IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 26 March 2024, accepted 8 May 2024, date of publication 23 May 2024, date of current version 4 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3404613

RESEARCH ARTICLE

DIRBW-Net: An Improved Inverted Residual Network Model for Underwater Image Enhancement

YONGLI AN[1], YAN FENG[1], NA YUAN[2], ZHANLIN JI[1,3,4], (Member, IEEE), AND IVAN GANCHEV[4,5,6], (Senior Member, IEEE)

[1]College of Artificial Intelligence, North China University of Science and Technology, Tangshan 063000, China
[2]Intelligence and Information Engineering College, Tangshan University, Tangshan 063000, China
[3]College of Mathematics and Computer Science, Zhejiang A&F University, Hangzhou 311300, China
[4]Telecommunications Research Centre (TRC), University of Limerick, Limerick, V94 T9PX Ireland
[5]Department of Computer Systems, University of Plovdiv "Paisii Hilendarski," 4000 Plovdiv, Bulgaria
[6]Institute of Mathematics and Informatics–Bulgarian Academy of Sciences, 1040 Sofia, Bulgaria

Corresponding authors: Na Yuan (yuanna@tsc.edu.cn), Zhanlin Ji (zhanlin.ji@gmail.com), and Ivan Ganchev (ivan.ganchev@ul.ie)

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFE0135700; in part by the High Level Talent Support Project of Hebei Province under Grant A201903011; in part by the Natural Science Foundation of Hebei Province under Grant F2018209358; in part by the Telecommunications Research Centre (TRC) of the University of Limerick, Ireland; and realized with the help of the infrastructure purchased under the National Roadmap for Scientific Infrastructure, financially coordinated by the Ministry of Education and Science of Bulgaria, under Grant D01-325/01.12.2023.

ABSTRACT Underwater photography is challenged by optical distortions caused by water absorption and scattering phenomena. These distortions manifest as color aberrations, image blurring, and reduced contrast in underwater scenes. To address these issues, this paper proposes a novel underwater image enhancement model, called DIRBW-Net, leveraging an improved inverted residual network. In order to minimize the interference of the Batch Normalization (BN) layer on color information, newly designed Double-layer Inverted Residual Blocks (DIRBs) are introduced, which omit the BN layer and extract deep feature information from the input images. Subsequently, each input image is fused with the intermediate feature map using skip connections to ensure consistency between local and global image information, thus effectively enhancing the image quality. In the concluding phase, effects of diverse activation functions are studied, opting for the *h*-swish activation function to further boost the overall model performance. DIRBW-Net is evaluated on a public dataset, with comparisons drawn against existing representative models. The experiments showcase a notable success in enhancing the underwater image quality when using the proposed model.

INDEX TERMS Underwater image enhancement, convolutional neural network (CNN), residual network, deep learning.

I. INTRODUCTION

During the process of image generation, factors such as lighting, motion, occlusion, and exposure inevitably affect the image quality, leading to degradation. Underwater images are particularly affected by uneven lighting, low visibility, refraction, and scattering of light, as well as the existence of small particles and suspended matter, resulting in color deviations, color artifacts, and blurred details. Marine surveying, underwater biology, underwater cultural heritage preservation, and ship maintenance tasks all require high-quality images as to be performed well. However, the aforementioned issues significantly affect the accuracy and reliability of underwater visual tasks and pose enormous challenges to underwater operations. Therefore, research, development, and application of techniques for enhancing underwater images are particularly important.

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

Underwater image enhancement techniques are primarily designed to elevate image quality through the meticulous adjustment of various image attributes. Enhancement methods of a traditional nature can be categorized into two main groups: (1) employing non-physical models; and (2) based on physical models. Common methods for non-physical models include histogram equalization [1], wavelet transform [2], white balance [3], etc. These methods mainly adjust the distribution of pixel values in the images to achieve a uniform distribution and enhance the visual fidelity of images. However, given the existence of factors such as detail blurring and color distortion in underwater images, relying solely on a single enhancement method proves inadequate. Consequently, several fusion methods have been developed to address this limitation. For instance, Jia et al. [4] transform the image color space into a hue-saturation-intensity (HSI) space and apply a wavelet transform to distinguish the low- and high-frequency bands of the luminance channel. The low-frequency band undergoes adjustment through Retinex, while the high-frequency band is fine-tuned using multi-channel filtering techniques to enhance images. However, during the image conversion process, there may be a loss of image information and color distortion. Sanila et al. [5] transform images into a Lab color model and apply white balance to adjust the brightness component, compensating for other components and eliminating unwanted color casts, generating blurred images using a Gaussian filter, and then utilize the images obtained through white balance for unsharp masking to ensure edge enhancement. Additionally, these authors incorporate contrast-limited adaptive histogram equalization to achieve contrast enhancement. While non-physical models can generate improved images with enhanced visual quality, they often neglect the intrinsic characteristics of underwater physical imaging. As a consequence, this oversight can result in color deviations and other related issues.

The methods for underwater image enhancement, employing physical models, aim to mathematically model the image degradation process and estimate parameters to invert and obtain clear images. Existing models include scattering models [6], color balance models [7], degradation models [8], etc. For instance, Chen et al. [9] employ a dark channel prior (DCP) algorithm for super-pixel processing. The DCP algorithm leverages frequency domain information from underwater images, examining color variations and frequency shifts in adjacent regions within the scene. By estimating transmittance and applying it for color correction, the algorithm achieves good image enhancement. Liu et al. [10] preprocess underwater images using color channel transmission. They estimate the transmission rate based on the pixel distribution along the curve, simulate the underwater light attenuation process, and incorporate saturation adjustment on the three-color channels to achieve color enhancement while preserving edge details. This type of method relies on mathematical modeling of the underwater imaging process and requires the addition of certain prior knowledge, which has limitations for different underwater scenes. Simultaneously, the parameter estimation introduces heightened computational complexity to the algorithm, potentially causing deviation in later estimation and poor enhancement effects.

The rise of deep learning has opened up a new developmental avenue for image processing and computer vision applications. When addressing image enhancement, traditional methods frequently depend on strong assumptions or predefined conditions, while deep learning methods resolve the issue by learning the intrinsic distribution characteristics of the data. For complex underwater scenes, traditional enhancement methods neglect the underwater mapping relationship, concentrating solely on color improvement. Conversely, deep learning methods not only extract relevant features from the original images by means of artificial neural networks but also acquire knowledge of the distinctive mapping patterns unique to underwater imagery. This addresses the limitations of traditional underwater enhancement methods.

Based on the neural network type, underwater enhancement networks in deep learning can be categorized into two groups: convolutional neural networks (CNNs) and generative adversarial networks (GANs). Li et al. [11] proposed a network model rooted in the priori of underwater scenes, that integrates the underlying principles of underwater imaging, encompassing the physical aspects and optical characteristics of underwater scenes. By synthesizing degraded datasets from various scenes, a CNN was trained to effectively reconstruct clear images. Wang et al. [12] proposed achieving underwater image enhancement through the fusion of two different color spaces, performing fundamental enhancement operations on pixel blocks within the red-green-blue (RGB) color space. Furthermore, global adjustment blocks are integrated in the hue-saturation-value (HSV) color space to finetune the saturation and brightness of underwater images. Both operations are combined within a unified CNN architecture. Islam et al. [13] introduced a method that incorporates underwater scene priors and trains on a synthesized dataset to better adapt the generator to the specific characteristics exhibited by underwater images. This serves as the fundamental groundwork for subsequent adversarial training, enabling the learning of distinctive features and textures inherent in underwater images. However, this approach presents challenges such as high model complexity and extended training times. Lin et al. [14] primarily adjusted a GAN network to enhance visual perception by adding a U-Net structure to the discriminator and employing dilated convolution in the generator. This model effectively removes color artifacts and generates high-quality underwater images.

While numerous methods for underwater image enhancement have been proposed, they still encounter challenges. Traditional methods rely on prior knowledge, limiting their applicability. Conversely, deep learning based methods yield varying effects in different environments, leading to

instability in specific situations. In addition, some deep learning models have complex structures and demand a substantial number of parameters for training, restricting deployment on devices with limited computing resources.

Addressing the aforementioned challenges, this paper proposes a novel underwater image enhancement model grounded on inversion residual principles. The model refines the original Inverted Residual Block (IRB) to mitigate the BN layer interference on image color information and eliminate redundant BN layers. The traditional residual block is improved by designing two layers to deal with the problem of gradient disappearance. The distinctive advantage of this method lies in its detailed supplementation of depth feature information during extraction, contributing to enhanced model performance. The network model initially acquires shallow feature information through standard convolution operations, followed by the extraction of deep feature information via a double-inversion residual process. Ultimately, the deep feature information is globally integrated with the input image through skip connections, yielding an ultimate enhancement map.

Evaluation, performed on the Underwater Image Enhancement Benchmark (EUVP) dataset, demonstrates that, in comparison with other underwater image enhancement models, the proposed model adeptly eliminates the blue-green background in the visual output and effectively addresses issues such as color deviation, low contrast, and blurring. Simultaneously, there is a noteworthy enhancement in objective evaluation indices. Additionally, the effectiveness of various components, utilized by the proposed model, is confirmed through ablation study experiments.

The rest of the paper is organized as follows. Section II describes the proposed model. Section III presents the experimentally obtained performance evaluation results along with their analysis. Finally, section IV concludes the paper.

II. PROPOSED DIRBW-Net MODEL

To address color cast issues arising from uneven illumination in underwater photography, this paper introduces a novel underwater image enhancement model built upon the inverted residual network framework. To streamline computational complexity, the network omits the up-down sampling step, conducting all operations directly on the feature map. The proposed model, illustrated in Fig. 1, consists of three main components, namely a shallow feature extraction module, a Double-layer Inverted Residual Block (DIRB), and a feature fusion module, described in detail in the next subsections.

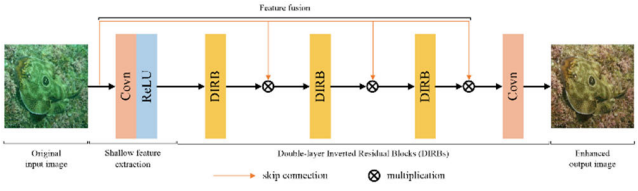

FIGURE 1. The overall structure of the proposed model.

A. SHALLOW FEATURE EXTRACTION MODULE

Comprised of a 3×3 standard convolution and an activation layer employing the ReLU function, this module extracts shallow information from the input RGB images and generates the requisite feature map for subsequent deep feature extraction, as follows:

$$I_0 = Conv_{3\times3}(x) \tag{1}$$

where x denotes the input underwater degraded image, $Conv_{3\times3}$ denotes a standard convolution operation with a kernel size of 3×3, and I_0 denotes the outcome obtained from the input image after passing through the shallow feature extraction module.

During the enhancement process, each input degraded image undergoes preprocessing, resizing it to a 256×256 3-channel image. This modification not only reduces computational costs but also ensures consistency of the feature extraction map in terms of shape and spatial characteristics. Following the convolution operation, the output feature map is transformed into 32 channels, serving as input for subsequent deep feature extraction. The image size remains constant throughout the entire feature extraction process.

B. DOUBLE-LAYER INVERTED RESIDUAL BLOCK

Generally, the accuracy of a neural network model can be enhanced by stacking multiple convolutional layers or by adding extra nodes. However, as the number of nodes and layers increases, issues like gradient disappearance or gradient explosion may arise in the propagation of gradient information, leading to a significant decline in the model performance. Therefore, He et al. [15] introduced a method known as residual learning. By introducing skip connections, information can be rapidly propagated in the deep network, allowing the network to learn the previous feature information for creating a residual mapping. This approach mitigates the issue of vanishing gradients associated with information propagation.

The Inverted Residual Block (IRB) [16] is a variation of the residual network commonly applied in image processing tasks. It usually converts low-resolution images into high-resolution images and has a good ability to restore lost detail information and improve clarity. Moreover, it can recognize the category of pixels in the images, which helps capture local and global features. The computational complexity and parameter count are reduced while still maintaining good accuracy. However, in image enhancement tasks, the primary role of the BN layer is to normalize the color distribution, typically resulting in input data with zero mean and unit variance. This leads to a shift in the color distribution of the input images. Additionally, the BN layer can alter the contrast of the input images, disrupting the original contrast information and impacting the visual perception of the images. Considering this appealing phenomenon, we propose a novel block, called DIRB, shown in Fig. 2.

DIRB is composed of two branches. The upper branch extracts deep features from the input feature map through

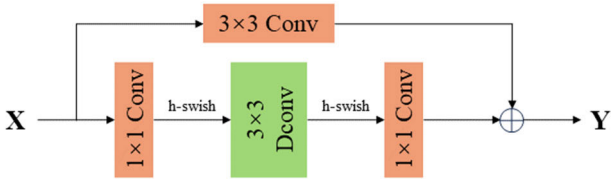

FIGURE 2. The newly designed double-layer inverted residual block (DIRB), utilized by the proposed model.

a 3 × 3 convolution. The lower branch includes a 3 × 3 depthwise separable convolution and two 1 × 1 convolutions. The *h*-swish function [17] is used as an activation function between the two convolutions. Firstly, the extracted feature map is input into the 1 × 1 convolutional layer to expand the channel information so as to increase the depth of the feature map and help extract high-dimensional features. Then, the extended feature map undergoes filtering through a 3 × 3 depthwise separable convolution, extracting intricate features for deep learning. Subsequently, the feature map undergoes adjustment through a 1 × 1 convolution to align its final output channel count with that of the input, completing the linear mapping of the feature map. This process ensures the size and depth of the feature map both align with the input. Finally, residual learning is performed by adding the feature maps of the upper branch and the feature maps of the lower branch by means of a skip connection. This way, more detailed features of the images can be obtained. Compared to IRB, DIRB eliminates redundant BN layers, and while the upper branch's feature extraction increases the network's computational complexity, it prevents the occurrence of vanishing gradients.

In the proposed model, the deep feature extraction is performed by means of three DIRBs, as follows:

$$F_3 = h^3_{DIRB}\left(h^2_{DIRB}\left(h^1_{DIRB}\left(I_0\right)\right)\right) \tag{2}$$

where h^n_{DIRB} $(n = \{1, 2, 3\})$ denotes the n-th DIRB function and F_3 denotes the feature map of the third output. The DIRB feature extraction is performed as follows:

$$h^n_{DIRB} = f_{com}(f_{dep}(f_{\exp}(h^{n-1}_{DIRB}))) + Conv_{3\times3}(h^{n-1}_{DIRB}) \tag{3}$$

$$h^1_{DIRB} = f_{com}(f_{dep}(f_{\exp}(I_0))) + Conv_{3\times3}(I_0) \tag{4}$$

where f_{com} $(\cdot)$ denotes the feature compression, f_{dep} $(\cdot)$ denotes the depthwise separable convolution operation, and f_{exp} $(\cdot)$ denotes the feature expansion.

C. FEATURE FUSION MODULE

To augment the model's capability in handling the degradation of underwater image characteristics, a global feature fusion module is incorporated into it. This module merges the feature maps from each DIRB with the input image using skip connections. This approach facilitates the fusion of global and local information, effectively enhancing the quality of the images and mitigating the issue of gradient disappearance. It is essential to emphasize that this feature fusion operation is not constrained by the number of channels in the feature map. The output y_n of the n-th concatenation operation is obtained as follows:

$$y_n = S(h^n_{DIRB}, x) \tag{5}$$

where $S(\cdot)$ denotes the concatenation operation. As this operation is performed only on the channel dimension, the image size does not change. After the last feature fusion, an enhanced image is generated using a 3 × 3 convolutional layer, which ensures that the output has the correct number of channels, while smoothing the gradually refined deep feature information and retaining the appropriate feature information.

D. ACTIVATION FUNCTIONS

The lack of a loss function causes a multi-layer neural network to behave as an equivalent form of a single-layer linear model, which is unable to handle intricate non-linear relationships. To enhance the network's expressive capability and adaptability, introducing an activation function into it becomes imperative. As one of the key components of neural networks, activation functions can introduce non-linear characteristics so that the network can better capture and represent non-linear data patterns. This further extends the representation power of neural networks, enabling them to learn and represent more complex functional relationships. Commonly used activation functions include ReLU [18], Leaky ReLU [19], Sigmoid [20], and GELU [21], depicted in Fig. 3.

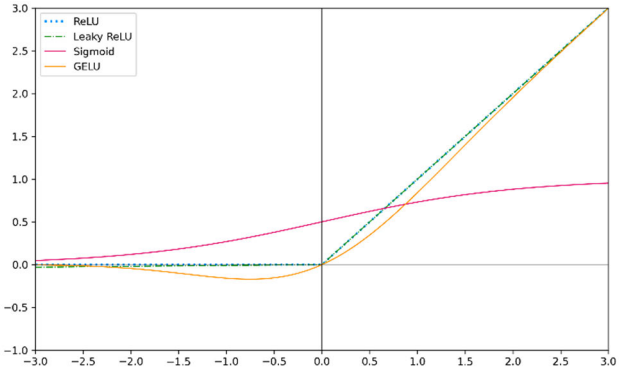

FIGURE 3. Four popular activation functions.

ReLU is a widely utilized activation function. Various modifications of it exist, one of which is ReLU6 [22]. This function restricts the part of the input value greater than 6 to the constant value of 6, which helps suppress the gradient explosion problem, while the output in the negative number region is set to 0, as follows:

$$ReLU6\,(x) = \begin{cases} 0, & x \le 0 \\ x, & 0 < x < 6 \\ 6, & x \ge 6 \end{cases} \tag{6}$$

As could be inferred from (6), when the input is negative, the neurons in this region are in a deactivated state, which causes the model to fail in learning the feature information of this region. This reduces the expressive power of the network, affecting negatively the model. In order to prevent the issue of dead neurons, the proposed model employs the *h*-swish activation function, as follows:

$$h\_swish\,(x) = \frac{x * ReLU6\,(x+3)}{6} \tag{7}$$

As shown in Fig. 4, compared to ReLU6, *h*-swish introduces nonlinearity in the negative region, which makes the curve of activation function smoother. At the same time, *h*-swish has a certain value in the negative region, which helps alleviate the problem of neurons death.

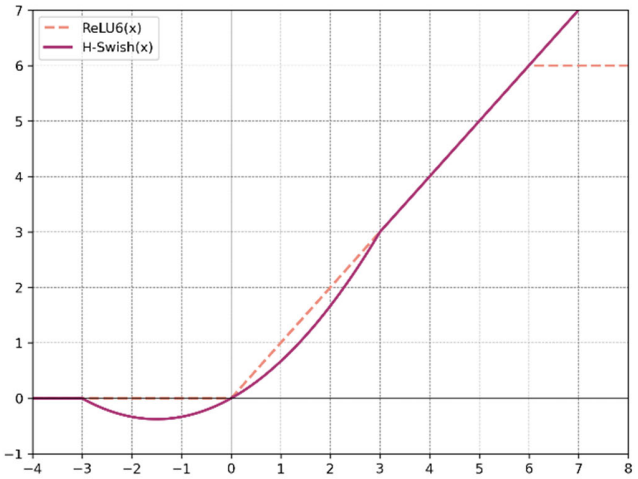

FIGURE 4. ReLU6 and *h*-swish activation functions.

E. LOSS FUNCTIONS

The proposed model utilizes the mean square error (MSE) loss function [23] to assess the disparity between the enhanced images and the original images. Additionally, the MSE loss is sensitive to brightness, which helps retain the consistency of the overall brightness or brightness distribution information. The MSE loss is calculated as follows:

$$L_{MSE} = \frac{1}{N}\sum_{i=1}^{N}(x_i - t_i)^2 \tag{8}$$

where N denotes the number of input images, x_i denotes the i-th original input image, and t_i denotes the i-th enhanced output image.

In the experiments, the MSE loss is combined with the content loss [13] that extracts high-level semantic information through a VGG-16 network pre-trained on ImageNet [24]. This content loss function preserves the semantic content of images and generates high-dimensional visual perception images, allowing the generated images to acquire realistic detail information with enhanced quality and realism. The content loss is calculated as follows, [13]:

$$L_{content} = \frac{1}{C_j H_j W_j}\left\|\phi_j(\hat{y}) - \phi_j(y)\right\|_2^2 \tag{9}$$

where j represents the feature map of shape $C_j \times H_j \times W_j$ extracted by convolution at the j-th layer of the pre-trained VGG-16 network, $\hat{y}$ denotes an enhanced output image, y denotes the original input image, and $\phi_j(\hat{y})$ and $\phi_j(y)$ denote the pixel values of the corresponding image features, respectively.

Combining these two loss functions in a linear fashion, as shown below, allows enhancing the robustness of the proposed model:

$$L = \lambda_1 * L_{MSE} + \lambda_2 * L_{content} \tag{10}$$

where λ_1 and λ_2 denote the weight of the MSE loss and content loss, with values set to 5 and 0.01, respectively, based on a large number of experiments.

III. EXPERIMENTAL RESULTS AND ANALYSIS

The proposed model was implemented by means of the TensorFlow deep learning framework using a 14-core Intel(R) Xeon(R) CPU E5-2680v4@2.40GHz processor. The computation acceleration was performed by an NVIDIA GeForce RTX 3090 (24GB) GPU. During the model training process, the ADAM optimizer was used for parameter optimization, the learning rate was set to 0.0001, the batch size was set to 1, and the number of training epochs was set to 30.

The EUVP dataset [25] was used in the experiments, as it contains color images featuring various underwater sediments, organisms, seaweed, and corals, within a variety of underwater environments. In addition, this dataset provides both high-quality and low-quality images in a pairwise manner. From these, 256 × 256 underwater images with diverse color tones were selected for use in the experiments. 1500 pairs of these images were randomly selected for model training and another 515 pairs of images were used for model testing.

The performance of the proposed model was evaluated in comparison to three traditional image enhancement models (Underwater Dark Channel Prior (UDCP) [26], Histogram Equalization (HE) [27], Underwater Image Enhancement using LIME and ACP (ULAP) [28]), and four deep learning enhancement models (UWCNN [11], UDnet [29], WaterNet [25], FUnIE-GAN [13]).

A. SUBJECTIVE EFFECT ANALYSIS

Sample subjective effect results, obtained in the conducted experiments, are illustrated in Fig. 5. As can be seen, the UDCP model preserves the complete texture information of the images by utilizing the minimum value to estimate pixel brightness. However, the presence of atmospheric light leads to neglecting or slight enhancement of dark areas, which makes some image regions excessively dark. The HE model enhances image contrast by remapping pixels to improve brightness, but an excessive high contrast may lead

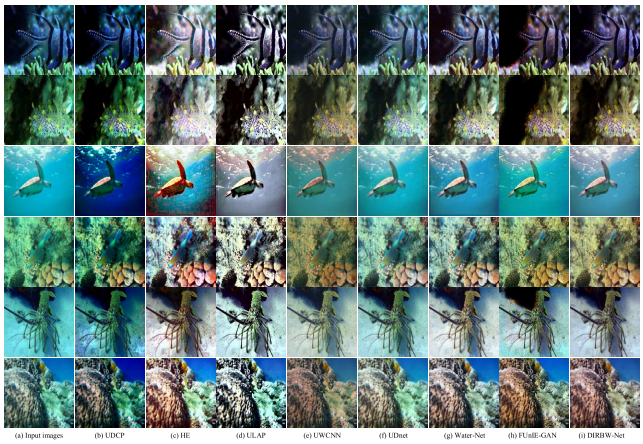

FIGURE 5. Sample subjective effect comparison of different models.

to overexposure. Images output by the ULAP model exhibit a deeper blue-green background than the original images in certain regions. The UWCNN model effectively removes the blue and green background, but the enhanced images exhibit color distortion and noticeable dark areas. The UDnet model successfully eliminates the blue and green background, yet the white patch overall brightens the enhanced image and introduces color distortion. The Water-Net model efficiently eliminates the blue tone and reinstates a favorable color effect. However, for color shift improvement under the blue-green background, the red channel tends to be overcompensated, resulting in blurred image details. The FUnIE-GAN model achieves comprehensive color correction, effectively removing the blue and green backgrounds. However, certain areas in the enhanced images may exhibit excessive darkness, leading to a loss of some image details. In comparison to these models, the proposed DIRBW-Net model excels not only in effectively eliminating the blue or green background in underwater images, yielding natural and clear enhancements, but also restores the true color distribution for various underwater image types. This improves image contrast and brightness, resulting in the production of high-quality images.

B. OBJECTIVE EFFECT ANALYSIS

For a thorough model performance evaluation, both objective evaluation metrics using real image values and metrics without real image values were selected for use. Within the first group, the Peak Signal to Noise Ratio (PSNR) and the Structural Similarity Index Measure (SSIM) [30] were utilized, as these offer a quantitative assessment of the improved image quality with reference to the real images.

PSNR is employed to measure the quality of enhanced images compared to reference images. It offers a numerical gauge of the similarity between the two images, signifying the degree of fidelity in preserving the original image details. A higher PSNR value signifies lower distortion and higher image quality. PSNR is calculated as follows:

$$\mathrm{PSNR} = 10 * \lg\left(\frac{255^2}{\mathrm{MSE}}\right) \tag{11}$$

where MSE, denoting the mean square error between the enhanced image and the original image, is calculated as follows:

$$\mathrm{MSE} = \frac{1}{H * W} \sum_{i=1}^{H} \sum_{j=1}^{W} (X(i,j) - Y(i,j))^2 \tag{12}$$

where W and H denote the numbers of pixels in the images, counted horizontally and vertically, respectively.

SSIM is utilized to assess the similarity between enhanced images and original images, by taking into account their structural information. It produces a value between 0 and 1, where a value closer to 1 signifies less distortion, higher preservation of image information, and greater similarity in terms of structure. SSIM is calculated as follows:

$$\mathrm{SSIM} = \frac{\left(2\mu_x\mu_y + C_1\right)\left(2\sigma_{xy} + C_2\right)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \tag{13}$$

where μ_x and μ_y denote the mean brightness of images x and y, respectively, σ_{xy} denotes the covariance of x and y, σ_x^2 and σ_y^2 denote the standard deviations of x and y, respectively, and C_1 and C_2 are constants with the values of $(255 \times 0.01)^2$ and $(255 \times 0.03)^2$, respectively.

The Underwater Color Image Quality Evaluation (UCIQE) [31] is a non-reference evaluation metric designed specifically for underwater images. It measures the saturation, color richness, and contrast of enhanced underwater images. A higher UCIQE value signifies better quality of the underwater image in terms of color reproduction. UCIQE is calculated as follows:

$$\mathrm{UCIQE} = c_1 * \sigma_c + c_2 * con_l + c_3 * \mu_s \tag{14}$$

where σ_c denotes the standard deviation of chroma, con_1 denotes the contrast of brightness, and μ_s denotes the average saturation value. Basically, the UCIQE score is calculated using a linear combination of these factors, with weights given by c_1, c_2, and c_3. Specifically, in the experiments, these weights were set to the following values: $c_1 = 0.4680$, $c_2 = 0.2745$, and $c_3 = 0.2576$.

TABLE 1. Performance comparison of different models, based on EUVP dataset.

Model	PSNR (*dB*)	SSIM	UCIQE	Parameters (*number*)
UDCP	15.98	0.639	**0.514**	-
HE	13.74	0.631	0.494	-
ULAP	19.63	0.747	0.494	-
UWCNN	19.76	0.773	0.370	39,972
UDnet	19.24	0.787	0.407	16,116,602
Water-Net	20.65	0.799	0.422	1,090,668
FUnIE-GAN	21.40	0.703	0.451	7,716,259
DIRBW-Net (*proposed*)	**22.57**	**0.817**	0.427	**26,372**

The obtained averaged results of the model performance comparison, performed on EUVP dataset, are shown in Table 1 (the best results are highlighted in **bold**). As can be seen, the proposed model is superior to all other compared models, based on PSNR and SSIM, surpassing the second-ranked model by 1.17 dB and 0.018, respectively. Only in terms of UCIQE, the proposed model does not perform so well. Here, the traditional enhancement models achieved the highest results, which is attributed to UCIQE values being derived from the calculation of three indicators (i.e., the image color, brightness, and saturation). Traditional enhancement models directly adjust image pixel values, leading to quadrants with high color indicators due to excessive enhancement of pixel values and color differences. Higher color values contribute to improved UCIQE values, enabling traditional enhancement models to achieve higher UCIQE. Compared to the deep learning models, the proposed model ranks second (after FUnIE-GAN), according to UCIQE. However, FUnIE-GAN, like other deep learning models, does not address factors such as color deviation, resulting from excessive enhancement. Although FUnIE-GAN scores higher on UCIQE, its visual enhancement outcomes are still suboptimal. Concerning the number of model parameters, the proposed model exhibits the lowest parameter count compared to the deep learning models. This characteristic is advantageous for model deployment on devices with limited computing resources, affirming the efficacy of the proposed model in this respect.

C. QUANTITATIVE COMPARATIVE ANALYSIS OF ACTIVATION FUNCTIONS

In deep learning networks, different activation functions exhibit distinct effects, playing a critical role in both the model's performance and training process. Among the numerous effects, the activation function notably impacts the model training speed. Specifically, the Sigmoid function could easily lead to a gradient disappearance problem, causing the model training to become slow or even stagnated, while the ReLU function has better properties for gradient propagation and can accelerate the model training process. In addition, the activation function has a profound impact in several other aspects, including the model convergence and expressiveness, and the gradient stability. Therefore, it is particularly important to choose an appropriate activation function. By experimenting with various activation functions, we sought to identify the most suitable one for the proposed model as to optimize its performance. As depicted in Table 2, different activation functions have varying impacts on the quality of generated images. Among these functions, the proposed model exhibits the most substantial performance improvement with images generated under the *h*-swish function.

D. ABLATION STUDY

In order to assess the individual components contribution to the overall performance of the proposed model, ablation study experiments were conducted. w.r.t. three main components – DIRBs, skip connections, and content loss. The subjective comparison results obtained are shown in Fig. 6. From the visual perspective, the generated enhanced images can effectively eliminate the blue-green background even when using the original inverted residual block, or when the skip connections or perceptual content loss are absent. However, under the blue background, the enhanced images will have different degrees of color cast. Table 3 displays the

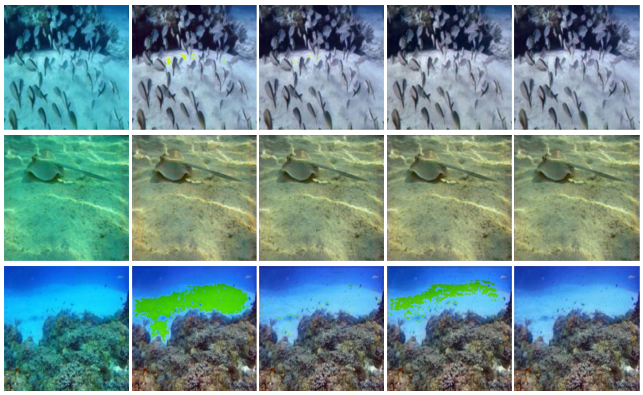

FIGURE 6. Qualitative results of ablation study experiments: (a) original images; (b) images, enhanced without using DIRBs; (c) images, enhanced without using skip connections; (d) images, enhanced without using content loss; (e) images, enhanced by the proposed DIRBW-Net model.

TABLE 2. Quantitative comparison of different activation functions, performed on EUVP dataset.

Activation function	PSNR (*dB*)	SSIM	UCIQE
ReLU	22.37	0.813	0.422
ReLU6	22.36	0.813	0.423
Leaky ReLU	22.45	0.815	0.424
Sigmoid	22.25	0.809	0.425
GELU	22.51	0.816	0.424
h-swish	**22.57**	**0.817**	**0.427**

TABLE 3. Ablation study results.

	PSNR (*dB*)	SSIM	UCIQE
skip connections + content loss (i.e., without DIRBs)	21.76	0.791	**0.444**
content loss + DIRBs (i.e., without skip connections)	22.39	0.813	0.429
DIRBs + skip connections (i.e., without content loss)	21.98	0.811	0.422
DIRBs + skip connections + content loss (i.e., the proposed DIRBW-Net model).	**22.57**	**0.817**	0.427

average results of the conducted ablation study experiments, with the best result for each objective evaluation metric highlighted in **bold**. As can be observed, the use of all three components allows to achieve the best PSNR and SSIM values, while maintaining structural recognition between the enhanced image and the corresponding original image. The replacement of the original IRBs with DIRBs is not tolerated by the third metric only, i.e., UCIQE.

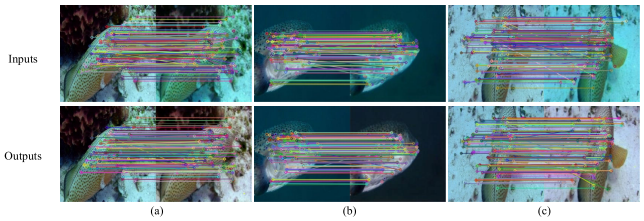

FIGURE 7. Sample image feature point matching test visualization results.

TABLE 4. Sample image feature point matching test results.

Example	Matched feature points in the original image	Matched feature points in the enhanced image	Increased quantity
Fig. 7a	282	306	+24
Fig. 7b	346	354	+8
Fig. 7c	294	317	+23

E. APPLICATION ANALYSIS

Underwater image enhancement is the key to image preprocessing to provide high-quality image input for subsequent high-level vision tasks such as classification tasks. Image feature point matching is a widely used technique for identifying corresponding feature points across different images, significantly impacting vision tasks. The quantity of feature points extracted from images plays a crucial role in influencing the efficacy of advanced vision tasks like underwater target recognition, classification, tracking, etc. Therefore, as a final experiment, the Oriented FAST[1] and Rotated BRIEF[2] (ORB) matching algorithm [32] was employed to determine the number of feature points between original images and true reference images, as well as between the true reference images and enhanced images. The results of the feature point matching test are illustrated in Fig. 7, with the corresponding numbers of matched feature points presented in Table 4. The test results reveal a substantial increase in the number of feature points in the enhanced images generated by the proposed model compared to the original images. This outcome confirms the effectiveness of the proposed model in enhancing the local features of underwater images, thereby providing a robust foundation for subsequent advanced underwater vision tasks.

[1]FAST: Features from Accelerated Segment Test.
[2]BRIEF: Binary Robust Independent Elementary Features.

IV. CONCLUSION

To address challenges posed by illumination variations in the underwater image capturing process, this paper has proposed a novel network model, named DIRBW-Net, for underwater image enhancement by utilizing the inversion residual method. Through the optimization of the inversion residual structure, the proposed model employs a two-layer architecture to replace the redundant BN layer, effectively extracting feature information from the input images using an up and down branch strategy. This model design not only preserves more detailed information during the image enhancement process but also mitigates the issue of gradient vanishing during model training. The model effectiveness is ensured through the selection of an appropriate activation function. By incorporating skip connections, the proposed model captures global information from the images, thereby enhancing overall image quality. Comparative evaluations, performed on the EUVP dataset, using image quality metrics such as PSNR, SSIM, and UCIQE, affirm that the proposed DIRBW-Net model surpasses the traditional image enhancement models and deep learning enhancement models, participating in the experiments. Ablation study experiments further validated the effectiveness of each main component of the proposed model. It is noteworthy that the dataset utilized for model training was synthetic, generated by simulating underwater environments on real photos rather than directly collecting data from real underwater settings. Future research will focus on exploring lightweight networks for real underwater image enhancement tasks and endeavoring to achieve real-time enhancement of underwater videos.

REFERENCES

[1] G. Ulutas and B. Ustubioglu, "Underwater image enhancement using contrast limited adaptive histogram equalization and layered difference representation," *Multimedia Tools Appl.*, vol. 80, no. 10, pp. 15067–15091, Apr. 2021.
[2] J. Wang, H. Wang, G. Gao, H. Lu, and Z. Zhang, "Single underwater image enhancement based on LP-norm decomposition," *IEEE Access*, vol. 7, pp. 145199–145213, 2019.
[3] R. Vijay Anandh and S. Rukmani Devi, "Visual enhancement of underwater images using transmission estimation and multi-scale fusion," *Comput. Syst. Sci. Eng.*, vol. 44, no. 3, pp. 1897–1910, 2023.
[4] P. Jia, B. Li, and X. Zhao, "Improved retinex underwater image enhancement algorithm based on HSI model," *Res. Exploration Lab.*, vol. 39, no. 12, pp. 1–4+14, 2020.
[5] K. H. Sanila, A. A. Balakrishnan, and M. H. Supriya, "Underwater image enhancement using white balance, USM and CLHE," in *Proc. Int. Symp. Ocean Technol. (SYMPOL)*, Dec. 2019, pp. 106–116.
[6] T. Zhang, Q. Li, Y. Li, and X. Liu, "Underwater optical image restoration method for natural/artificial light," *J. Mar. Sci. Eng.*, vol. 11, no. 3, p. 470, Feb. 2023.
[7] W. Zhang, X. Pan, X. Xie, L. Li, Z. Wang, and C. Han, "Color correction and adaptive contrast enhancement for underwater image enhancement," *Comput. Electr. Eng.*, vol. 91, May 2021, Art. no. 106981.
[8] Y. Wu, F. Li, H. Bai, W. Lin, R. Cong, and Y. Zhao, "Bridging component learning with degradation modelling for blind image super-resolution," *IEEE Trans. Multimedia*, to be published.
[9] H. Chen, H. He, and X. Feng, "Underwater image enhancement method based on color correction and dark channel prior," *J. Phys., Conf. Ser.*, vol. 2066, no. 1, Nov. 2021, Art. no. 012050.
[10] K. Liu and Y. Liang, "Underwater image enhancement method based on adaptive attenuation-curve prior," *Opt. Exp.*, vol. 29, no. 7, p. 10321, Mar. 2021.

[11] 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.
[12] 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.
[13] M. J. Islam, Y. Xia, and J. Sattar, "Fast underwater image enhancement for improved visual perception," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 3227–3234, Apr. 2020.
[14] J.-C. Lin, C.-B. Hsu, J.-C. Lee, C.-H. Chen, and T.-M. Tu, "Dilated generative adversarial networks for underwater image restoration," *J. Mar. Sci. Eng.*, vol. 10, no. 4, p. 500, Apr. 2022.
[15] 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.
[16] S. C. Hoo, H. Ibrahim, and S. A. Suandi, "ConvFaceNeXt: Lightweight networks for face recognition," *Mathematics*, vol. 10, no. 19, p. 3592, Oct. 2022.
[17] G. Wang, W. Yang, Y. Liu, X. Yang, Q. Wang, S. Yang, B. Feng, W. Sun, and H. Li, "Potato malformation identification and classification based on improved YOLOv3 algorithm," *Electronics*, vol. 12, no. 21, p. 4461, Oct. 2023.
[18] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 1–8.
[19] A. K. Dubey and V. Jain, "Comparative study of convolution neural network's ReLU and Leaky-ReLU activation functions," in *Applications of Computing, Automation and Wireless Systems in Electrical Engineering* (Lecture Notes in Electrical Engineering), vol. 553, S. Mishra, Y. Sood, and A. Tomar, Eds. Singapore: Springer, 2019, pp. 873–880.
[20] B. Yun, "A neural network approximation based on a parametric sigmoidal function," *Mathematics*, vol. 7, no. 3, p. 262, Mar. 2019. [Online]. Available: https://www.mdpi.com/about/announcements/784
[21] M. Lee, "Mathematical analysis and performance evaluation of the GELU activation function in deep learning," *J. Math.*, vol. 2023, pp. 1–13, Aug. 2023.
[22] C. Chen, B. Wu, and H. Zhang, "An image recognition technology based on deformable and CBAM convolution Res-net50," *J. Math.*, vol. 50, no. 1, pp. 1–8, 2023.
[23] J. Zhou, X. Li, T. Ding, C. You, Q. Qu, and Z. Zhu, "On the optimization landscape of neural collapse under MSE loss: Global optimality with unconstrained features," in *Proc. Int. Conf. Mach. Learning. PMLR*, 2022, pp. 27179–27202.
[24] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Miami, FL, USA, 2009, pp. 248–255.
[25] 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.
[26] G. Hou, J. Li, G. Wang, H. Yang, B. Huang, and Z. Pan, "A novel dark channel prior guided variational framework for underwater image restoration," *J. Vis. Commun. Image Represent.*, vol. 66, Jan. 2020, Art. no. 102732.
[27] K. G. Dhal, A. Das, S. Ray, J. Gálvez, and S. Das, "Histogram equalization variants as optimization problems: A review," *Arch. Comput. Methods Eng.*, vol. 28, no. 3, pp. 1471–1496, May 2021.
[28] 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 *Proc. Adv. Multimedia Inf. Process. PCM*, 2018, pp. 678–688.
[29] A. Saleh, M. Sheaves, D. Jerry, and M. R. Azghadi, "Adaptive uncertainty distribution in deep learning for unsupervised underwater image enhancement," 2022, *arXiv:2212.08983*.
[30] A. Horé and D. Ziou, "Is there a relationship between peak-signal-to-noise ratio and structural similarity index measure?" *IET Image Process.*, vol. 7, no. 1, pp. 12–24, Feb. 2013.
[31] P. Guo, L. He, S. Liu, D. Zeng, and H. Liu, "Underwater image quality assessment: Subjective and objective methods," *IEEE Trans. Multimedia*, vol. 24, pp. 1980–1989, 2022.
[32] S. A. Khan Tareen and R. H. Raza, "Potential of SIFT, SURF, KAZE, AKAZE, ORB, BRISK, AGAST, and 7 more algorithms for matching extremely variant image pairs," in *Proc. 4th Int. Conf. Comput., Math. Eng. Technol. (iCoMET)*, Mar. 2023, pp. 1–6.

YONGLI AN received the Ph.D. degree from the Institute of Information Science, Beijing Jiaotong University, in 2015. She is currently a Professor with North China University of Science and Technology, China. She was a Visiting Scholar with Oakland University, in 2011, and Griffith University, in 2018. Her research interests include key technologies of physical layer of wireless communication, signal processing, and image processing.

YAN FENG was born in 1997. He received the B.S. degree from the Qinggong College, North China University of Science and Technology, in 2020, where he is currently pursuing the master's degree. His research interests include machine vision and graphic image processing.

NA YUAN received the bachelor's degree from Heilongjiang University, in July 2011, and the master's degree from Hebei University of Technology, in January 2014. She is currently a Lecturer with Tangshan University. Her research interests include intelligent control, machine vision, and graphic image processing.

ZHANLIN JI (Member, IEEE) received the M.Eng. degree from Dublin City University, in 2006, and the Ph.D. degree from the University of Limerick, Ireland, in 2010.

He is a Professor with North China University of Science and Technology, China, and an Associated Researcher with the Telecommunications Research Centre (TRC), University of Limerick. He has authored/coauthored more than 100 research papers in refereed journals and conferences. His research interests include ubiquitous consumer wireless world (UCWW), the Internet of Things (IoT), cloud computing, big data management, and data mining.

IVAN GANCHEV (Senior Member, IEEE) received the Engineering and Ph.D. degrees (summa cum laude) from Saint-Petersburg University of Telecommunications, in 1989 and 1995, respectively. He is an International Telecommunications Union (ITU-T) Invited Expert and an Institution of Engineering and Technology (IET) Invited Lecturer, currently affiliated with the University of Limerick, Ireland; the University of Plovdiv "Paisii Hilendarski," Bulgaria; and the Institute of Mathematics and Informatics–Bulgarian Academy of Sciences, Bulgaria. He has participated in more than 40 international and national research projects. He has authored/coauthored one monographic book; three textbooks; four edited books; and more than 300 research articles in refereed international journals, books, and conference proceedings. He has served on the TPC for more than 400 prestigious international conferences/symposia/workshops. He is on the editorial board of and has served as a guest editor for multiple renowned international journals.