IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 23 May 2023, accepted 26 June 2023, date of publication 3 July 2023, date of current version 19 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3291449

RESEARCH ARTICLE

SCAUIE-Net: Underwater Image Enhancement Method Based on Spatial and Channel Attention

YUANHAO ZHONG[1], JI WANG[1,2], AND QINGJIE LU[1]
[1]School of Electronics and Information Engineering, Guangdong Ocean University, Zhanjiang, Guangdong 524088, China
[2]Guangdong Smart Ocean Sensor Network and its Equipment Engineering Technology Research Center, Zhanjiang, Guangdong 524088, China
Corresponding authors: Qingjie Lu (1123493362@qq.com) and Ji Wang (gdouwangii@163.com)

ABSTRACT Underwater image enhancement is a Low-Level Vision task that plays an important role in marine resource development, but the light absorption and scattering cause severe underwater image quality degradation. To solve these problems, this paper proposes a neural network based on a spatial and channel attention module that reinforces the network's attention to channel and spatial information. The network's Confidence Generator can precisely extract feature maps from multi-scale underwater images. Meanwhile, we propose a new training loss function by mixing perceptual, MS-SSIM and MAE loss functions to further improve the contrast in high-frequency, colors and luminance. For training, this paper also uses a feature fusion strategy: Firstly, augmenting the training underwater images by Gamma Correction, White Balance and Histogram Equalization algorithms to remove color cast, lighten up dark regions and improve the contrast. Then, fusing the enhancing images with confidence maps predicted from the Generator. The network was validated in the UIEB dataset and obtains efficient improvements on Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) metrics, yielding a PSNR of 22.9286 and SSIM of 0.9290. Experimental results on real-world underwater images demonstrate that the proposed method performs well on different underwater scenes.

INDEX TERMS Underwater image enhancement, low-level vision, attention mechanism.

I. INTRODUCTION

In recent years, underwater image enhancement plays an important role in underwater resource exploration, aquatic robotics inspection [1] and underwater archaeology [2]. Although underwater images are good for marine resource development, there are still urgent issues to be solved, such as image distortion caused by light absorption, and image blur caused by scattering including forward scattering and backward scattering [3]. Furthermore, underwater light attenuation also rises some underwater image problems, such as low contrast, color casts, low visibility and blurred details. These issues significantly reduce the efficiency of marine resource development. Therefore, it's vital to improve the visual quality, contrast and color properties of underwater images to accurately excavate the underwater world.

There have been several approaches to enhance the visual quality of underwater images. Mainly underwater enhancement methods are traditional visual enhancement algorithms [21], [22], [23], [24], physical-based methods [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35] and Deep Learning-based methods [20], [36], [37], [38], [39], [40], [41]. Traditional visual enhancement algorithms mainly concentrate on modifying underwater images' pixel values to adapt contrast, saturation and brightness. However, the lack of physical degradation process causes the inability to achieve better enhancement quality. In addition, physical-based methods use channel prior [26], [27], [30], [32] to accurately estimate the medium transmission [28], [29], [35]. Through the estimation of medium transmission and some other important physical parameters such as homogeneous background light, a physical underwater formation model can be constructed. Putting raw underwater images into the physical underwater formation model could reverse clean images.

The associate editor coordinating the review of this manuscript and approving it for publication was Shadi Alawneh.

Although the physical-based methods could perform well in specific underwater environments, it is limited in different geographical and temporal underwater environments. The changes in physical parameters would restrict the model's performance. In the past few years, the approaches based on deep learning obtain tremendous progress which make great effects on image enhancement and dehazing. In the underwater image enhancement domain, the effectiveness of Deep Learning-based methods depends on the model's capability and the existing underwater image datasets. Some image enhancement and dehazing CNNs perform well in ground scenes but obtain bad results in underwater scenes due to the degradation process between underwater and the ground being different. CNNs model constructed for ground scenes could not fit underwater scenes. Also, the lack of a high-quality real-world underwater image dataset, such as a small number of images, not enough underwater scenes and not enough real-world scenarios underwater images restricts the performance of Deep-Learning based method. A large-scale real-world underwater image dataset makes CNNs trained easily to improve underwater results. However, it is practically impossible to photograph a real underwater scene and the corresponding ground truth image for different underwater scenarios at the same time.

To solve the above disadvantages of Deep Learning-based enhancement methods, this paper proposes a network based on Spatial and Channel Attention for Underwater Image Enhancement, termed SCAUIE-Net. SCAUIE-Net is a gated fusion framework including both Confidence Map Generator and Image Refiner parts. The two parts are as follows: Confidence Map Generator uses U-Net [4] architecture as the backbone for feature extracting and confidence map prediction. Image Refiner uses simple convolutional layers to denoise, remove color cast and adjust underwater image's saturation and brightness. Although the U-Net [4] is already performing very well in extracting feature information, there are still exists some issues such as insufficient texture details and local color cast. Inspired by CBAM [5] and SK-NET [6], we applied Spatial Attention Module and Selective Kernel Block to fully extract underwater image context information so that the underwater images could be enhanced more effectively.

To make further improvements in visual quality, we focus on the loss function used to train a neural network for underwater image enhancement. Due to the human visual system being more sensitive to luminance and color variations in texture-less regions, SCAUIE-Net uses a loss function mixed with Perceptual Loss [7], MS-SSIM Loss [8] and MAE Loss. These loss functions are more compliant with the human vision system. Mix Loss could maintain high-frequency regions, colors and luminance information. Meanwhile, Perceptual loss [7] in Mix Loss measures the similarity of images matching with the Human Visual System. It can express image details well.

SCAUIE-Net is trained on a large-scale real-world underwater image enhancement benchmark (i.e., UIEB [9]) dataset which contains 950 real underwater images from different light sources, such as natural light, artificial light or a mixture of natural light and artificial light. Compared to various baseline networks, SCAUIE-Net obtains visual quality improvement in the UIEB dataset.

This paper introduces the following main contributions:

1. We use a gated fusion framework trained by the UIEB dataset for the underwater image enhancement task including Confidence Map Generator and Image Refiner. Confidence Map Generator based on U-Net predicts the enhancing confidence maps. Image Refiner fuses the raw images and enhanced images to remove color cast.

2. We use Selective Kernel Block and Spatial Attention Module for underwater image enhancement to improve the model's representation capability. Spatial Attention Module decides 'where' is more informative to focus on. Selective Kernel Convolution has a dynamic selection mechanism to adaptively adjust the local receptive field sizes of neurons.

3. We combine MS-SSIM loss, perceptual loss, MAE loss as mixed loss function. MS-SSIM loss and MAE loss are used to maintain high-frequency regions, colors and luminance information. Perceptual loss measures the similarity of images matching with the Human Visual System. It can express image details well.

II. RELATED WORK

A. UNDERWATER IMAGE DATASET

Deep Learning based underwater enhancement methods need to be heavily data-driven. The datasets can be divided into two categories: real-world underwater image datasets and synthetic underwater image datasets. Unlike in-air image enhancement tasks, complicated underwater environments (e.g., turbidity and lighting conditions) are hard to synthesize plenty of realistic underwater images for deep learning. Consequently, synthesizing underwater images becomes a challenge. There are some ways to synthesize underwater images including GAN-based method, underwater image formation model. GAN-based method could obtain paired images, Wang et al. [10] proposed an unsupervised GAN-based method called UWGAN to synthesize underwater images from in-air RGB-D images and depth maps pairs. More recently, Zhao et al. [11] used an image-to-image framework for underwater image synthesis and depth map estimation in underwater conditions which eliminates the challenge to convert a single underwater image into an underwater depth map due to the lack of paired data. On the other hand, using an underwater image formation model has gradually attracted attention recently. Blasinski et al. [12] provided a three-parameter underwater image formation model [13]. Anwar et al. [14] incorporate a new underwater image synthesis method that simulates 10 different categories of underwater images using NYU-v2 indoor dataset [15]. Despite the similarity, there are still gaps

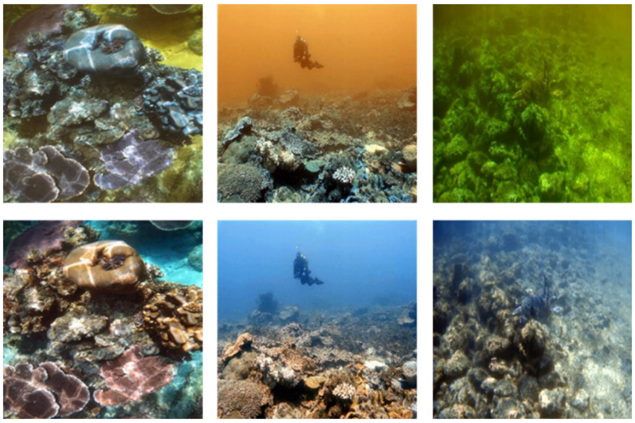

FIGURE 1. Sampling of real-world underwater image enhancement by SCAUIE-Net. Top row: raw underwater images taken in diverse underwater scenes; Bottom row: the corresponding reference results.

between synthetic and real-world underwater images. Real-world underwater datasets are Fish4Knowlege dataset for underwater target detection and recognition, SUN dataset for scene recognition and object detection [16], MARIS dataset for marine autonomous robotics, Sea-thru dataset including 1100 underwater images with range maps [17], Haze-line dataset providing raw images, TIF files, camera calibration files, and distance maps [18], Liu et al. [19] proposed the RUIE dataset, which encompasses varied underwater lighting, depth of filed blurriness and color cast scenes and Peng et al. [20] proposed the LSUI dataset including 4279 real-world underwater images with more abundant scenes. However, most of the existing real-world datasets still have some problems, such as limited scenes, few degradation characteristics, insufficient data and no corresponding ground truth images. Therefore, we use the UIEB [9] dataset, a dataset including 950 real-world images. To overcome issues of scene monotony and insufficient data, 890 images of them have the corresponding ground truth images. These potential reference images are produced by 12 enhancement methods and voted by 50 volunteers to select the final references.

B. UNDERWATER IMAGE ENHANCEMENT METHOD

There are various effective underwater image enhancement solutions, such as traditional visual enhancement algorithms, physical model-based and Deep Learning-based methods.

1) TRADITIONAL VISUAL ENHANCEMENT ALGORITHMS

In the earlier stage, traditional visual enhancement algorithms perform well in underwater image enhancement. This branch of underwater image enhancement methods concentrates on operating enhancement algorithms in color spaces. Hitam et al. [21] proposed a method called mixture Contrast Limited Adaptive Histogram Equalization colors models, which operates CLAHE [52] on RGB and HSV color models. Ma et al. [22] proposed a fusion algorithm in different color spaces based on CLAHE. This algorithm converts RGB color space to two different color spaces YIQ and HIS and operates CLAHE in both of them. Then the algorithm converts two color spaces back to RGB color space to fuse out the enhanced image. Abdul et al. [23] enhance underwater images through dual-intensity images and Rayleigh-stretching. The underwater image is applied with modified Von Kreis hypothesis and stretched into two different intensity images at the average value with respects to Rayleigh distribution. Then, the image is applied with color correction in HSV color model to obtain the enhanced result. Huang et al. [24] proposed a relative global histogram stretching algorithm, which is mainly based on the equalization of G-B channels and histogram stretching in the RGB color model.

2) PHYSICAL MODEL-BASED METHODS

Physical-based methods are accurately estimating the medium transmission and some other important physical parameters such as attenuation and diffusion coefficients. The steps of physical model-based methods can be explained as follow: 1) constructing a physical model with underwater prior conditions; 2) estimating the key parameters; 3) putting raw underwater images and reversing the degradation process to obtain a clean image.

The essence of the physical model-based methods is to establish an underwater image formation model with prior knowledge, like Dark Channel Prior (DCP) [25]. Chiang et al. [26] used DCP combined with the wavelength-dependent compensation algorithm to restore underwater images. In [27], an Underwater Dark Channel Prior (UDCP) was proposed based on the fact that the information of the red channel in an underwater image is undependable. Based on the observation that the dark channel of the underwater image tends to be a zero map, Liu and Chau [28] formulated a cost function and minimized it to find the optimal transmission map, which can maximize the image contrast. Instead of the DCP, Li et al. [29] employed the random forest regression model to estimate the transmission of the underwater scenes. Peng et al. [30] proposed a Generalized Dark Channel Prior (GDCP) for image restoration, which incorporates adaptive color correction into an image formation model. Carlevaris-Bianco et al. [31] proposed a prior that exploits the difference in attenuation among three color channels in RGB color space to predict the transmission of an underwater scene. Galdran et al. [32] proposed a Red Channel method, which recovers the lost contrast of an underwater image by restoring the colors associated with short wavelengths. Li et al. [33], [34] proposed an underwater image enhancement method based on the minimum information loss principle and histogram distribution prior. Peng et al. [35] proposed a depth estimation method for underwater scenes based on image blurriness and light absorption, which is employed to enhance underwater images. The physical-based methods could perform well in specific underwater environments. However, there are some disadvantages to these methods. For example, the degradation process is estimated inaccurately when the underwater scenes change and physical-model based methods do not take the human visual system into account. Therefore, deep learning-based methods attract researchers' attention in recent years.

3) DEEP LEARNING-BASED METHODS

With the development of arithmetic power and the expansion of data volume, deep learning-based methods are used to improve the quality of underwater images. A variety of methods based on deep learning can be divided into two main categories, which are GAN-based methods and CNN-based methods.

There are GAN-based methods. Li et al. [36] proposed WaterGAN which first simulates underwater images from the in-air image and depth pairings in an unsupervised pipeline. The network includes a two-stage network for color cast removal. Li et al. [37] proposed UWCNNs trained by ten types of underwater images, where the underwater images are synthesized from a revised underwater image formation model [38] and the corresponding underwater scene parameters. More recently, Li et al. [39] proposed a weakly supervised underwater color transfer model called Water CycleGAN which is based on CycleConsistent Adversarial Networks [40]. Although the GAN-based methods have made great progress in underwater image enhancement, they still cannot solve the disadvantages of unstable outputs from GANs. As for CNN-based methods, Wang et al. [41] proposed a CNN model with two color spaces RGB and HSV called UIEC^2. UIEC^2 uses RGB color space for denoising and removing color cast and HSV color space for globally adjusting underwater image luminance, color and saturation. Li et al. [59] proposed Ucolor that uses medium transmission-guided multi-color space embedding to solve wavelength and distance-dependent attenuation and scattering. More recently, a transformer architecture [20] called U-shape transformer was proposed. U-shape transformer was designed with two transformer modules to reinforce the network's attention to color channels and space areas and combines RGB, LAB and LCH color spaces as loss function. In summary, many CNN-based methods improve the quality of underwater images by increasing the network's feature extraction capability. However, there are still some shortcomings in CNN-based methods discussed above. CNN models rely on large number of high-quality datasets but such datasets are difficult to obtain. Also, the structure of these CNN models is often too complex resulting in a large model size. Therefore, in this paper, we derive the inputs with multiple preprocessing operations for data augmentation and we use a simpler model structure to keep the model lightweight.

C. UNDERWATER IMAGE ENHANCEMENT USING ATTENTION MECHANISM

In previous studies, attention mechanism [42] has been widely used in various low-level vision tasks, such as image restoration [43], image super resolution [44], image segmentation [45] and image enhancement [46]. It biases the allocation of the most informative feature expression and simultaneously suppresses the less useful ones. From the above researches, we could see that attention mechanism has the advantages of feature representation and visual compensation, which are important for the low-level tasks. Because of these advantages, attention has been applied to many underwater image enhancement tasks. Li et al. [47] proposed UDA-Net using unsupervised attention mechanism for region-wise underwater image enhancement, Wang et al. [48] used class-condition attention, in which an underwater image is classified first and then the class label guides the generating of enhanced images and Fu et al. [49], used residual two-fold attention, in which non-local attention and channel attention are embedding to extract and enhance features. Furthermore, Qi et al. [50] proposed SGUIE-Net using semantic information as high-level guidance across different images that share common semantic regions. Given the above introduction, using attention mechanism shows better performance in underwater image enhancement both synthetic and real-world. Therefore, in this paper, we use spatial and channel attention in underwater image enhancement to achieve higher-quality images. Inspired by Selective Kernel Network [6] and Convolutional Block Attention Module [5], we combine both and apply them in U-Net [4]. Our Selective Kernel Block and Spatial Attention Module can guide the network to pay more attention to the more serious attenuated color channels and spatial areas. Unlike other models, our network could adaptively adjust receptive filed sizes of spatial areas and color channel weights.

III. PROPOSED NETWORK

In this section, we discuss the details of the proposed CNN-based underwater image enhancement model using spatial and channel attention, called SCAUIE-Net. Firstly, we introduce the input generation, a preprocessing module for underwater images. Then we depict the network architecture including Selective Kernel Block and Spatial Attention Module. Finally, we introduce the loss function used in SCAUIE-Net.

A. INPUT GENERATION

To fulfill lighting conditions and complex underwater scenes needed for the training data, we derive the inputs with multiple preprocessing operations. Inspired by the degradation process of underwater images, we generate three inputs by respectively applying White Balance [51], Histogram Equalization [52] and Gamma Correction algorithms [53]. Then we use the fusion strategy of blending color features to obtain decent results. We directly apply the white balancing technique proposed in [51] which minimizes this effect of color casts for the entire scene. We employ Histogram Equalization on Lab color space for

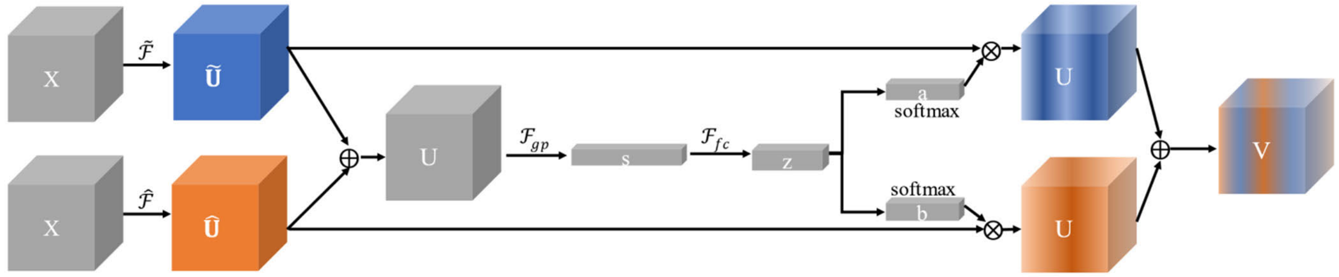

FIGURE 2. Selective Kernel Convolution. ⊕ denotes element-wise summation and ⊗ denotes element-wise product.

improving the contrast and lightening up dark regions. In the Gamma Correction algorithm, we set the Gamma value to 0.7 empirically.

B. SELECTIVE KERNEL CONVOLUTION

Receptive field sizes play an important role in image color perception. For underwater images, the enhancement effect could be benefited from adaptively adjusting receptive field sizes. Therefore, we use an automatic selection operation, "Selective Kernel Convolution", among multiple kernels with different kernel sizes. Specifically, we implement the SK convolution via three operators – Split, Fuse and Select, as illustrated in Fig. 2, where a two-branch case is shown. Therefore, in this example, there are only two kernels with different kernel sizes, but it is easy to extend to multiple branches case.

1) SPLIT

For any given feature map $\mathbf{X} \in R^{H' \times W' \times C'}$, by default, we first conduct two transformations $\tilde{\mathcal{F}} : \mathbf{X} \rightarrow \tilde{\mathbf{U}} \in R^{H \times W \times C}$ and $\hat{\mathcal{F}} : \mathbf{X} \rightarrow \hat{\mathbf{U}} \in R^{H \times W \times C}$ with kernel sizes 3 and 5, respectively. Note that both $\tilde{\mathcal{F}}$ and $\hat{\mathcal{F}}$ are composed of efficient grouped/depthwise convolutions, Batch Normalization [53] and ReLU [54] function in sequence. For further efficiency, the conventional convolution with a 5 × 5 kernel is replaced with the dilated convolution with a 3 × 3 kernel and dilation size 2.

2) FUSE

As stated in the Introduction, our goal is to enable neurons to adaptively adjust their receptive field sizes according to the stimulus content. The basic idea is to use gates to control the information multiple branches carrying different scales of information into neurons in the next layer. To achieve this goal, the gates need to integrate information from all branches. We first fuse results from multiple (two in Fig.2) branches via an element-wise summation:

$$\mathbf{U} = \tilde{\mathbf{U}} + \hat{\mathbf{U}} \tag{1}$$

then we embed the global information by simply using global average pooling to generate channel-wise statistics as $s \in R^{C}$. Specifically, the c-th element of s is calculated by shrinking $\mathbf{U}$ through spatial dimensions $H \times W$:

$$s_c = \mathrm{F}_{gp}\left(\mathbf{U}_c\right) = \frac{1}{H \times W} \sum_{\mathrm{i}=1}^{\mathrm{H}} \sum_{\mathrm{j}=1}^{\mathrm{W}} \mathrm{U_c(i, j)}. \tag{2}$$

Further, a compact feature $z \in R^{d \times 1}$ is created to enable the guidance for the precise and adaptive selections. This is achieved by a simple fully connected (fc) layer, with the reduction of dimensionality for better efficiency:

$$z = \mathrm{F}_{fc}(s) = \delta(\mathrm{B}(\mathbf{W}s)), \tag{3}$$

where δ is the ReLU function, B denotes the Batch Normalization, $\mathbf{W} \in R^{d \times C}$. To study the impact of d on the efficiency of the model, we use a reduction ratio r to control its value:

$$d = \max(C/r, L), \tag{4}$$

where L denotes the minimal value of d ($L = 32$ is a typical setting in our experiments).

3) SELECT

A soft attention across channels is used to adaptively select different spatial scales of information, which is guided by the compact feature descriptor z. Specifically, a softmax operator is applied on the channel-wise digits:

$$a_c = \frac{e^{\mathbf{A}_c z}}{e^{\mathbf{A}_c z} + e^{\mathbf{B}_c z}}, \quad b_c = \frac{e^{\mathbf{B}_c z}}{e^{\mathbf{A}_c z} + e^{\mathbf{B}_c z}} \tag{5}$$

where $\mathbf{A}, \mathbf{B} \in R^{C \times d}$ and a, b denote the soft attention vector for $\tilde{\mathbf{U}}$ and $\hat{\mathbf{U}}$, respectively. Note that $\mathbf{A}_c \in R^{1 \times d}$ is the c-th row of $\mathbf{A}$ and a_c is the c-th element of $\mathbf{a}$, likewise $\mathbf{B}_c$ and b_c. In the case of two branches, the matrix $\mathbf{B}$ is redundant because $a_c + b_c = 1$. The final feature map $\mathbf{V}$ is obtained through the attention weights on various kernels:

$$\mathbf{V}_c = a_c \cdot \tilde{\mathbf{U}}_c + b_c \cdot \hat{\mathbf{U}}_c, a_c + b_c = 1, \tag{6}$$

where $\mathbf{V} = [\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_{\mathrm{c}}], \mathbf{V}_{\mathrm{c}} \in R^{H \times W}$. Note that here we provide a formula for two-branch cases and on can easily deduce situations with more branches by extending equals (1) (5) (6).

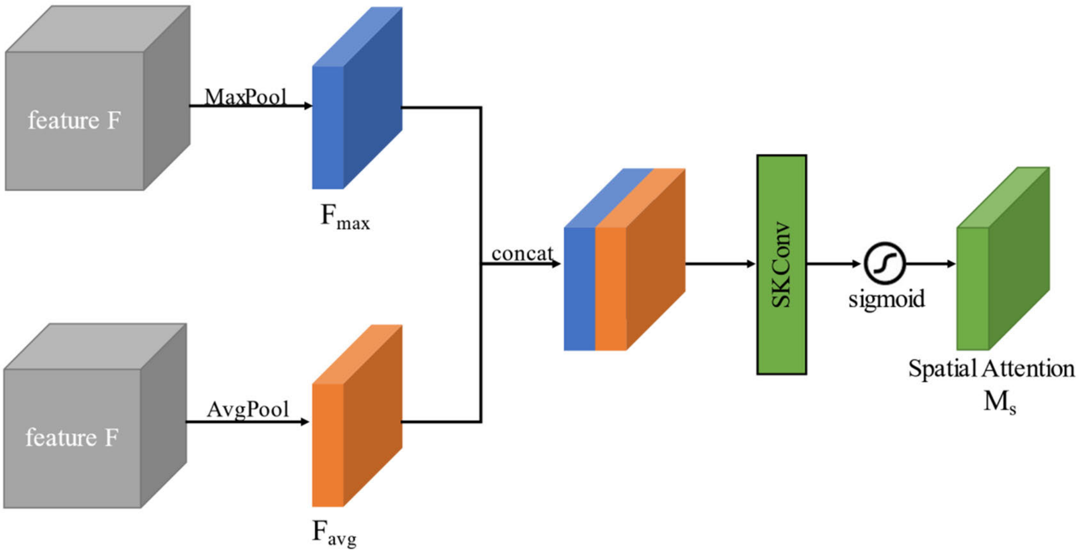

FIGURE 3. Spatial Attention Module. The spatial attention module uses max-pooling and average-pooling along the channel axis to obtain outputs. Then, the spatial attention module forwards the outputs to a selective kernel convolution.

C. SPATIAL ATTENTION MODULE

We generate a spatial attention map by utilizing the inter-spatial relationship of features. Different from channel attention, spatial attention focuses on 'where' is an informative part, which is complementary to channel attention. To compute the spatial attention, we first apply average-pooling and max-pooling operations along the channel axis and concatenate them to generate an efficient feature descriptor. Applying pooling operations along the channel axis is shown to be effective in highlighting informative regions. On the concatenated feature descriptor, we apply a selective kernel convolution layer to generate a spatial attention map $\mathbf{M_s}(\mathbf{F}) \in \mathbf{R}^{H\times W}$ which encodes were to emphasize or suppress. Selective kernel convolution employs adaptive channel weights, which could select the weights of the concatenated feature descriptor to improve the channel representation capability of the Spatial Attention Module. The Spatial Attention Module is illustrated in Fig. 3. We describe the detailed operation below.

We aggregate channel information of a feature map by using two pooling operations, generating two 2D maps: $\mathbf{F}^{\mathbf{s}}_{avg} \in R^{1\times H\times W}$ and $\mathbf{F}^{\mathbf{s}}_{max} \in R^{1\times H\times W}$. Each denotes average pooled features and max-pooled features across the channel. Those are then concatenated and convolved by a standard convolution layer, producing our 2D spatial attention map. In short, the spatial attention is computed as:

$$\begin{aligned}\mathbf{M_s}(\mathbf{F}) &= \sigma\left(f^{7\times 7}([\mathrm{AvgPool}(\mathbf{F}); \mathrm{MaxPool}(\mathbf{F})])\right),\\ &= \sigma\left(f^{7\times 7}\left(\left[\mathbf{F}^{\mathbf{s}}_{\mathrm{avg}}; \mathbf{F}^{\mathbf{s}}_{\mathrm{max}}\right]\right)\right)\end{aligned} \tag{7}$$

where σ denotes the sigmoid function and $f^{7\times 7}$ represents a selective kernel convolution operation with the filter size of 7×7.

D. NETWORK ARCHITECTURE

1) OVERALL ARCHITECTURE

SCAUIE-Net is illustrated in Fig.4. It's a gated fusion network to learn three confidence maps that indicate the most significant features of inputs respectively. Then, the inputs are fused with the confidence maps to get the fused images. The sum of the fused images is the enhanced result.

The architecture of the proposed SCAUIE-Net consists of two parts: Image Refiner and Confidence Map Generator. The components used in SCAUIE-Net are Selective Block and Spatial Attention Module. Image Refiner is a plain fully CNN. Confidence Map Generator uses U-Net [4] as the backbone. To reduce the color casts and artifacts brought by the White Balance [51], Histogram Equalization [52] and Gamma Correct algorithms, we add three Image Refiners and feed the three derived inputs and original input to the Image Refiner. Then, we separately feed the refined inputs to the Confidence Map Generator to predict confidence maps. At last, the refined three inputs are multiplied by the three learned confidence maps to achieve the final enhanced result:

$$I_{en} = R_{WB} \odot C_{WB} + R_{HE} \odot C_{HE} + R_C \odot C_{GC} \tag{8}$$

where I_{en} is the enhanced result; $\odot$ indicates the element-wise production of matrices; $\mathrm{R_{WB}}$, $\mathrm{R_{HE}}$ and $\mathrm{R_{GC}}$ are the refined results of input after processing by White Balance [51], Histogram Equalization [52] and Gamma Correct algorithms; C_{WB}, C_{HE} and $\mathrm{C_{GC}}$ are the learned confidence maps.

2) IMAGE REFINER

Image Refiner is a shallow CNN. It consists of Selective Kernel Convolution and 2D convolutional layers, each followed by a ReLU [54]. At the first layer, 1 × 1 convolution

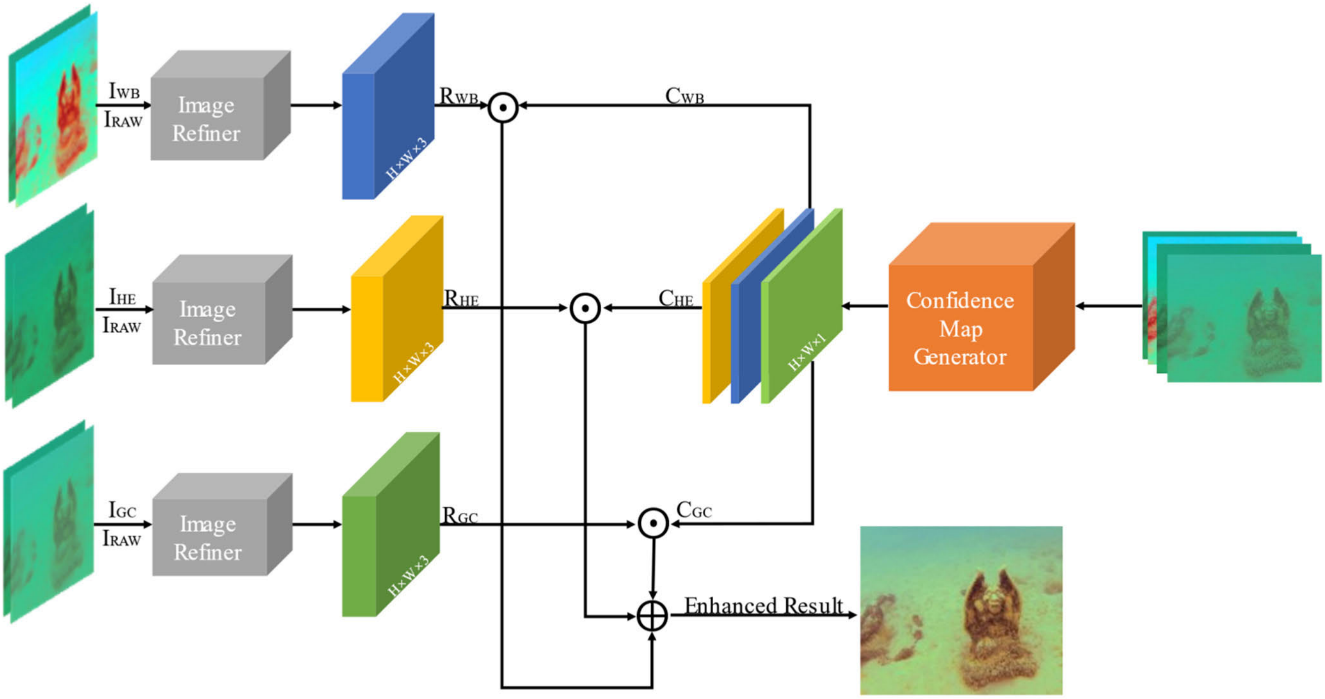

FIGURE 4. The network structure of SCAUIE-Net. SCAUIE-Net consists of a Confidence Map Generator for predicting the confidence maps and three Image Refiners for removing color cast.

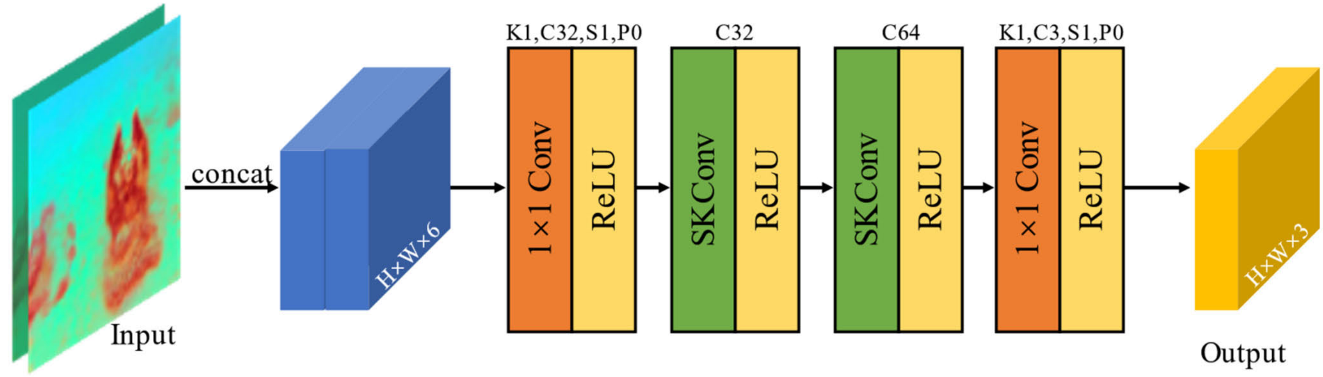

FIGURE 5. The structure of Image Refiner that is a shallow CNN.

is used to increase the number of feature channels to 32. Then, Selective Kernel Convolution is applied twice to focus on the channel information. the second Selective Kernel Convolution layer will double the number of the feature channels. At the final layer, 1×1 convolution is used to reduce dimension mapping 64 feature channels to a refined image with 3 channels.

3) CONFIDENCE MAP GENERATOR

The backbone of Confidence Map Generator is U-Net which performs well in image processing. Similar to U-Net [4], Confidence Map Generator consists of a contracting path and an expanding path. The contracting path consists of the repeated application of two 3×3 convolutions, each followed by a ReLU [54], a 2×2 maxpooling operation with stride 2 for downsampling, a Selective Kernel Block that is a basic residual block [55] (shown in Fig. 7) built by Selective Kernel Convolution and a Spatial Attention Module for extracting spatial attention weights. At each downsampling step, we double the number of feature channels. The expanding path consists of an upsampling of the feature map followed by a 2×2 convolution, a Selective Kernel Block and a Spatial Attention Module. The usage of Selective Kernel Block and Spatial Attention Module is similar to the contracting step. At each expanding step, we halve the number of feature channels.

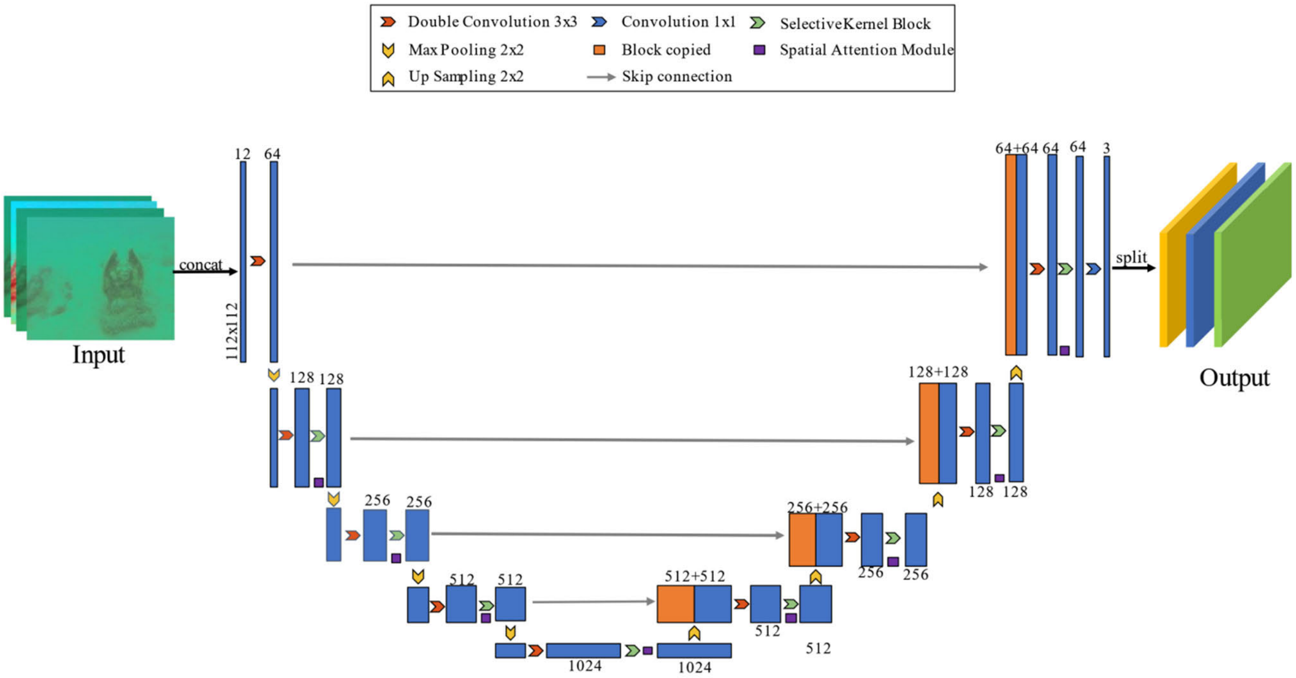

FIGURE 6. The structure of Confidence Map Generator. Each blue block represents a multi-channel feature map. The number of channels is denoted at the top of the box. The symbols in the box above are the operations that we use in the confidence map generator.

E. NETWORK LOSS FUNCTION

The end-to-end training of SCAUIE-Net is supervised by three loss components, which consist of $\mathcal{L}^{MS-SSIM}$, $\mathcal{L}^{\ell_1}$ and $\mathcal{L}^{Perceptual}$.

1) SSIM LOSS

To enhance underwater images from the perspective of luminance, contrast and structure, the error function of perceptually motivated SSIM is effective. SSIM for pixel p is defined as

$$\begin{aligned}\mathrm{SSIM}(p) &= \frac{2\mu_x\mu_y + C_1}{\mu_x^2+\mu_y^2+C_1}\cdot\frac{2\sigma_{\mathrm{xy}}+\mathrm{C}_2}{\sigma_{\mathrm{x}}^2+\sigma_{\mathrm{y}}^2+\mathrm{C}_2}\\ &= l(p)\cdot \mathrm{cs}(p)\end{aligned} \tag{9}$$

where x, y denotes the position of pixel p; μ_x, μ_y and σ_x, σ_y separately denotes the mean and standard deviation of pixel p; σ_{xy} represents the covariance of x, y; C_1, C_2 are small constants used to maintain the stability of $l\,(p)$, $c\,(p)$ and $s\,(p)$. The loss function for SSIM can be then written setting $\varepsilon\,(p) = 1 - SSIM\,(p)$:

$$\mathrm{L}^{SSIM}(P) = \frac{1}{N}\sum_{p\in P} 1 - SSIM(p). \tag{10}$$

2) MS-SSIM LOSS

In practice, the subjective evaluation of a given image varies due to the different factors for different images. The single-scale method SSIM described in 4.5.1 may be appropriate only for specific settings. Multi-scale method is convenient to incorporate image details at different resolutions. Rather than fine-tuning settings, we propose to use the multiscale version of SSIM, MS-SSIM. Given a dyadic pyramid of M levels, MS-SSIM is defined as

$$\text{MS-SSIM}(p) = l_M^{\alpha}(p)\cdot\prod_{j=1}^{M} cs_j^{\beta_j}(p) \tag{11}$$

where l_M and cs_j are the terms that we defined in Selection 4.5.1 at scale M and j, respectively. For convenience, we set $\alpha = \beta_j = 1$, for $j = \{1, \ldots, M\}$. Like Equation (10), the loss function for MS-SSIM can be written as follow:

$$\mathrm{L}^{\mathrm{MS}-SSIM}(P) = 1 - \text{MS-SSIM}(p). \tag{12}$$

3) PERCEPTUAL LOSS

Perceptual loss can produce visually pleasing and realistic results. Inspired by [7] and [58], We define the perceptual loss based on the ReLU activation layers of the pretrained 19 layers VGG network [56]. Due to the deep layer could represent semantic information well and can fully preserve the image content and overall spatial structure, we select layer 5_4 from VGG19 to make it sensitive to semantics. The perceptual loss is expressed as the distance between the feature representations of the enhanced underwater image I_{en} and the reference underwater image I_{gt}:

$$L_j^{\phi} = \frac{1}{C_jH_jW_j}\sum_{i=1}\parallel \phi_{\mathrm{j}}\left(\mathrm{I}_{\mathrm{en}}^{\mathrm{i}}\right) - \phi_{\mathrm{j}}\left(\mathrm{I}_{\mathrm{gt}}^{\mathrm{i}}\right)\parallel, \tag{13}$$

where $\phi_j\,(x)$ denotes the jth convolution layer (after activation) of the VGG19 network pretrained on the ImageNet [57]

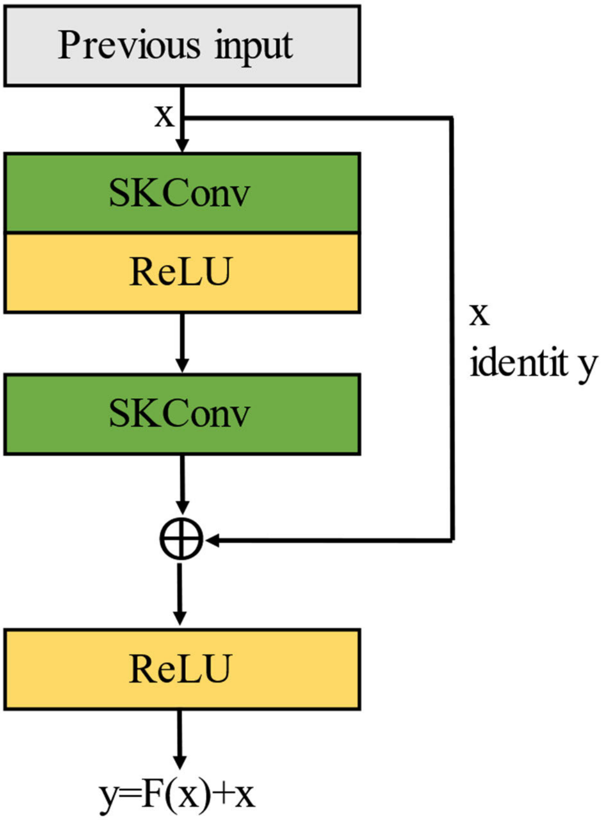

FIGURE 7. Selective Kernel Block. A basic residual block built by Selective Kernel Convolution.

dataset; N is the number of each batch in the training procedure; $C_jH_jW_j$ represents the dimension of the feature maps of the jth convolution layer within the VGG19 network; C_j, H_j, and W_j are the number, height, and width of the feature map.

4) MAE LOSS

Due to the ℓ_2 loss function will cause artifacts, ℓ_1 is applied ℓ_1 instead of ℓ_2. The loss function ℓ_1 is simply defined as follows:

$$\mathrm{L}^{\ell_1}(P) = \frac{1}{N}\sum_{P} |x(p) - y(p)|, \tag{14}$$

where p is the index of the pixel and P is the patch; $x(p)$ and $y(p)$ are the values of the pixels in the processed patch and the ground truth respectively. The derivatives for the back-propagation are also simple, since $\partial \mathcal{L}^{\ell 1}(p)/\partial q = 0, \forall \{q\} \neq$ p. Therefore, for each pixel p in the patch,

$$\partial \mathrm{L}^{\ell 1}(P)/\partial x(p) = sign(x(p) - y(p)). \tag{15}$$

The derivative of $\mathcal{L}^{\ell_1}$ is not defined at 0. Thus, we use the convention that sign $(0) = 0$. The network will not update the weights when $\mathcal{L}^{\ell_1} = 0$.

5) LOSS TERM WEIGHTS

MS-SSIM preserves the contrast in high-frequency regions, ℓ_1 preserves colors and luminance, and Perceptual Loss preserves semantic information. To capture the best characteristics of these functions, we propose to combine them, and each loss term has a weight hyperparameter: α, β, γ:

$$\mathrm{L}^{\mathrm{Mix}} = \alpha \cdot \mathrm{L}^{\text{MS-SSIM}} + \beta \cdot \mathrm{L}^{\ell_1} + \gamma \cdot \mathrm{L}^{\mathrm{Perceptual}}, \tag{16}$$

where we empirically set $\alpha = 2, \beta = 0.000025$ and $\gamma = 0.0025$.

IV. EXPERIMENTS

In this section, we first introduce the training details of the SCAUIE-Net. Then, we train our network model with the UIEB dataset. Furthermore, we perform qualitative and quantitative comparisons with traditional, physical-based, and recent deep-learning-based methods to evaluate our proposed network. These methods include Histogram Equalization [52], GDCP [30], UDCP [27], UWGAN [10], Water-Net [9], Ucolor [59]. Finally, we conduct ablation studies to demonstrate the effectiveness of each component in SCAUIE-Net.

A. IMPLEMENTATION DETAILS

For training, the inputs of our network are real-world underwater images. A random set of 800 pairs of real-world images extracted from the UIEB dataset are used to train our network. We resize the input images to size 112 × 112 due to our limited memory. Flipping and rotation are used to obtain 7 augmented versions of the original training data. For testing, the rest 90 pairs of real-world images are treated as the testing set.

We implemented the proposed SCAUIE-Net with PyTorch on Ubuntu20 with an Nvidia 2080Ti GPU. During training, a batch-mode learning method with a batch size of 16 was applied and the epoch was set to 300. The filter weights of each layer were initialized by standard Gaussian distribution. Bias was initialized as a constant. We trained our model using ADAM and set the learning rate to 0.0001. We used ReduceLROnPlateau as the learning rate decay strategy. The learning rate decreased by a factor of 0.50 when the loss stopped declining over 10 epochs.

B. EXPERIMENT ON UIEB DATASET

We first select underwater images from the UIEB and then divide these images into five categories: greenish and bluish images, yellowish images, low backscatter scenes (short distance between camera and scene), and high backscatter scenes (long distance between camera and scene). Then, we enhance images of the various categories in different methods. Moreover, we qualitatively compare the enhanced results of different methods and the corresponding images are shown in Fig. 8.

Due to the different attenuation ratios of red, green, and blue lights, photographs taken underwater always show color casts, such as greenish color, and bluish color shown in

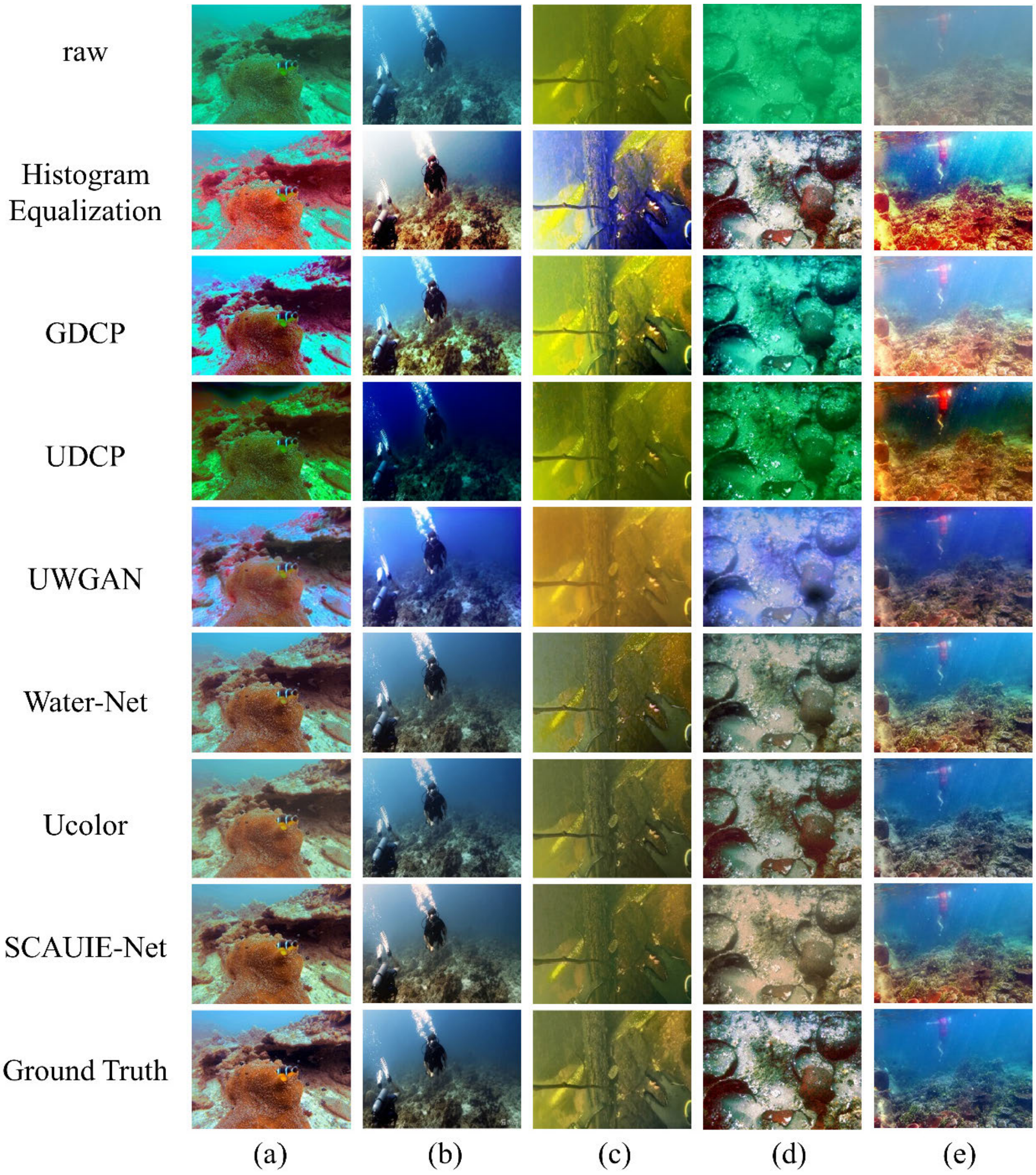

FIGURE 8. Qualitative comparisons on underwater samples from the UIEB dataset. (a) denotes the greenish image; (b) denotes the bluish image; (c) denotes the yellowish image; (d) denotes the low backscatter image; (e) denotes the high backscatter image. From the top row to the bottom row are raw underwater images, the results of Histogram Equalization [52], GDCP [30], UDCP [27], UWGAN [10], Water-Net [9], Ucolor [59], the proposed SCAUIE-Net and reference images.

Fig. 8 (a) and (b). Also, the particles that are suspended underwater will absorb blue lights which cause yellowish color casts shown in Fig. 8 (c). With the distance that light travels farther in underwater, the yellowish color cast will be deepened. Additionally, because of the light coming from atmospheric light reflected by the suspended particles [3], the backscatter will cause foggy veiling in underwater images. The low backscatter underwater image and high backscatter underwater image are separately shown in Fig.8 (d) and (e). Histogram Equalization [52] effectively improves the contrast of images and performs well in Fig. 8 (d). However, Histogram Equalization causes significant over-saturation. GDCP [30] brightens the underwater images. UDCP [27] could significantly dehaze underwater images but aggravate the color casts. UWGAN [10] improves the brightness and contrast of underwater images but the enhanced images are bluish. Water-Net [9] could effectively reduce artifacts but has local over-saturation. Ucolor [59] has fewer color casts but introduces artifacts shown in Fig. 8 (d). Our proposed method improves proper contrast and saturation making the foreground more natural but still exists obvious color casts. In conclusion, most methods could effectively remove

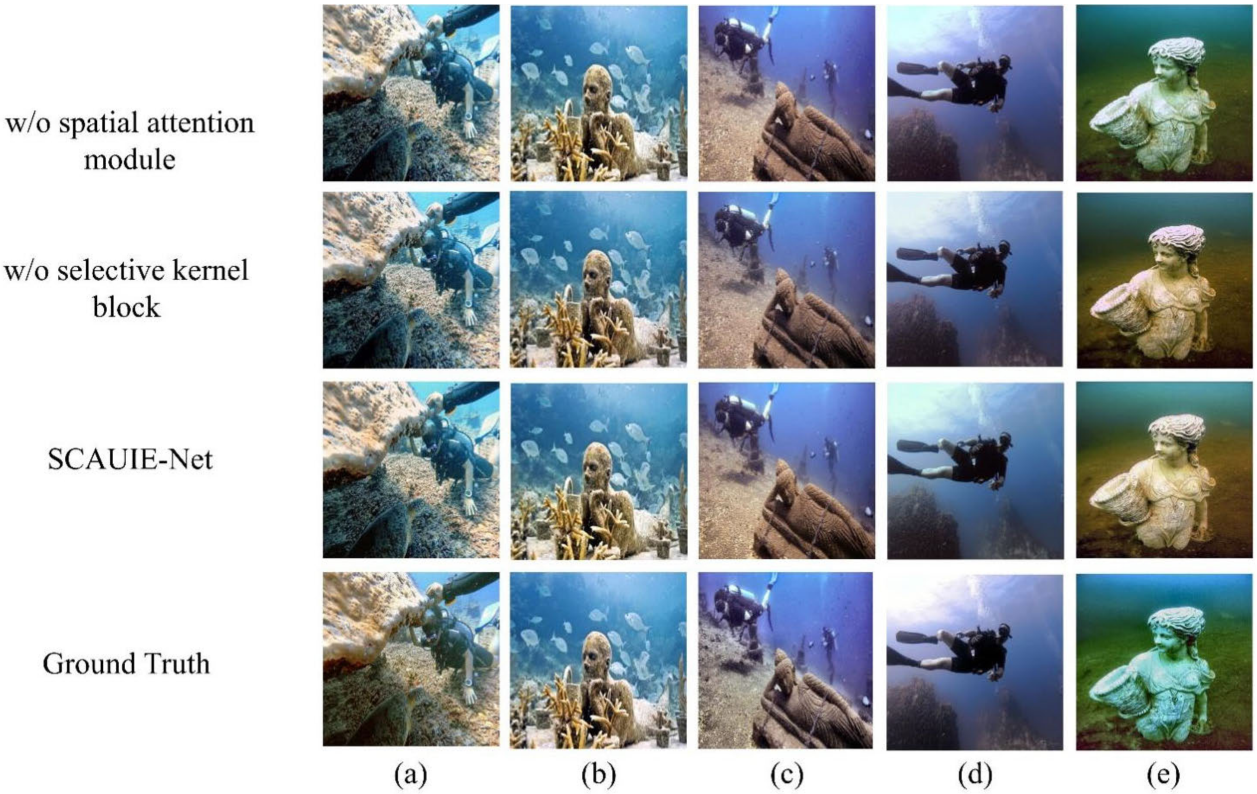

FIGURE 9. The enhanced results of without using Spatial Attention Module, without using Selective Kernel Block and SCAUIE-Net. (a)-(e) are underwater images randomly selected from the UIEB dataset and enhanced by each component. Top row: the results of without using Spatial Attention Module; Middle row: the results of without using Selective Kernel Block; Bottom row: the results of SCAUIE-Net.

the haze and improve the quality of underwater images. However, for deep-learning based methods, introducing artifacts, over-enhancement and color casts are still issues to be overcome.

To quantitatively evaluate the performance of different methods, we choose two commonly used full-reference metrics (*i.e.,* PSNR and SSIM) to assess the enhanced results on the UIEB dataset. A higher PSNR score means that the enhanced image is closer to the reference image in terms of image content. A higher SSIM denotes that the enhanced image is more like the reference image in the image structure. Meanwhile, we choose underwater color image quality evaluation (UCIQE [60]) and underwater image quality measures (UIQM [61]) as non-reference image quality metrics. UCIQE evaluates underwater quality by color density, saturation, and contrast. UIQM measures the underwater image quality by underwater colorfulness, underwater image sharpness and underwater image contrast.

The full-reference results of different methods on the UIEB dataset are reported in Table 1. Also, the non-reference results of different methods on the UIEB dataset are reported in Table 2. A higher UCIQE or UIQM score denotes a better human visual perception. We highlight the best performance in red, the second best is in blue. As shown in Table 1, our proposed SCAUIE-Net stands as the performer across all metrics and the Ucolor [59] performs the second best in full-reference metrics. The highest scores obtained by SCAUIE-Net demonstrate that our method could process details better. The scores of UCIQE and UIQM are shown in Table 2. In Table 2, the Histogram Equalization [52] (HE) performs best in UCIQE and GDCP [30] performs the second best in UCIQE; the UWGAN [10] ranks the best in UIQM and the Ucolor [59] achieves the second best in UIQM. The poor non-reference metrics generated from SCAUIE-Net show that the underwater non-reference metrics could not provide a good measure of human eyes' perception.

TABLE 1. Full-reference image quality evaluation in terms of PSNR, and SSIM on the UIEB dataset.

Method	PSNR (dB)↑	SSIM↑
raw	15.7996	0.7370
HE	16.5905	0.7836
GDCP [30]	14.7138	0.7246
UDCP [27]	11.4849	0.5373
UWGAN [10]	16.3993	0.7894
Water-Net [9]	19.1130	0.7971
Ucolor [59]	20.7296	0.8806
SCAUIE-Net	22.9286	0.9290

To further measure the performance between the different deep-learning-based methods (i.e., UWGAN [10], Water-Net [9], Ucolor [59], and SCAUIE-Net), we also compare

TABLE 2. Non-reference image quality evaluation in terms of UCIQE, UIQM on UIEB dataset.

Method	UCIQE [60]↑	UIQM [61]↑
raw	0.5325	2.3469
HE	0.6625	2.6318
GDCP [30]	0.6350	2.2831
UDCP [27]	0.5837	1.8664
UWGAN [10]	0.5811	2.9776
Water-Net [9]	0.6070	2.7118
Ucolor [59]	0.5772	2.7565
SCAUIE-Net	0.6192	2.6908
Reference	0.6214	2.9842

TABLE 3. FLOPs and Params for different algorithms.

Method	FLOPs (M)	Params (M)
UWGAN [10]	10455.35	8.78
Water-Net [9]	13670.75	1.09
Ucolor [59]	727053.15	148.77
SCAUIE-Net	9835.40	17.46

TABLE 4. Full-reference Image Quality Assessment without Spatial Attention Module and Selective Kernel Block.

Method	PSNR (dB)	SSIM
w/o Spatial Attention Module	22.6180	0.9252
w/o Selective Kernel Block	22.8665	0.9281
SCAUIE-Net	22.9286	0.9290

TABLE 5. Non-reference Image Quality Assessment without Spatial Attention Module and Selective Kernel Block.

Method	UCIQE	UIQM
w/o Spatial Attention Module	0.6161	2.6979
w/o Selective Kernel Block	0.6141	2.6819
SCAUIE-Net	0.6192	2.6908

FLOPs and Params for different algorithms in Table 3. As shown in Table 3, SCAUIE-Net has the fewest FLOPs compared to other deep-learning-based methods although we do not have the fewest Params. Because we introduce the selective kernel blocks that allow the network to reduce the FLOPs with increasing the number of Params. Water-Net employs a wider network, resulting in high FLOPs. Ucolor employs multi-color space encoder network, which causes too large network Params and FLOPs.

C. ABLATION STUDY

To demonstrate the effect of Spatial Attention Module and Selective Kernel Block in our network, we compared the proposed SCAUIE-Net without (w/o) Spatial Attention Module and without Selective Kernel Block as an ablation study. As shown in Table 4 and Table 5, Spatial Attention Module significantly improves the whole model performance though it decreases the performance of UIQM; Selective Kernel Block increases the performance of UIQM even though the improvements are not as obvious as Spatial Attention Module.

In Fig. 9, we select five images (a)-(e) that are performed on the network with different components. Fig. 9 (a), (b) shows that Spatial attention module and Selective Kernel Attention are effective in removing the background color cast which makes the images more realistic. Fig. 9 (c) shows that although the components could effectively process color cast, they perform not well in preserving image details and edge contour information. Fig. 9 (d) shows that Spatial Attention Module is not sensitive to the local color of the image, and the background color is relatively monotonous. Fig. 9 (e) shows that Selective Kernel Block compared to Spatial Attention module could obtain reasonable underwater images even though Spatial Attention Module could obtain more visually pleasing images.

V. CONCLUSION

This paper proposes an underwater image enhancement method called SCAUIE-Net, using both Spatial Attention Mechanism and Channel Attention Mechanism. The network uses a gated fusion strategy and attention mechanism on the UIEB dataset. Compared to Water-Net, this network uses U-Net architecture as the backbone which enlarges the network's depth and width. Besides, the network's Spatial Attention Module and Selective Block could perceive color differences of underwater images in different color channels and space regions. Combined with the multiple image quality loss function, the contrast and saturation of output images are further improved. In terms of full-reference metrics, the PSNR of SCAUIE-Net is 3.8156 higher than that of Water-Net, and the SSIM of SCAUIE-Net is 0.1289 higher than that of Water-Net. In the section on experiments, we validate the model's effectiveness through qualitative comparisons and quantitative comparisons with other underwater image enhancement methods. Furthermore, we conduct ablation studies to demonstrate the effectiveness of each component in SCAUIE-Net. Nevertheless, there is a lack of reference metrics to measure the underwater image quality. Therefore, appropriate evaluation metrics for measuring the performance and effectiveness of underwater image enhancement algorithms is an important direction for the future work.

REFERENCES

[1] R. Cui, L. Chen, C. Yang, and M. Chen, "Extended state observer-based integral sliding mode control for an underwater robot with unknown disturbances and uncertain nonlinearities," *IEEE Trans. Ind. Electron.*, vol. 64, no. 8, pp. 6785–6795, Aug. 2017.

[2] M. Ludvigsen, B. Sortland, G. Johnsen, and H. Singh, "Applications of geo-referenced underwater photo mosaics in marine biology and archaeology," *Oceanography*, vol. 20, no. 4, pp. 140–149, Dec. 2007.

[3] Y. Wang, W. Song, G. Fortino, L. Qi, W. Zhang, and A. Liotta, "An experimental-based review of image enhancement and image restoration methods for underwater imaging," *IEEE Access*, vol. 7, pp. 140233–140251, 2019.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.

[5] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.

[6] X. Li, W. Wang, X. Hu, and J. Yang, "Selective kernel networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 510–519.

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

[8] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, vol. 2, 2003, pp. 1398–1402.

[9] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao, "An underwater image enhancement benchmark dataset and beyond," *IEEE Trans. Image Process.*, vol. 29, pp. 4376–4389, 2020.

[10] N. Wang, Y. Zhou, F. Han, H. Zhu, and J. Yao, "UWGAN: Underwater GAN for real-world underwater color restoration and dehazing," 2019, *arXiv:1912.10269*.

[11] Q. Zhao, Z. Xin, Z. Yu, and B. Zheng, "Unpaired underwater image synthesis with a disentangled representation for underwater depth map prediction," *Sensors*, vol. 21, no. 9, p. 3268, May 2021.

[12] H. Blasinski, T. Lian, and J. Farrell, "Underwater image systems simulation," in *Proc. Imag. Syst. Appl.*, 2017, pp. 1–2, Paper ITh3E.3.

[13] H. Blasinski and J. Farrell, "A three parameter underwater image formation model," *Electron. Imag.*, vol. 28, no. 18, pp. 1–8, Feb. 2016.

[14] S. Anwar, C. Li, and F. Porikli, "Deep underwater image enhancement," 2018, *arXiv:1807.03528*.

[15] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from RGBD images," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2012, pp. 746–760.

[16] J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba, "SUN database: Large-scale scene recognition from abbey to zoo," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 3485–3492.

[17] D. Akkaynak and T. Treibitz, "Sea-thru: A method for removing water from underwater images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1682–1691.

[18] D. Berman, D. Levy, S. Avidan, and T. Treibitz, "Underwater single image color restoration using haze-lines and a new quantitative dataset," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 8, pp. 2822–2837, Aug. 2021.

[19] R. Liu, X. Fan, M. Zhu, M. Hou, and Z. Luo, "Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 12, pp. 4861–4875, Dec. 2020.

[20] L. Peng, C. Zhu, and L. Bian, "U-shape transformer for underwater image enhancement," 2021, *arXiv:2111.11843*.

[21] M. S. Hitam, E. A. Awalludin, W. N. J. H. W. Yussof, and Z. Bachok, "Mixture contrast limited adaptive histogram equalization for underwater image enhancement," in *Proc. Int. Conf. Comput. Appl. Technol. (ICCAT)*, Jan. 2013, pp. 1–5.

[22] J. Ma, X. Fan, S. X. Yang, X. Zhang, and X. Zhu, "Contrast limited adaptive histogram equalization-based fusion in YIQ and HSI color spaces for underwater image enhancement," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 32, no. 7, Jul. 2018, Art. no. 1854018.

[23] 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, Dec. 2014.

[24] 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 *Proc. Int. Conf. Multimedia Modeling*. Cham, Switzerland: Springer, 2018, pp. 453–465.

[25] 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.

[26] 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.

[27] 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. 2016.

[28] H. Liu and L. Chau, "Underwater image restoration based on contrast enhancement," in *Proc. IEEE Int. Conf. Digit. Signal Process. (DSP)*, Oct. 2016, pp. 584–588.

[29] C. Li, J. Guo, C. Guo, R. Cong, and J. Gong, "A hybrid method for underwater image correction," *Pattern Recognit. Lett.*, vol. 94, pp. 62–67, Jul. 2017.

[30] Y. Peng, K. Cao, and P. C. Cosman, "Generalization of the dark channel prior for single image restoration," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2856–2868, Jun. 2018.

[31] N. Carlevaris-Bianco, A. Mohan, and R. M. Eustice, "Initial results in underwater single image dehazing," in *Proc. OCEANS MTS/IEEE SEATTLE*, Sep. 2010, pp. 1–8.

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

[33] C. Li, J. Guo, S. Chen, Y. Tang, Y. Pang, and J. Wang, "Underwater image restoration based on minimum information loss principle and optical properties of underwater imaging," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 1993–1997.

[34] 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.

[35] 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.

[36] 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.

[37] 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.

[38] D. Akkaynak and T. Treibitz, "A revised underwater image formation model," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6723–6732.

[39] C. Li, J. Guo, and C. Guo, "Emerging from water: Underwater image color correction based on weakly supervised color transfer," *IEEE Signal Process. Lett.*, vol. 25, no. 3, pp. 323–327, Mar. 2018.

[40] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2242–2251.

[41] Y. Wang, J. Guo, H. Gao, and H. Yue, "UIEC^2-Net: CNN-based underwater image enhancement using two color space," *Signal Process., Image Commun.*, vol. 96, Aug. 2021, Art. no. 116250.

[42] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–15.

[43] Y. Zhang, K. Li, K. Li, B. Zhong, and Y. Fu, "Residual non-local attention networks for image restoration," 2019, *arXiv:1903.10082*.

[44] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 286–301.

[45] J. M. J. Valanarasu, P. Oza, I. Hacihaliloglu, and V. M. Patel, "Medical transformer: Gated axial-attention for medical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2021, pp. 36–46.

[46] F. Lv, Y. Li, and F. Lu, "Attention guided low-light image enhancement with a large scale low-light simulation dataset," *Int. J. Comput. Vis.*, vol. 129, no. 7, pp. 2175–2193, Jul. 2021.

[47] Y. Li and R. Chen, "UDA-Net: Densely attention network for underwater image enhancement," *IET Image Process.*, vol. 15, no. 3, pp. 774–785, Feb. 2021.

[48] J. Wang, P. Li, J. Deng, Y. Du, J. Zhuang, P. Liang, and P. Liu, "CA-GAN: Class-condition attention GAN for underwater image enhancement," *IEEE Access*, vol. 8, pp. 130719–130728, 2020.

[49] B. Fu, L. Wang, R. Wang, S. Fu, F. Liu, and X. Liu, "Underwater image restoration and enhancement via residual two-fold attention networks," *Int. J. Comput. Intell. Syst.*, vol. 14, no. 1, pp. 88–95, 2021.

[50] Q. Qi, K. Li, H. Zheng, X. Gao, G. Hou, and K. Sun, "SGUIE-Net: Semantic attention guided underwater image enhancement with multi-scale perception," 2022, *arXiv:2201.02832*.

[51] 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.*, Jun. 2012, pp. 81–88.

[52] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics Gems IV*, 1994, pp. 474–485.

[53] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[54] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. ICML*, 2010, pp. 807–814.

[55] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[56] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[57] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[58] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Trans. Comput. Imag.*, vol. 3, no. 1, pp. 47–57, Mar. 2017.

[59] C. Li, S. Anwar, J. Hou, R. Cong, C. Guo, and W. Ren, "Underwater image enhancement via medium transmission-guided multi-color space embedding," *IEEE Trans. Image Process.*, vol. 30, pp. 4985–5000, 2021.

[60] M. Yang and A. Sowmya, "An underwater color image quality evaluation metric," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 6062–6071, Dec. 2015.

[61] 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.

YUANHAO ZHONG is currently pursuing the bachelor's degree in electronic information engineering with the School of Electronics and Information Engineering, Guangdong Ocean University, China. His current research interests include low-level vision and AI-generated content.

JI WANG received the B.S. degree in electronics and communication technology from Liaoning University, China, in 1994, and the M.S. degree in engineering from the Guangdong University of Technology, in 2010. He is currently a Professor with the Institute of Electronics and Information Engineering, Guangdong Ocean University. He is also the Director of the Guangdong Intelligent Ocean Sensor Network and its Equipment Engineering Technology Research Center. His research interests include wireless sensor networks and ocean Internet of Things information processing, and communication systems. He is a Senior Member of the China Electronics Society and a member of the Guangdong Electronic Information Education and Reference Committee.

QINGJIE LU received the Ph.D. degree in optical engineering from the University of Shanghai for Science and Technology, China, in 2020. He is currently a Teacher with the Institute of Electronics and Information Engineering, Guangdong Ocean University. His research includes wireless communication and laser inteference mesurement.