module.

Von Kries hypothesis can be used to express the relationship between underwater images and ground-truth images:

\[
\begin{bmatrix}
I'_{R} \\
I'_{G} \\
I'_{B}
\end{bmatrix} = \begin{bmatrix}
e_R & 0 & 0 \\
0 & e_G & 0 \\
0 & 0 & e_B
\end{bmatrix} \begin{bmatrix}
I_R \\
I_G \\
I_B
\end{bmatrix}
\]  

(1)

Where \( I \) denotes pixels values of underwater images, while \( I' \) denotes those in ground-truth images, and \( e_R, e_G, e_B \) denote scale factors that are applied independently to the three color channels.

However, Von Kries hypothesis considers the R, G, B color channels independently. More generally, we need to consider the relationship between color channels and correction of each channel, so as Fig. 4 shown that we can use pointwise convolution to generalize Von Kries hypothesis as:

\[
I' = \text{Conv}_{1 \times 1}(I)
\]  

(2)

Where \( \text{Conv}_{1 \times 1} \) means the convolution operation, subscript \( i \) means the size of convolutional kernel.

However, it is difficult to learn the relationship between underwater images and enhanced images only by pointwise convolution (Eq. (2)). Inspired by two white balance algorithms, Gray World and White Patch, we add average and maximum of images into Eq. (2) to help model learn this relationship more easily.

The Gray World theory assumes that the average value of color object in a perfect image is gray, which means the average pixels of the three channels R, G, and B tend to the same value. Hence, scale factors of each channel, \( e_R, e_G, e_B \) can be set based on GW theory:

\[
A_c = \frac{1}{N} \sum_{i=1}^{N} I_c(x), c \in \{R, G, B\}
\]  

(3)

\[
e_R = \frac{0.5}{A_R}, e_G = \frac{0.5}{A_G}, e_B = \frac{0.5}{A_B}
\]  

(4)

Where \( A_c \) denotes the average value of \( c \) channel in original image.

Therefore, inspired by Eq. (4), we add average of each color channel in Eq. (2), which can be expressed:

\[
I^G_W = \text{Conv}_{1 \times 1}(I) \odot \mathbf{A}
\]  

(5)

where \( \mathbf{A} = [\frac{1}{A_R}, \frac{1}{A_G}, \frac{1}{A_B}] \in \mathbb{R}^{3 \times 1} \), \( \odot \) denotes pixel-wise multiplication, \( I^G_W \) means image enhanced by the trainable white balance module based on Gray World algorithm. Gray World algorithm can be considered a special case of Eq. (5).

White Patch algorithm assumes the maximum response of the RGB color channel is caused by the white patch in the scene. Theoretically, the white patch can reflect the color of the scene light, so the largest value in the RGB channel will be used as the light source. According to this hypothesis, scale factors of each channel can be expressed as:

\[
M_c = \max I_c(x), c \in \{R, G, B\}
\]  

(6)

\[
e_R = \frac{1}{M_R}, e_G = \frac{1}{M_G}, e_B = \frac{1}{M_B}
\]  

(7)

Where \( M_c \) denotes the maximum value of \( c \) channel in original image.

Same as before, inspired by Eq. (7), we add maximum of each color channel in Eq. (2), which can be expressed:

\[
I^{WP} = \text{Conv}_{1 \times 1}(I) \odot \mathbf{M}
\]  

(8)

where \( \mathbf{M} = [\frac{1}{M_R}, \frac{1}{M_G}, \frac{1}{M_B}] \in \mathbb{R}^{3 \times 1}, I^{WP} \) means image enhanced by the trainable white balance module based on White Patch algorithm. White Patch algorithm can be considered a special case of Eq. (8).

Finally, two enhanced image are merged together by going through 3×3 convolutional layer and pixel-wise adding together.

\[
I_{\text{final}} = \text{Conv}_{3 \times 3}(I^G_W) + \text{Conv}_{3 \times 3}(I^{WP})
\]  

(9)

where \( I_{\text{final}} \) means the final result enhanced by dual-statistic white balance module. \( \oplus \) means pixel-wise addition.

B. Multi-color Space Stretching Module

Due to the low histogram range, underwater images often suffer from low contrast and visibility. Therefore, histogram stretching is adopted to provide a wider pixel distribution of the image channels to improve the image contrast. The following equation is used for linear contrast stretching:

\[
P_{\text{out}} = \frac{P_{\text{in}} - i_{\text{min}}}{i_{\text{max}} - i_{\text{min}}}(o_{\text{max}} - o_{\text{min}}) + o_{\text{min}}
\]  

(10)

Where the \( P_{\text{in}} \) and \( P_{\text{out}} \) are the input and output pixels intensity values, respectively. \( i_{\text{max}} \) and \( i_{\text{min}} \) represent the maximum and minimum intensity value of the input image, respectively. \( o_{\text{max}} \) and \( o_{\text{min}} \) represent the maximum and minimum intensity value of the desired output image, respectively.

In histogram stretching-based methods, \( i_{\text{max}} \) and \( i_{\text{min}} \) are easy to get, but \( o_{\text{max}} \) and \( o_{\text{min}} \) must be defined artificially. For example, Kashif Iqbal [3] used global histogram stretching to enhance the contrast of the image, setting \( o_{\text{max}}, o_{\text{min}} \) to 1, 0. However, such a simple setup may not achieve good results. We can use pointwise convolution to make Eq. (10) trainable by next steps:

\[
\tilde{I}_c(x) = \frac{I_c(x) - \min I_c(x)}{\max I_c(x) - \min I_c(x)}, c \in \{R, G, B\}
\]  

(11)

\[
I' = \text{Conv}_{1 \times 1}(\tilde{I})
\]  

(12)

Where \( \tilde{I} = [\tilde{I}_R, \tilde{I}_G, \tilde{I}_B] \in \mathbb{R}^{3 \times H \times W}, I' \) denotes the histogram stretched pixel value in RGB color space. Global histogram stretching can be considered a special case of Eq. (12).

Fig. 5 presents the details of the trainable stretch module in RGB space.

In addition to operating in the RGB color space, inspired by traditional algorithms [3], [13], [26] and Ucolor [8], we also stretch images in HSI and Lab spaces. HSI color space
reflects the way the human visual system perceives color, which intuitively reflects the hue, saturation and intensity. And Lab color space is designed to approach human perceptual vision, which contains almost every shade of every color in nature. As Fig. 6 shown that the same image has different nature. As Fig. 6 shown that the same image has different
color distribution in various color spaces. Thus, the multiple
color spaces stretch can facilitate the adjustment of underwater
color images in these color spaces is not linear [27], which enables
the deep-learning model to learn more complex mapping
relationships. We also design ablation study to verify the
contribution of multi-color space stretch operation.

Concretely, we convert the image into HSI and Lab color
spaces and then stretch the S, I channels and L, a, b channels,
respectively. Note that we do not stretch H channel, because our
ablation studies verify that changing the value of the H channel will damage the color of the image. Interestingly, this
conclusion is consistent with the statistic-based methods like
[3], [13] and [28].

At last, three stretched images are converted back to RGB
color space and merged together by going through 3x3 convo-
colutionary layer and pixel-wise add up.

The kernel size of convolutional layer is set as 3x3, because
the main convolutional layers in dual-statistic white balance
module and multi-color space stretch module are pointwise
convolution, which ignores the relationship between adjacent
pixels. Therefore, the 3x3 convolutional layers in residual-
enhancement module can alleviate this problem. For activation
function, we chose Tanh activation function, because the
output of Tanh activation is between -1 and 1, which means that residual-enhancement module can choose to reduce or
increase the intensity of the enhanced pixels based on the
features.

D. Loss Function

Inspired by Hang Zhao’s work [30] and Johnson’s work
[31], we used combined loss to train our model, which consists of
MAE loss, SSIM loss and VGG perceptual loss.

MAE loss is to calculate the mean of the absolute difference
between all pixels of the output image, \( \hat{y} \), and the label, \( y \), it can be expressed as:

\[
L_{MAE} = \frac{1}{W \times H} \sum_{m=1}^{W} \sum_{n=1}^{H} |\hat{y}(m, n) - y(m, n)|
\]  

SSIM Loss is to improve the structural similarity between
the feature representations of the output image, \( \hat{y} \), and the
label, \( y \):

\[
L_{SSIM} = 1 - \frac{(2u_y \sigma_y + \mu_1)(2\sigma_{\hat{y}y} + \mu_2)}{(u_y^2 + \sigma_y^2 + \mu_1)(\sigma_{\hat{y}y}^2 + \sigma_{\hat{y}y}^2 + \mu_2)}
\]  

where \( u \) and \( \sigma \) denote the mean, standard deviation, and
covariance of images. \( \mu_1 \) and \( \mu_2 \) are used to stabilize
the division.

VGG perceptual loss is based on pretrained VGG networks.
Let \( \phi(\cdot) \) be the last convolutional layer of VGG-19 networks
[27]. The perceptual loss is calculated the MAE loss of the
feature representations of output image, \( \hat{y} \), and the label, \( y \),
assuming that the total number of pixels is \( M \):

\[
L_{VGG} = \frac{1}{M} \sum_{m=1}^{W} \sum_{n=1}^{H} |\phi(\hat{y})(m, n) - \phi(y)(m, n)|
\]

The final combined loss is the linear combination of the
MAE loss, SSIM loss and VGG perceptual loss:

\[
L_{final} = L_{MAE} + \lambda_1 L_{SSIM} + \lambda_2 L_{VGG}
\]

where \( \lambda_1 \) and \( \lambda_2 \) are empirically set to 0.25 and 1, respectively,
for balancing the scales of different losses.

The validity of our combined loss function will be demonstrated in later ablation experiments.

E. Perspective of Attention Mechanism

From perspective of attention mechanism, we think that
our model can be seen as a special case of channelwise
attention networks. Classic channelwise attention networks, SE
block [32] can be roughly divided into three stages: squeeze,
excitation and scale. SE block first squeezes global spatial
information into a channel descriptor, then the global information is embedded by channelwise multiplying the excited
channel descriptor and feature map together. In our model, the
squeezed channel descriptors include average, maximum, and
IV. EXPERIMENTS

In this section, we first describe the datasets and evaluation metrics, then introduce our implementation details. We compare our model with the current underwater image enhancement models and provide ablation experiments to verify the contribution of each module. Cases of failure are presented at the end of this section.

A. Benchmarks

In our work, we adopted four benchmarks for evaluation, Uimagine [9], UIEB [6], UFO [33], and Color-Check7 [34]. The scale of these datasets is various, so that we can compare these methods comprehensively.

In UIEB [6], Uimagine [9], and UFO [33], there are many paired images and unpaired images which were divided as shown in Tab. II.

Color-Check7 [34] contains seven original underwater images with Color Checker. These images are taken by seven different cameras, Canon D10, Fuji Z33, Olympus Tough 6000, Olympus Tough 8000, Pentax W60, Pentax W80, and Panasonic TS1, denoted as D10, Z33, T6000, T8000, W60, W80, TS1. Considering there is no corresponding train set in Color-Check7, we got three results from models trained on Train-800, Train-3145 and Train-1500, respectively. Finally, the best score is used to evaluate the performance.

B. Evaluation Metrics

For test sets in UIEB, Uimagine and UFO, we chose three full-reference evaluation metrics and one no-reference evaluation metric. Concretely, 1) The full-reference evaluation metrics consist of Peak Signal-to-Noise Ratio (PSNR) [35], Structural Similarity (SSIM) [35] and Mean Squared Error (MSE), which are used for paired test sets (Test-90, Test-555, Test-120). A higher PSNR or a lower MSE score means that the output image and the label image are closer in content, while a higher SSIM score means that the two images are more structurally similar. 2) The no-reference evaluation metric is UIQM [36], which is used for unpaired test sets (Test-60, Test-1270). A higher UIQM score suggests a better human visual perception. UIQM is the linear combination of UICM, UISM, and UIConM, which reflect the degree of color deviation, brightness, and contrast, respectively. However, it is worth mentioning that these no-reference metrics cannot accurately reflect the quality of an image in some cases [6] [37], so scores of UIQM (UICM, UISM, UIConM) are only provided as references for our study. We will present enhanced unpaired images in Visual Comparisons section for readers to score.

For Color-Check7, each colorcheck in Color-Check7 consists of 24 color patches. We used CIEDE2000 [38] to evaluate the enhanced results. Concretely, we cropped the 24 color patches for each enhanced result and measured the average value of each color patch, then calculated the perceptual differences between corresponding color patches of ground-truth Color Checker and those of enhanced results.

C. Implementation Details

We used python and pytorch framework via NVIDIA RTX3080 on windows 10. Adam was used for network optimization, while the initial learning rate was set as 0.01 with 5% reduction per epoch. We trained our model for 100 epochs and the batch size was set as 10.

Different from fixed size (256×256) of images in Uimagine, the size of images in UIEB is unfixed. In order to use batch learning, all images in UIEB were resized to 256×256. For
modules. 2) Our model shows excellent performance on small images enhanced by FUnIE-GAN and Ucolor are greenish and bluish, respectively. In comparison, our model recovers the reddish, respectively. Note that although FUnIE-GAN can be trained with unpaired data, we also trained this model in paired train sets to ensure the fairness of the experiment.

In addition to the quality of the enhancement results, for deep-learning-based models, the amount of parameters are also important evaluation metrics. As shown in Table III, our model takes the least parameters and GFLOPs. Compared with the second-lightest model, the amount of parameters and GLOPs of our model are only 0.22% of UWCNN (amount of parameters) and 7.89% of FUnIE-GAN (GFLOPs).

**D. Compared Methods**

We selected eight works to make fair comparisons with our model, including two statistic-based methods (UCM [3], RGHS [4]), two physical-model-based methods (Li et al. [12], ULAP [5]), four deep-learning-based methods (Water- net [6], Ucolor [8], UWCNN [7], FUnIE-GAN [9]). UCM integrates White Balance and histogram stretch to improve image quality, which is highly similar to our model. The comparison with UCM aims to demonstrate the advantages of our trainable modules.

For these traditional unsupervised methods, we used them directly on the test sets. For these deep-learning-based methods, we used the code and training strategy provided by the authors. Note that although FUnIE-GAN can be trained with unpaired data, we also trained this model in paired train sets to ensure the fairness of the experiment.

**E. Quantitative Comparisons**

We first make a quantitative comparison on Test-90, Test-555, and Test-120. The average scores of MSE, PSNR, and SSIM of different methods are shown in Table IV. As shown in Table IV, although there are only few parameters in our model, our method still performs better than the complex deep-learning-based models. Our model achieves the lowest MSE score and highest PSNR, SSIM scores in all three datasets. In addition, we also found that 1) although our method principle is roughly the same as UCM model, which is based on white balance and histogram stretch, our model performs much better than UCM, which suggests the superiority of our trainable modules. 2) Our model shows excellent performance on small datasets UIFB. Compared with the second-best performer, our USLN achieves the percentage gain of 16%, 8% and 5% in terms of MSE, PSNR and SSIM, respectively. This means that our model can learn the mapping of underwater images and enhanced images well with small dataset.

Then, for Color-Checker7, we show the average CIEDE2000 scores of different methods in Table V. Our model achieves highest color similarity in Z33, T6000 and T8000. Moreover, our model achieve the lowest average CIEDE2000 score (10.71 under CIEDE2000 metric). These results show that our model is robust to different cameras and can restore the real color of underwater images with high accuracy. Interestingly, many compared methods get worse average CIEDE2000 score than raw images, which suggests that these methods even destroy the images color.

Next, we use UIQM as evaluation metric for unpaired test set (Test-60, Test-1270). The average scores of different methods are shown in Table V. As Table VI shown, Our model gets highest UICON in Test-90, and second best UICM and UICON scores in Test-1270. There are two interesting findings from the Table V. 1) Different from full-reference evaluation metrics, deep-learning-based models have not achieve much better results than traditional methods in no-reference evaluation metrics. Conversely, for UICM, Li et al and UCM rank the best and second best respectively in both datasets. 2) Although FUnIE-GAN and Ucolor perform best in Test-90 and Test-1270, respectively, they do not perform well in another dataset. In comparison, our model achieves robust performance in both two test sets.

**F. Visual Comparisons**

In this section, we conduct several visual comparisons on testing datasets.

We first present raw images and enhanced images in paired datasets in Fig. 9. Compared with other SOTA models, our result is closest to the ground-truth images and achieves the highest PSNR and SSIM scores. Note that, as table.1 shown, traditional methods get much lower PSNR/SSIM scores than deep-learning-model, so due to length limitation, we only present images enhanced by deep-learning-based models in Fig.7.

The underwater images suffer from different color deviations are shown in Fig. 9(a). The first two images (left) are only distorted slightly, while the other images are more distorted. For slightly distorted images, all compared methods fail to restore the complete scene structure, these methods all get even lower PSNR or SSIM than raw images. In comparison, our model improves the quality of these images effectively without over-enhancement. For the rest images which are deviated by greenish and blueish color. The compared models either fail to remove the greenish and blueish color completely or introduce undesirable color artifacts. For example, these images enhanced by FUnIE-GAN and Ucolor are greenish and reddish, respectively. In comparison, our model recovers the color of image and achieves the best scores.

To demonstrate the robustness of our model, we then present the enhanced Color Checker images of different methods.
raw image shown in Fig. 8 is taken by Panasonic TS1, which suffers from severe red color cast. All compared methods cannot achieve satisfactory results, some of them even reduce the quality of image. According to CIEDE2000 metric, our method gets the best score (9.90) and restores the most accurate colors of the raw image.

We then show raw images and enhanced images in unpaired datasets, Test-60 and Test-1270 in Fig. 10, Fig. 11, and Fig. 12. Noted that we only use UIQM as the reference to evaluate these images, the accurate perceptual quality of these enhanced images need the readers to score.

For Fig. 10, the raw image sampled from Test-1270 suffers from serious blueish deviation. Our result gets the best UIQM. Compared with SOTA models, our model successes to recover the natural color and structural details.

For Fig. 11 and Fig. 12 which are sampled from Test-60, these two images are much more challenging. As Fig. 11 shown, the obvious light limitation of the raw image results in low contrast. The compared models tend to make the image darker, some models even introduce reddish color. In comparison, our model increases both brightness and contrast, making the details of the image clear. Fig. 12(a) suffers from obvious yellow deviation, all compared methods cannot achieve satisfactory results. For example, although Fig. 12(f) which is enhanced by Waternet achieves the best UIQM, it is obvious that the left part of the image is over-enhanced, introducing unwanted blue, while the right part is under-enhanced, failing to remove yellow completely. In comparison, our model removes the yellowish color and makes the color balanced.

G. Ablation

We perform extensive ablation studies to analyze the advantage of the core modules, including residual-enhancement module (REM), multi-color space stretch module (MCSM), and dual-statistic white balance module (DSBM). In addition, we analyze the combined loss in our model. More specifically,

- w/o Lab+SI, w/o Lab, and w/o SI mean that the multi-color space stretch module without both Lab and SI color channels, Lab channels, and SI channels, respectively. w/ H means that we do not separate H channel out in multi-color space stretch module.
- w/o avg and w/o max, and w/o max+avg stand for dual-statistic white balance module without average, maximum, and both average and maximum, respectively.
- w/o RES refers to USLN without residual-enhancement module, while w/o Tanh refers to the Residual-Enhancement Module without Tanh activation function.
TABLE VI
THE COLOR DISSIMILARITY COMPARISONS OF DIFFERENT METHODS ON COLOR-CHECK7 IN TERMS OF THE CIEDE2000. WE REPRESENT THE BEST TWO RESULTS IN RED AND BLUE COLORS.

| Models       | D10 | Z33 | T6000 | T8000 | TS1 | W60 | W80 | Avg |
|--------------|-----|-----|-------|-------|-----|-----|-----|-----|
| raw          | 11.62 | 16.00 | 13.37 | 16.44 | 14.46 | 10.84 | 9.80 | 13.27 |
| UCM [3]      | 12.23 | **10.87** | 12.32 | **11.79** | 12.75 | 11.99 | 8.04 | **11.43** |
| RGHS [4]     | 13.68 | 12.91 | 14.20 | 13.84 | 11.64 | 13.66 | 11.08 | **13.00** |
| UAP [5]      | 20.98 | 17.06 | 20.36 | 15.56 | 24.47 | 20.09 | 20.53 | **19.86** |
| Li et al. [12] | 17.51 | 24.70 | 16.71 | 16.94 | **11.47** | 13.48 | 13.93 | **16.39** |
| Ucolor [8]   | **10.73** | 12.90 | **11.30** | 12.25 | 14.34 | **10.26** | 10.79 | **11.79** |
| Waternet [6] | 12.18 | 15.25 | 11.66 | 15.74 | 19.89 | 12.14 | 11.31 | **14.02** |
| FUnIE-GAN [9] | 10.29 | 14.48 | 12.67 | 13.81 | 15.91 | **10.11** | 9.51 | **12.40** |
| UWCNN [7]    | 12.97 | 13.29 | 14.12 | 14.33 | 12.99 | 13.29 | 10.66 | **13.09** |
| USLN         | **10.88** | **10.69** | **10.59** | 12.89 | 9.90 | **11.00** | **9.03** | **10.71** |

Fig. 8. Visual comparisons on an underwater image taken by a Panasonic TS1 camera sampled from Color-Check7. Our model removes color cast and improves contrast effectively. Compared with the existing methods, the image produced by our USLN has the best score in terms of the CIEDE2000.

- w/o VGG+SSIM, w/o SSIM, and w/o VGG mean that USLN is trained with MAE loss, MAE+SSIM loss, and MAE+VGG loss, respectively.

We used MSE, PSNR, and SSIM to evaluate the results on Test-90, which are shown in Table VI. Compared to the ablation models, full model achieves the best quantitative performance on all metrics. The visual comparisons of the contributions of dual-statistic white balance module, multi-color stretch model, and residual-enhancement model are presented in Fig. 13, Fig. 14, and Fig. 15, respectively. The conclusions drawn from the ablation study for the each module are listed as follows:

1) The ablated models w/o avg and w/o max produce comparable performance as shown in Table VII and Fig. 13. Dual-statistic balance module help model to balance the images base on two representative values, average and maximum, while removing either of them will degrade performance as Table VII shown. As Fig. 13 shown, the full model removes the color casts much more completely than ablated models.

2) As Table VII shown, the lack of stretch in either HSI or Lab space will decrease performance. In addition, as Fig. 14 shown, the full model enhances images from three color spaces, which is effective to restore the natural color of image. Interestingly, we find blindly add all channels in HSI color space will reduce the performance, while adding only S and I channels can improve performance effectively.

3) Residual-enhancement module can preserve the image fidelity, making the enhanced images look more natural. As presented in Fig. 15, w/o Res model over-enhances the image, making aquatic weeds be greyish, while full model restores normal green color of aquatic weeds. Moreover, the lack of Tanh activation function will degrade the performance as Table VII presented.

4) Ablation study towards the combined loss is presented in Table VII. SSIM loss in the combined loss function can improve the structural similarity of the output image as shown in Table VII, while VGG perceptual loss can improve the visual quality of output. By adding perceptual loss and SSIM loss to the MAE loss, the visual quality of final result is improved.
Fig. 9. Visual comparisons on challenging underwater images sampled from Test-90 and Test-555. The number on the top-right corner of each image refers to its PSNR/SSIM (the larger, the better).

H. Failure Case

Enhancing the underwater image with severe color cast is very challenging. Both our model and other SOTA models...
fail to work well. As Fig. 16 shown, although our model and Ucolor remove the yellow cast, it is far from achieving the desired visual quality. Ucolor makes image much darker, while USLN introduces unwanted red. The main reason is the lack of similar images in train sets, limiting the ability of deep-learning-based methods to deal with such images.

V. CONCLUSION

In this paper, we introduce a statistically guided lightweight networks for underwater image enhancement via dual-statistic white balance and multi-color space stretch. Our model introduces several statistic-based methods into networks, including Gray World, White Patch and histogram stretching. Our model aims to enhance images by adjusting the pixel values directly, instead of processing high-dimensional image features. As a result, our model takes only 894 trainable parameters, but provides state-of-the-art performance in diverse representative datasets(UIEB, EUVP, UFO). We also conduct extensive ablation experiments to verify the effectiveness of the main components.

ACKNOWLEDGMENTS

The authors would like to thank Chongyi Li, Chunle Guo, Wenqi Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao for providing the UIEB dataset. They would also like to thank Md Jahidul Islam, Youya Xia, Peigen Luo and Junaed Sattar for providing EUVP and UFO datasets. Finally, they would also like to thank Codruta O. Ancuti, Cosmin Ancuti, Christophe De Vleeschouwer, and Philippe Bekaert for providing the underwater Color Checker images.
**TABLE VII**

| Modules  | Baselines | MSE ↓ | PSNR ↑ | SSIM ↑ |
|----------|-----------|-------|--------|--------|
| full model | 479 | 23.78 | 0.92 |
| w/o Lab+SI | 540 | 23.38 | 0.91 |
| w/o Lab | 505 | 23.51 | 0.91 |
| w/o SI | 494 | 23.57 | 0.92 |
| w/ HSI | 486 | 23.48 | 0.91 |
| w/o max | 556 | 22.98 | 0.91 |
| w/o avg | 679 | 22.33 | 0.90 |
| w/o max+avg | 763 | 21.20 | 0.90 |
| w/o RES | 611 | 22.97 | 0.91 |
| w/o Tanh | 506 | 23.51 | 0.91 |
| w/o VGG+SSIM | 491 | 23.38 | 0.91 |
| w/o VGG | 470 | 23.73 | 0.92 |
| w/o SSIM | 493 | 23.77 | 0.91 |

**REFERENCES**

[1] W. Zhang, L. Dong, T. Zhang, and W. Xu, “Enhancing underwater image via color correction and bi-interval contrast enhancement,” Signal Processing: Image Communication, vol. 90, p. 116030, 2021.

[2] F. Liu and M. Fang, “Semantic segmentation of underwater images based on improved deepLab,” Journal of Marine Science and Engineering, vol. 8, no. 3, p. 188, 2020.

[3] K. Iqbal, M. Odetayo, A. James, R. A. Salam, and A. Z. H. Talib, “Enhancing the low quality images using unsupervised colour correction method,” in 2010 IEEE International Conference on Systems, Man and Cybernetics. IEEE, 2010, pp. 1703–1709.

[4] D. Huang, Y. Wang, W. Song, J. Sequeira, and S. Mavromatis, “Shallow-water image enhancement using relative global histogram stretching based on adaptive parameter acquisition,” in International conference on multimedia modeling. Springer, 2018, pp. 453–465.

[5] W. Song, Y. Wang, D. Huang, and D. Tjondronegoro, “A rapid scene depth estimation model based on underwater light attenuation prior for underwater image restoration,” in Pacific Rim Conference on Multimedia. Springer, 2018, pp. 676–688.

[6] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao, “An underwater image enhancement benchmark dataset and beyond,” IEEE Transactions on Image Processing, vol. 29, pp. 4376–4389, 2019.

[7] C. Li, S. Anwar, and F. Porikli, “Underwater scene prior inspired deep underwater image and video enhancement,” Pattern Recognition, vol. 98, p. 107038, 2020.

[8] C. Li, S. Anwar, J. Hou, R. Cong, C. Guo, and W. Ren, “Underwater image enhancement via transmission-guided multi-color space embedding,” IEEE Transactions on Image Processing, vol. 30, pp. 4985–5000, 2021.

[9] M. J. Islam, Y. Xia, and J. Sattar, “Fast underwater image enhancement for improved visual perception,” IEEE Robotics and Automation Letters, vol. 5, no. 2, pp. 3227–3234, 2020.

[10] A. Galdran, D. Pardo, A. Picón, and A. Alvarez-Gila, “Automatic red-channel underwater image restoration,” Journal of Visual Communication and Image Representation, vol. 26, pp. 132–145, 2015.

[11] P. Drews, E. Nascimento, F. Moraes, S. Botelho, and M. Campos, “Transmission estimation in underwater single images,” in Proceedings of the IEEE international conference on computer vision workshops, 2013, pp. 825–830.
Fig. 14. Ablation study of the contributions of multi-color space stretch module. The full model which enhances image in three colors can restore the inherent color. The number on the top-right corner of each image refers to its PSNR/SSIM (the larger, the better).

Fig. 15. Ablation study of the contributions of residual-enhancement module. The residual-enhancement module can preserve the image fidelity, making the enhanced images look more natural. The number on the top-left corner of each image refers to its PSNR/SSIM (the larger, the better).

Fig. 16. Failure case. The input underwater image suffers from severe yellow color cast. USLN removes the yellow color but introduces unwanted red on the sands.

[12] C.-Y. Li, J.-C. Guo, R.-M. Cong, Y.-W. Pang, and B. Wang, “Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior,” *IEEE Transactions on Image Processing*, vol. 25, no. 12, pp. 5664–5677, 2016.
[13] K. Iqbal, R. A. Salam, A. Osman, and A. Z. Talib, “Underwater image enhancement using an integrated color model,” *IAENG International Journal of Computer Science*, vol. 34, no. 2, 2007.
[14] A. S. A. Ghani and N. A. M. Isa, “Enhancement of low quality underwater image through integrated global and local contrast correction,” *Applied Soft Computing*, vol. 37, pp. 332–344, 2015.
[15] ———, “Automatic system for improving underwater image contrast and color through recursive adaptive histogram modification,” *Computers and electronics in agriculture*, vol. 141, pp. 181–195, 2017.
[16] R. Liu, X. Fan, M. Zhu, M. Hou, and Z. Luo, “Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 12, pp. 4861–4875, 2020.
[17] X. Li, G. Hou, L. Tan, and W. Liu, “A hybrid framework for underwater image enhancement,” *IEEE Access*, vol. 8, pp. 197448–197462, 2020.
[18] G. D. Finlayson and E. Trezzi, “Shades of gray and colour constancy,” in *Color and Imaging Conference*, vol. 2004, no. 1. Society for Imaging Science and Technology, 2004, pp. 37–41.
[19] K. Zuiderveld, “Contrast limited adaptive histogram equalization,” *Graphics gems*, pp. 474–485, 1994.
[20] X. Sun, L. Liu, Q. Li, J. Dong, E. Lima, and R. Yin, “Deep pixel-to-pixel network for underwater image enhancement and restoration,” *IJT Image Processing*, vol. 13, no. 3, pp. 469–474, 2019.
[21] K. He, J. Sun, and X. Tang, “Single image haze removal using dark channel prior,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 12, pp. 2341–2353, 2010.
[22] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
[23] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” *Advances in neural information processing systems*, vol. 27, 2014.
[24] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
[25] A. Naik, A. Swarnakar, and K. Mittal, “Shallow-uwnet: Compressed model for underwater image enhancement (student abstract),” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 18, 2021, pp. 15 853–15 854.
[26] S. K. Naik and C. Murthy, “Hue-preserving color image enhancement without gamut problem,” *IEEE Transactions on image processing*, vol. 12, no. 12, pp. 1591–1598, 2003.
[27] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
[28] A. S. Abdul Ghani and N. A. Mat Isa, “Underwater image quality enhancement through composition of dual-intensity images and rayleigh-stretching,” *SpringerPlus*, vol. 3, no. 1, pp. 1–14, 2014.
[29] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
[30] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, “Loss functions for image restoration with neural networks,” *IEEE Transactions on computational imaging*, vol. 3, no. 1, pp. 47–57, 2016.
[31] J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” in *European conference on computer vision*. Springer, 2016, pp. 694–711.
[32] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.
[33] M. J. Islam, P. Luo, and J. Sattar, “Simultaneous enhancement and super-resolution of underwater imagery for improved visual perception,” *arXiv preprint arXiv:2002.01155*, 2020.
[34] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and P. Bekaert, “Color balance and fusion for underwater image enhancement,” *IEEE Transactions on image processing*, vol. 27, no. 1, pp. 379–393, 2017.
[35] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.
[36] K. Panetta, C. Gao, and S. Agaian, “Human-visual-system-inspired underwater image quality measures,” *IEEE Journal of Oceanic Engineering*, vol. 41, no. 3, pp. 541–551, 2016.
[37] D. Berman, D. Levy, S. Avidan, and T. Treibitz, “Underwater single image color restoration using haze-lines and a new quantitative dataset,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 8, pp. 2822–2837, 2020.
[38] G. Sharma, W. Wu, and E. N. Dalal, “The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations,” *Color Research & Application: Endorsed by International Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur*, vol. 30, no. 1, pp. 21–30, 2005.