Underwater Image Enhancement Using Dual Regression U-Structure Network

MA Chi, Wang Shaofan, Hu Hui

Abstract—**Recent advancements in underwater image enhancement have demonstrated notable progress through the application of convolutional neural networks. Despite these achievements, there remains an untapped potential in fully exploring and leveraging image features essential for representing semantic knowledge, as well as local and global information within the realm of image enhancement. Faced with this problem, we introduce a novel approach, the Dual Regression U-Structure Network (DRUSN), aimed at delving into multi-level and multi-scale collaborative feature representation for enhancing underwater images. Our method incorporates dual regression branches dedicated to estimating ambient light and transmission for optimal underwater image enhancement. Notably, we introduce an innovative attention model, seamlessly integrating a residual learning strategy and an attention module within a unified framework. Subsequently, we construct an underwater image enhancement framework using a U-Structure formulation, featuring our proposed attention model. This attention model adeptly leverages the attention module to extract intricate local details from different-level features, effectively modeling local information. To harness the learned features, we design dual branches dedicated to learning ambient light and transmission, respectively. The DRUSN adeptly integrates the features form multi-level and multi-scale perspectives, demonstrating effectiveness in achieving superior underwater image enhancement. Rigorous quantitative and qualitative experiments conducted on several datasets show the superiority of our proposed DRUSN over several cutting-edge methods for enhancing underwater imagery.**

Index Terms—**Underwater image, image enhancement, multi-level features, multi-scale features, attention model**

I. Introduction

RECENTLY , the academic community has witnessed a marked increase in research of underwater image enhancement [6], [8], garnering attention within computer underwater vision applications. The primary objective of enhancing underwater images is to obtain details obscured by the attenuation and scattering of light, which are dependent on the wavelength and the distance traveled [44]. As light propagates through water, the inevitable occurrences of scattering and absorption result in varying degrees of color deviations. Traditional approaches to mitigate color casts and distortions in underwater scenes involve estimating light transmission and ambient light [23], [5], [38].While effective

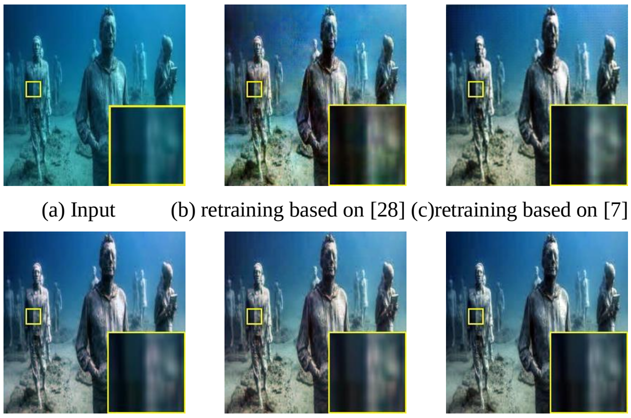
Manuscript received Oct 25, 2023; revised Oct. 15, 2024.

This paper is supported by Foundation of Guangdong Educational Committee under the Grant No. 2022ZDZX4052, 2021ZDJS082.

MA Chi is an associate professor of the School of Computer Science and Engineering, Huizhou University, Huizhou 516007, China (e-mail: machi@hzu.edu.cn).

Wang Shaofan is a engineer of the Derkee Information Co., Ltd, ShenZhen 518000, China (corresponding author to provide phone: 18241242603; e-mail: wsf19961230@163.com).

Hu Hui is a lecturer of the School of Computer Science and Engineering, Huizhou University, Huizhou 516007, China (e-mail: Huhui@hzu.edu.cn).



(a) Input (b) retraining based on [28] (c)retraining based on [7]

(d) retraining based on [24] (e) Ours (f) Reference

Fig. 1. A visual comparison of underwater image enhancement techniques The images are generated by us based on the concept mentioned in the quotation. Our model (e) can generate sharper and more natural color contents comparatively.

in certain scenarios, these methods face limitations in Challenging underwater conditions. Two key challenges include: 1) the accuracy of medium transmission estimation is hindered by multiple influencing factors, posing a nontrivial task, and 2) Such methodologies are heavily dependent on ancillary data, including but not limited to depth estimations and aquatic turbidity metrics, which are not invariably obtainable.

Recently, CNNs [43], [36], [10] demonstrated the capability of extracting image features, and this prevalent framework can leverage these distinctive features to address a multitude of visual processing tasks. Hou et al. [14] fulfill the transfer learning and image enhancement objectives employing a residual learning approach. Sun et al. [35] incorporate an encoder-decoder architecture. This architecture can effectively learn valuable underwater image features. The strategy has demonstrated efficacy in advanced visual tasks [32]. Since the deeper layers are capable of discerning a richer features, Liu et al. in [28] introduce a very deep network proposed by [20] in to the enhancement of underwater imagery. This method extract more information and attain significant improvements in the quality of the enhanced images as shown in Fig. 1 (b). Despite the advancements in underwater image enhancement methodologies, there remains significant potential for developing more potent frameworks aimed at extracting more resilient features. Within the domain of image enhancement, it is essential for image features to encapsulate semantic insights, as well as local and global details. Local details, which encompass textural and chromatic data, are instrumental for tasks such as color correction and light transmission assessment. Conversely, global details offer broader contextual semantics that are conducive to ambient light estimation.

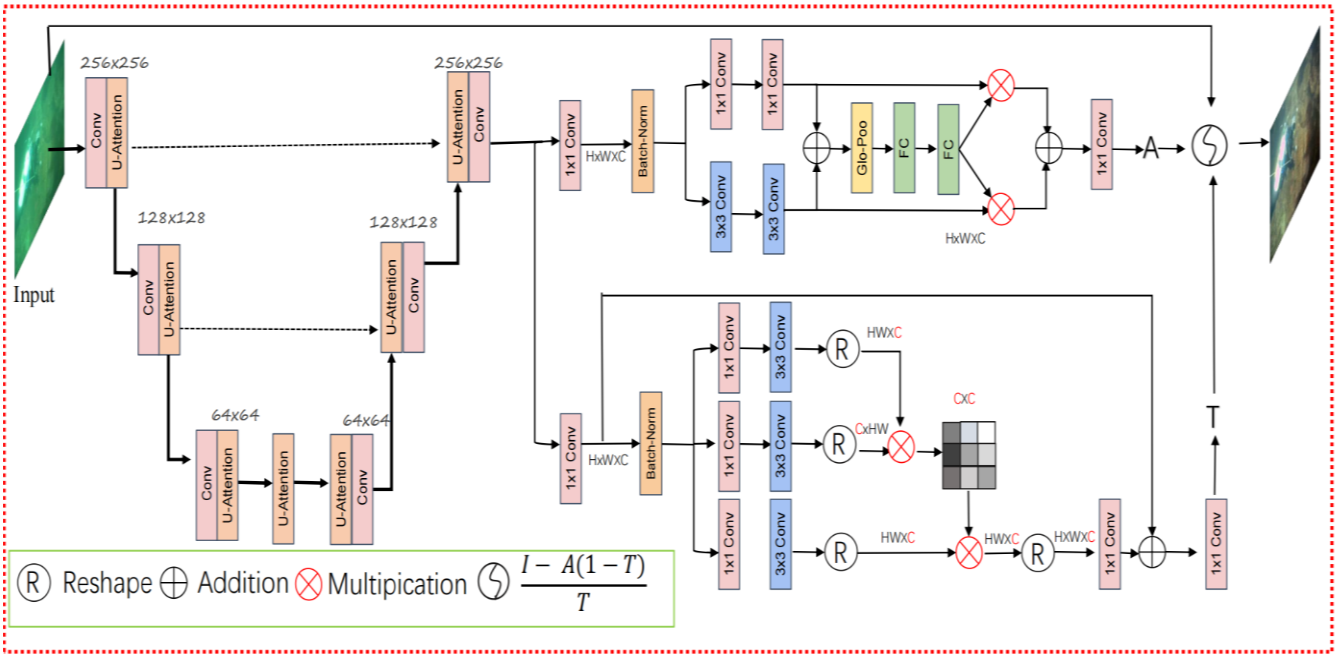


Fig. 2. Overview of our architecture for Dual Regression U-Structure Network.

With the advancement of Generative Adversarial Networks (GANs) [9], [37], there has been a surge in GAN-based methods extracting discriminative features [19], [42], [29] to generate visually appealing images. In [11], a multi-scale dense block has been incorporated into the generator for boosting the quality of underwater imagery, validating the effectiveness of multi-scale features in this context. Fabbri [7] employs a UGAN to enhance underwater images, utilizing Generative Adversarial Networks (GANs) to improve the visual appeal of underwater imagery. (refer to Fig. 1 (c)).

To address color bias issues, Li et al. [25] adopt a loosely adjustment approach to correct color distortions. Another approach, Water-Net, is proposed in [24] as an underwater image enhancement network, establishing a benchmark for training (depicted in Fig.1 (d)). In the quest to harness information across various color spaces, [23] presents a encoder network. The network introduce a multi-color space which can analyzes relationships within these diverse color spaces and leverages such attributes to boost the enhancement of underwater imagery.

However, it is noteworthy that these GAN-based methods predominantly emphasize semantic features and may overlook the significance of local information in image enhancement. This tendency sometimes results in generated images containing noticeable mesh effects, as illustrated in Fig. 1 (c).

Building on the insights gleaned from the analysis above, this paper introduces a novel approach: the Dual Regression U-Structure Network designed for underwater image enhancement. Our methodology leverages the U-Structure architecture to extract and harness multi-level features, effectively illustrating features from various levels and adeptly modeling local shallow features crucial for underwater image enhancement.

Within the U-Structure architecture, we integrate both residual learning and a pixel-wise attention module. This fusion facilitates the features extracting through channel and spatial attention strategy, augmenting the capacity to identify subtle details. Furthermore, a dual regression network is integrated to approximate ambient light and transmission respectively. The final step involves completing underwater image enhancement using the learned parameters, resulting in the generation of a clear image through our proposed model.

Through rigorous training of our model with profound supervision, we have conducted evaluations on an array of public benchmark datasets as well as real-world underwater imagery. The experimental data substantiate the efficacy and efficiency of our proposed Dual Regression U-Structure Network in achieving enhanced underwater image quality.

We summarize our contributions as follows:

- We introduce an innovative Attention module designed to effectively capture local multi-level and global information from underwater images. This is achieved by integrating residual learning with an attention mechanism within a U-structured framework.
- A dual regression net is designed for the underwater image enhancement, which can effectively capitalize on both multiple scales and levels features to achieve the estimation of ambient light and transmission in an iterative manner.
- Numerous benchmark dataset comparisons substantiate the efficacy of the Dual Regression U-Structure network, demonstrating its superiority over various cutting-edge methods for enhancing underwater imagery.

II. Related Work

Approaches to the enhancement and restoration of underwater imagery are broadly bifurcated into two distinct categories: traditional methodologies and deep learning-based methodologies.

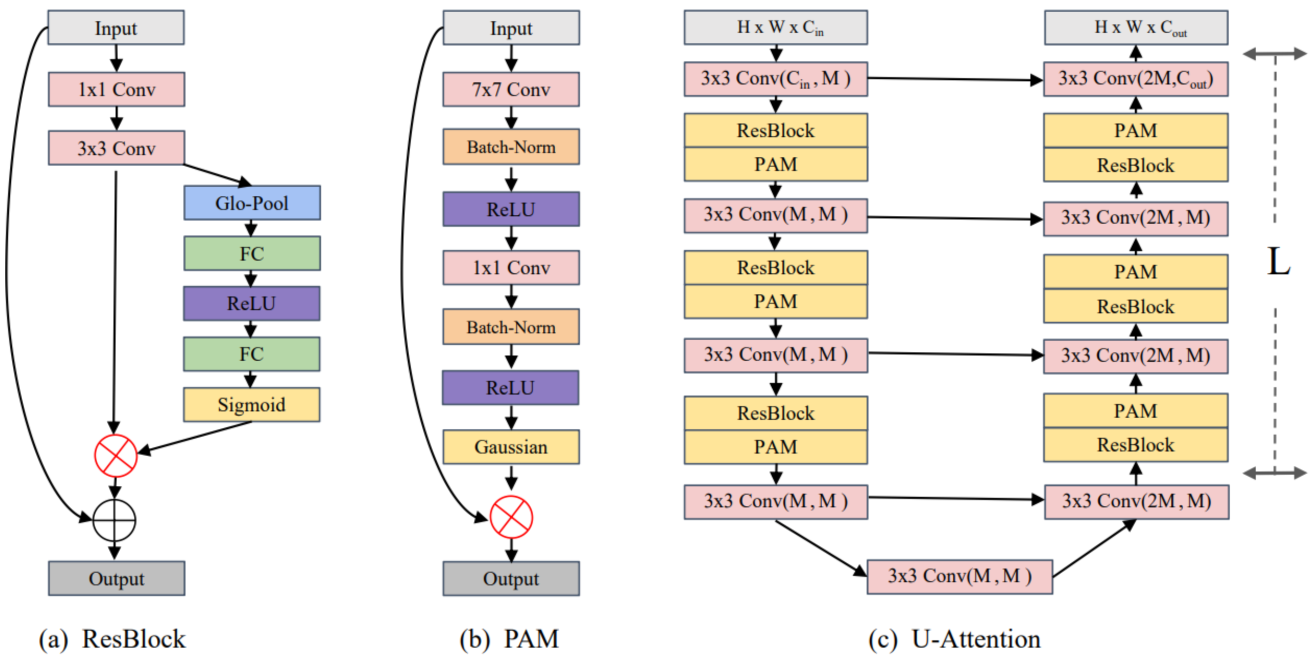


Fig. 3. Overview of our Attention module (c) with ResBlock (a) and pixel-wise attention module (b).

A. Traditional approaches

Enhancement of underwater imagery models that are traditional in nature frequently depend on models derived from physical phenomena, specifically for transmission and ambient light estimation. Methods like dark channel prior (DCP) [12,8] have been traditionally applied. DCP, initially designed for dehazing [30], [39], has been directly employed in enhancement of underwater imagery due to similarities between the two domains. However, the different environment characteristics, including absorption and scattering, pose challenges for the DCP algorithm's efficacy.

To address this, a universal approach based on underwater dark channel prior (UDCP) was introduced by [27], aiming for robust parameter estimation, including transmission and ambient light. [31] improved DCP by incorporating blurriness degree and differences in light absorption to estimate underwater ambient light and transmission. Given the rapid absorption of red light in underwater scenes, the red channel has been utilized as prior information in various studies [8]. Other researchers [1], [2] explored additional prior information to enhance parameter estimation.

Another approach involves enhancing underwater images by adjusting pixel values to improve visual quality. [3] proposed a method that increases visibility by adjusting pixel values using four weight maps. Building on this, [4] improved fusion-based underwater image enhancement strategies, deriving input images into color-compensated and white-balanced versions for separate enhancement with associated weight maps. Recognizing the importance of global information, [16] improved the quality of shallow-water images through the application of relative global histogram stretching (RGHS). [45] introduced a Bayesian retinex method, applying a maximum posteriori formulation to impose reflectance and illumination priors, effective for slightly distorted images. However, these methods may struggle to accurately estimate key underwater parameters in complex environments where prior information may not hold consistently.

B. CNN-based approaches

In recent times, researchers have harnessed the potent feature representation capabilities of CNN models to tackle the complexities inherent in enhancement of underwater imagery. These methods, which leverage Convolutional Neural Networks (CNNs), generally utilize end-to-end and data-driven training paradigms to directly generate clear underwater imagery. For instance, Yang et al. [41] presented a streamlined, adaptive network for feature fusion, facilitating easy enhancement through multiple adaptive feature fusion modules. Wang et al. [40] focused on integrating different color spaces within a CNN framework, employing the HSV color space for the enhancement of underwater images through deep learning approaches.

While these data-driven end-to-end methods excel in structural and texture recovery, they may introduce artifacts due to the absence of constraints from the imaging model. Recognizing the significance of depth information in light propagation underwater, some researchers have leveraged depth information as crucial prior knowledge to enhance the performance of the CNN-based model. Li et al. [26] presented a model which ingests natural scene photographs and depth data to synthesize corresponding underwater imagery. They trained two-stage fully convolutional networks for image restoration. In a similar vein, Li et al. [23] unified features from different spaces, enriching characteristic representations to explore relationships across various color spaces for effective underwater image enhancement.

While these CNN-based methods consider significant prior information, they rely on a data-driven strategy, overlooking the physical imaging characteristics inherent in the underwater environment. Therefore, we present a novel dual regression network that adeptly captures local multi-level and global information in underwater images by integrating residual learning and an attention module. The dual regression branches carefully consider key parameter characteristics, accurately estimating ambient light and transmission maps. Compared to other cutting-edge methods,

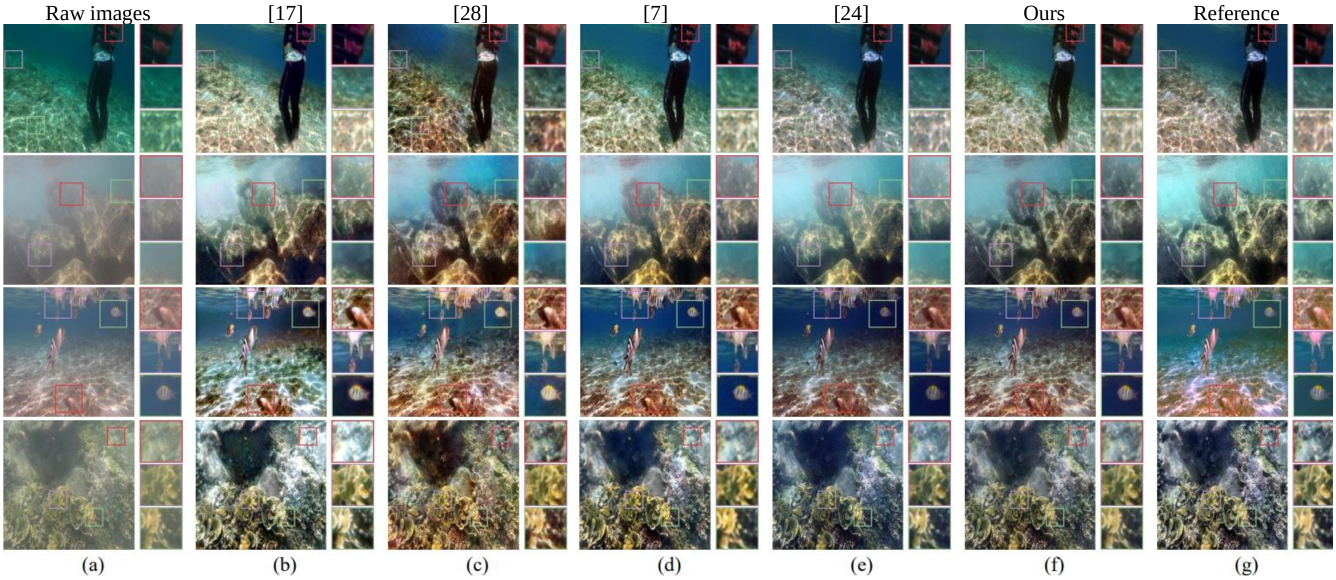


Fig. 4. Visual comparisons of the underwater image enhancement. The images are generated by us based on the concept mentioned in the quotation.

our approach exhibits superior performance.

III. Proposed Method

In this section, we will delineate the rationale and architectural framework of the proposed Dual Regression U-Structure Network (DRUN). Subsequently, we will explore the nuanced specifics of the DRUN structure, which is tailored for enhancing underwater images.

A. Our network

A successful underwater enhancement technique must tackle three essential components: 1)Global Semantic Understanding: The method needs to comprehend underwater images at a global semantic level. 2)Multiple levels and scales Feature Enrichment: The method should enrich useful feature representations by considering multiple levels and scales. 3)Optimal Utilization of Learned Features: The learned features must be effectively utilized for image enhancement.

To accomplish these objectives, Our model is designed as a Dual Regression U-Structure Network, illustrated in Fig. 2. The architecture of this model is bifurcated into two principal components: (1) U-Structure Model: The first component involves a U-structure model, integrating residual learning and an attention module. This combination forms a deeper network, and the attention module is adept at capturing features from multiple deep layers. (2)Dual Regression Network: The second component is a dual regression network, which separately estimates ambient light and transmission using corresponding topological networks.

The overall process involves utilizing the estimated parameters to enhance underwater images effectively.

1) Attention Module: As depicted in Fig. 3 (c), feature representation from multiple deep layers is crucial for effective image enhancement, as it enables the model to comprehend the image through multiple perspectives. In our methodology, we present an innovative module that combines residual learning with an attention mechanism to effectively capture multi-level features.

Our model's residual learning unit is crafted with inspiration from ResNet [13], which was originally conceived for the task of image recognition. The essence of our residual learning unit lies in the enhancement of feature representation by adaptively adjusting channel-wise feature responses, as depicted in Fig. 3 (a). This entails the creation of an attention map that highlights the varying emphasis on channel-wise features. Following this, the feature responses are recalibrated employing the derived attention vector. The mathematical expression of the residual learning block is detailed as follows:

$$F_{res} = F_{input} + Conv(F_{input} \times CA(Conv(F_{input})) \quad (1)$$

where F_{input} , F_{res} represent the input features and the features that have been refined by the residual learning module, respectively. The terms $Conv$ and CA refer to the fundamental convolutional operations and channel attention mechanisms embedded within the residual learning framework.

The residual learning block leverages channel-wise interactions to effectively enhance feature representation. However, it lacks an attention mechanism that can augment feature discrimination at the pixel level. We incorporate a pixel-wise attention module to enable our model to concentrate on pixel-level variations, as illustrated in Fig. 3 (b). Let X represent the feature set of the current layer. The pixel level attention is derived through a sequence of convolutional operations, succeeded by an activation function. The resulting response M_p encapsulates the spatial interdependencies for each element within the feature set X . The resulting weighted feature F_{pam} is calculated as follows:

$$F_{pam} = X \times Gaussian(M_p) \quad (2)$$

where X , F_{pam} signifies the input and the features that have been reassigned weights by the pixel-attention

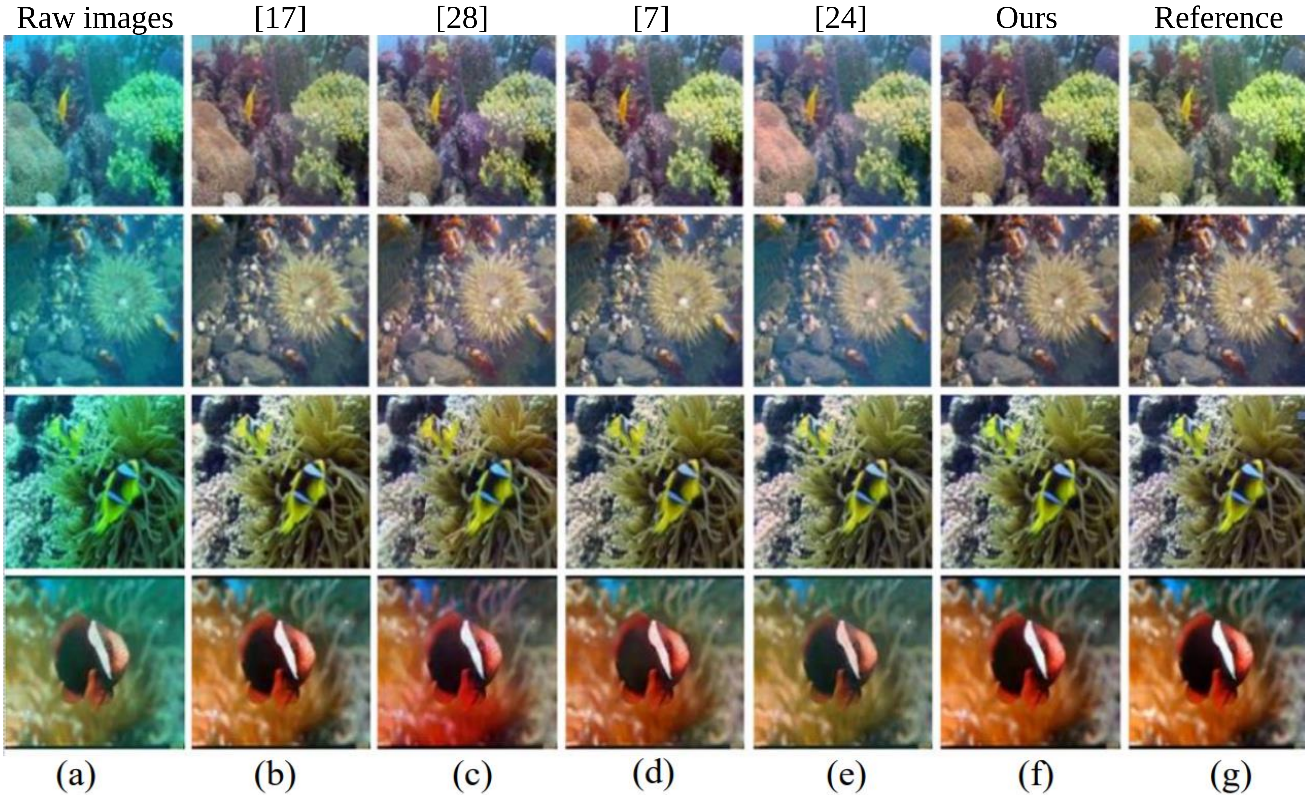


Fig. 5. Visual comparisons on Underwater Images. The images are generated by us based on the concept mentioned in the quotation.

mechanism. The final attention map, denoted as $Gaussian(M_p)$ is derived by applying a Gaussian gating function to the response map M_p.

Drawing from the aforementioned discourse, we introduce our attention module, designed to extract features across various hierarchical layers by amalgamating residual learning with a pixel-wise attention mechanism. The architecture of our attention module is depicted in Fig. 3 (c), where L indicates the profundity of the module, representing both the count of residual and attention blocks. C_{in}, C_{out} correspond to the input and output feature channel counts, respectively. Here, M signifies the channel count within the intermediate layers. Thus, our attention module, embodying a symmetric encoder-decoder configuration with depth L, ingests features $F_{in} \in W \times H \times C_{in}$ to distill and encode multi-level features $F_{out} \in W \times H \times C_{out}$. An increased value of L implies a greater number of residual learning blocks and attention modules, more profound layers, and a richer set of local features. The advantages of our attention module are two-pronged: 1) The amalgamation of residual learning and attention mechanisms effectively directs the network to investigate the channel-wise and spatial-wise feature inter dependencies. 2) The designed attention module, devoid of pooling operations, maintains local detailed information to the fullest extent possible.

2) U-Structure Network: U-Structure Network: The proposed U-Net model effectively captures multi-level features through residual learning. However, it might face challenges in capturing and integrating multi-scale features. As previously discussed, both semantic meaning and global information are essential. To address this, we introduce a U-Structure Network (USN) to leverage multi-scale information, incorporating the proposed attention module, as depicted in Fig. 2.

The proposed model commences by encoding the input through a succession of convolutional blocks. Each block has a convolutional layer, an attention module, and a strided layer, complemented by batch normalization and LReLU activation for the purpose of down-sampling. Consequently, each block expands its receptive field incrementally, allowing for the acquisition of features across various scales. The multi-scale features highlight the input's regulation from various perspectives. We integrate these multi-scale features using skip connections [15], enhancing the framework's comprehensive understanding of the input. In our empirical design, we utilize four blocks to achieve a threefold reduction in resolution, a strategy that has proven adequate for the image enhancement. The proposed USN offers two main advantages: 1) The multi-scale features gathered by the USN serve to augment the feature set of the U-attention mechanism; 2) It enhances network robustness through the iterative application of up- and down-sampling operations within the encoder-decoder framework, which complements the attention module in a multi-scale context. The U-structure process is articulated as follows:

$$F_u = USN(x) \tag{3}$$

where x, F_u denote the input underwater images and the learned robust features by the proposed U-structure net, which contain much global semantic and local detail information.

3) Dual Regression Network: The features learned by the proposed U-Structure Network encompass semantic knowledge, local details, and global information. Unlike

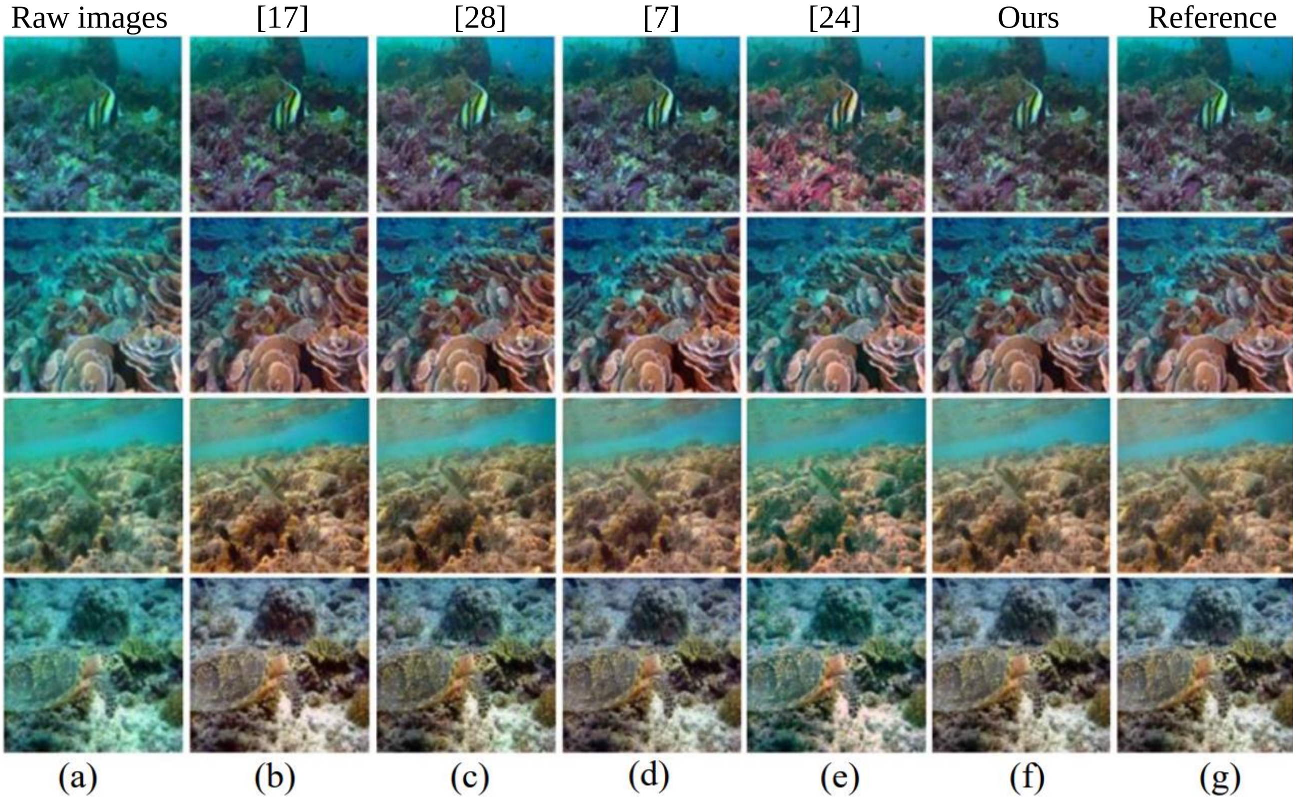


Fig. 6. Visual comparisons on Underwater Scenes. The images are generated by us based on the concept mentioned in the quotation.

previous methods that utilized these features to estimate parameters required for image enhancement or directly regress images through a unified topological network, we recognize that the parameters necessary for image enhancement hold distinct significance. Consequently, they demand different feature representations for accurate estimation. Ambient light estimation, for instance, necessitates more global semantic features to comprehensively understand the image. Conversely, transmission estimation may emphasize local feature representation, focusing on color and brightness information within limited regions. To address these diverse requirements, we introduce a Dual Regression Network, which is designed to independently estimate ambient light and transmission, tailoring the feature representation to the specific needs of each parameter.

The first branch is named AEN, which aims to predict the ambient light. We show the detail of the proposed AEN in Fig.2. We process the acquired features F_u using two sequential convolutional layers. The kernel of the first layer is 1 × 1. The kernel of the second one is 3 × 3. The outputs F_{one} , F_{two} can be obtain by:

$$F_{one}, F_{two} = Conv(F_u) \tag{4}$$

where F_u denotes the input features. F_{one} , F_{two} denote the learned two features with the different convolutional layers. Then, we fuse these two features with an addition operation, which can represent the features with different response fileds. In order to adeptly seize the holistic information embedded within the input features, we employ a global pooling operation to extract global characteristics from each channel. The last weight can be formulated as follows:

$$W = FC(FC(Gap(Add(F_{one}, F_{two})))) \tag{5}$$

where FC , Gap , Add denote the full connection, global pooling and addition operation respectively.

At last, we reweigh the features F_{one} , F_{two} with the learned weight W and add the reweighed features together, which can be used for the ambient light estimation. Since the features are enhanced by the global information, the enhanced features can well complete the global parameter estimation.

The second branch, named Transmission Estimation Network (TEN), is specifically used to predict the transmission of images in underwater environments. Transmission estimation relies more on spatial information. Hence, we introduce an effective network, TEN, to thoroughly explore spatial information and accurately estimate the transmission.

As illustrated in Fig.2, the model begins the process by extracting the input features through convolutional operations into three tensors: query (Q), key (K), and value (V) projections. Subsequently, we reshape the query and key projections, where $Q, K \in \Re^{HW \times C}$. For well explore the spatial correlation of these two projections, we generate a correlation map A with a dot-product interaction, where the size of map A is $\Re^{C \times C}$. At last, we also obtain the correlation between map A and tensor V with a dot-product interaction. The TEN processes are defined as:

$$T = Conv(X + Conv(Correlation(Q, K, V))) \tag{6}$$

TABLE I
QUANTITATIVE RESULTS COMPARING PSNR, SSIM, AND MSE METRICS.

Model	UD [31]	UI [31]	US [31]	UIEBD [24]	UGAN [7]
	PSNR/SSIM/MSE	PSNR/SSIM/MSE	PSNR/SSIM/MSE	PSNR/SSIM/MSE	PSNR/SSIM/MSE
IBLA [31]	16.66/0.772/1.68	16.09/0.634/2.29	19.65/0.723/1.31	15.31/0.647/3.33	15.80/0.627/2.57
RGHS [16]	15.91/0.789/1.99	16.51/0.714/1.65	18.42/0.751/1.09	19.72/0.839/1.06	16.30/0.708/1.77
ULAP [34]	17.36/0.771/1.48	18.39/0.707/1 15	19.93/0.747/0.78	16.33/0.758/2.24	18.02/0.700/1.32
FUnIE-GAN [17]	21.17/0.877/0.69	22.21/0.766/0.44	25.48/0.830/0.22	19.82/0.834/0.77	22.40/0.759/0.45
UresNet [28]	20.99/0.873/0.74	23.07/0.814/0.39	26.23/0.874/0.17	19.31/0.831/0.89	22.83/0.796/0.40
UGAN [7]	21.12/0.866/0.74	24.18/0.831/0.30	25.27/0.843/235	22.78/0.829/0.43	24.28/0.830/0.29
Water-Net [24]	20.80/0.864/0.86	22.50/0.817/0.44	22.65/0.820/0.46	23.71 /0.878/0.45	22.12/0.806/0.48
Ours	**21.24/0.880/0.65**	**24.67/0.833/0.29**	**26.29/0.879/0.16**	**23.78/0.886/0.42**	**24.88/0.834/0.28**

where X is the input features. $F_u, Q \in \Re^{HW \times C}$; $K \in \Re^{HW \times C}$; and $V \in \Re^{HW \times C}$ tensors are obtained after reshaping from the size $\Re^{H \times W \times C}$.

4) Image Enhanced: After the Dual Regression Network, we obtain two key parameters T and A, which are subsequently utilized for the processing of enhancement. We can learn from the imaging model proposed by [18], the image formation in an underwater setting can be mathematically expressed as follows:

$$I(x) = J(x)T(x) + A(1 - T(x)) \tag{7}$$

where $I(x)$ represents the underwater image and $J(x)$ denotes the corresponding clear image at location x which is the image we aim to enhance. The parameters of $T(x)$ and A are the key parameters at the position x ambient light of the holistic images in the underwater environment respectively, which also we try to estimate correctly. After the Dual Regression Network, we can enhance the underwater image with the learned key parameters T and A by deforming the equation Eq.7 as follows:

$$J(x) = \frac{I(x) - A(1 - T(x))}{T(x)} \tag{8}$$

Then, we can obtain the clear image $J(x)$.

B. Training

Let x denote an input underwater image , with y being the corresponding ground truth image of identical dimensions. Our proposed enhancement network is designed to learn a mapping function $F(x)$ that produces the enhanced underwater image $\hat{x}$, approximating the natural scene image y .

During the training phase, a pixel-wise loss [22] is employed for supervised learning, with the pixel loss defined as follows:

$$l_{pix} = \rho(\hat{x} - y) \tag{9}$$

In this context, the pixel penalty function is given by $\rho(x) = \sqrt{x^2 + \varepsilon^2}$, where x , y represent the enhanced and clear images, respectively. The value of ε is assigned as $1e^{-3}$. Essentially, we can impose additional constraints at various scales to perform multi-resolution image enhancement, but for simplicity, we only refine the underwater images at their original resolution.

We also use the loss l_{per} to direct the perceptual alignment between the outputs $\hat{x}$ and the corresponding y . The perceptual los l_{per} is derived from a VGG-19 network [33] that has been pre-trained on the ImageNet dataset [21]. We select some key features at the middle layers to compute the difference between these two images. The l_{per} is defined as follows:

$$l_{per} = \sum_{i=0}^{4} \varphi_i(\hat{x}) - \varphi_i(\hat{y}) \tag{10}$$

where $\varphi_i(\cdot)$ denotes extracted the features with the VGG-19 network, which was pre-trained on the ImageNet dataset, at the convolutional layer i .

By combining these two loss functions, the optimization problem becomes: (11).

$$l_c = \lambda_1 l_{pix} + \lambda_2 l_{per} \tag{11}$$

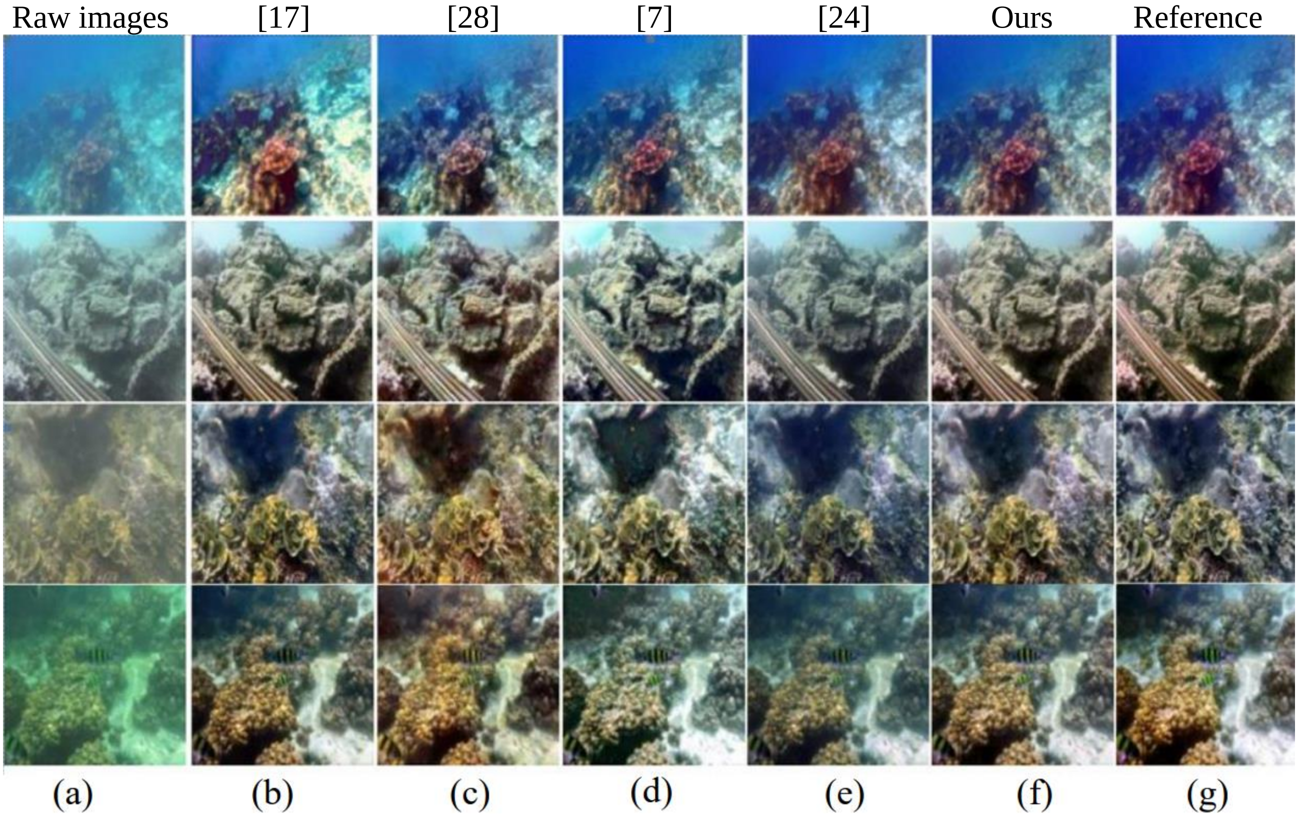


Fig. 7. Visual comparisons on UIEBD datasets. The images are generated by us based on the concept mentioned in the quotation.

where $\lambda_{1,2}$ represents pre-specified hyperparameters for the two loss components. Specifically, λ_1 is assigned a value of 1, while λ_2 is given a value of 0.01. The entire training regimen is executed via back propagation.

IV. EXPERIMENTS

In this segment, we exhibit a comparative analysis with various cutting-edge underwater image enhancement models that encompass both traditional and deep learning methodologies, namely IBLA [31], RGHS [16], ULAP [34], FUnIE-GAN [17], UresNet [28], UGAN [7], and Water-Net [24]. The evaluation is carried out utilizing five publicly accessible datasets: Underwater Dark [31], Underwater ImageNet [31], Underwater Scenes [31], UIEBD [24] and UGAN [7]. Underwater Dark: For the Underwater Dark dataset, we have curated a selection of 5,000 images for training and 500 for validation from various underwater environments. 1) Underwater ImageNet dataset: Comprising 6,128 pairs of underwater images, we randomly selected 2,000 pairs for training and 400 for validation. 2) UIEBD dataset: Out of the total 890 pairs of underwater images, 790 pairs were designated for training, with the rest held for testing. We fine-tuned our training strategy by setting the momentum to 0.9 and applying weight decay at a rate of 0.0001. Furthermore, we initialized a conservative learning rate of 0.001 for the training phase.

A. Comparison

To ensure an equitable comparison, we retrain the models in question and opt for the one that demonstrates the most superior performance for the comparative analysis. Initially, we conduct both the training and testing phases of our model using the Underwater Dark dataset [31], where the level of image distortion is low. As illustrated in Fig. 4, our model effectively restores detailed image information and corrects color cast induced by deep water. In contrast, traditional models like IBLA, RGHS, and ULAP struggle to achieve satisfactory enhancement due to limited parameter estimation. Leveraging the powerful learning capabilities of deep learning methods, our model, with its extensive parameters, outperforms others, resulting in enhanced images with rich color information that closely resemble reference images.

We further evaluate our model on the Underwater ImageNet dataset [31], which exhibits more serious image degradation compared to the Underwater Dark dataset [31]. The results, as shown in Fig.5, demonstrate that our model excels in restoring texture details and color information. Among other learning-based methods, FunIE-GAN [17] stands out for its performance, leveraging vision-driven behaviors to enhance degraded underwater images. However, some learning methods introduce artifacts and unnatural color information, as observed in the enhanced results of UGAN [7] and Water-Net [24] in Fig. 5.

The US dataset [31] poses the most challenging underwater scenes, making the enhancement task particularly difficult. Despite this complexity, our model, when compared to other models, achieves superior performance in color and contrast enhancement, as depicted in Fig. 6.

Finally, we conduct training and evaluation of our model on an authentic underwater dataset. The outcomes presented in Fig. 7 demonstrate that our model surpasses other enhancement techniques when utilized on actual underwater imagery. The key advantage lies in our model's ability to effectively leverage multi-level and multi-scale features for

TABLE II
NUMERICAL COMPARISON OUTCOMES WITHIN THE UIEB DATASET, WHERE WE RETRAIN OUR MODEL WITH DIFFERENT DEPTH (L= 2, 3, 4, 5) MODULE FOR COMPARISON.

	L=1	L=2	L=3	L=4
PSNR	23.36	23.78	23.83	23.88
SSIM	0.866	0.886	0.889	0.889

enhanced image representation, successfully eliminating color cast and generating vibrant images. This observation underscores the importance of incorporating valuable feature learning into image enhancement.

Water-Net, which conducts a comprehensive perceptual study focusing on semantic meaning, supports the idea that semantic learning contributes to improved underwater image enhancement. On the other hand, UGAN performs less effectively than our model, relying on a single U-structure that tends to prioritize global feature integration. This approach overlooks the fusion of multi-level features at specific scales, resulting in noticeable colored moir in the output images, as seen in the seventh column of Fig.7. UresNet [29] architecture capitalizes on a profoundly deep network to extract multi-level features, thereby attaining competitive performance metrics. In contrast, our proposed model harnesses a novel U-Attention mechanism designed to assimilate features from a multiplicity of perspectives. This approach optimizes the extraction and integration of multilevel and multi-scale features, culminating in an augmented image representation that surpasses existing methods.

From the discussion above, we notice that:

- Our model maximizes the utilization of local multi-level feature information by integrating residual learning and an attention module within a U-structure formulation. This approach enhances feature representation, providing substantial benefits for image enhancement.
- Our model acknowledges the nuances in key parameter estimation and addresses them with a dual regression network. The Ambient Estimation Network (AEN) focuses on global semantic information, accurately estimating ambient light, while the Transmission Estimation Network (TEN) places greater emphasis on spatial information, leading to precise transmission estimation.
- Relative to other methods predicated on learning, our model distinguishes itself through its adeptness in color correction and contrast enhancement. This achievement underscores the significance of our strategy, where the estimation of transmission and ambient light, aligned with their respective feature considerations, plays a crucial role in enhancing underwater information

B. Quantitative comparison

We evaluate the efficacy of our model through three established quantitative metrics: the Structural Similarity Index (SSIM), the Peak Signal-to-Noise Ratio (PSNR), and the Mean Square Error (MSE). The quantitative analysis detailed in Table I. illustrates that our model attains markedly higher scores in PSNR, SSIM, and MSE when juxtaposed with current leading models. This underscores the proficiency of the Dual Regression Network in furnishing robust feature representations for the precise rendition of underwater imagery. Contrasted with the top-performing model, DRUSN, our proposed model leads in PSNR, SSIM, and MSE across the entire spectrum of five datasets. Specifically, DRUSN manifests the optimal performance with respect to PSNR/SSIM/MSE in comparison to its peers. Of particular note, our proposed model bolsters the structural integrity and enhances the SSIM metric, outperforming the current benchmark WaterNet model by an average margin of 0.8%.

C. Effect of the Depth

To further evaluate the efficacy of the depth of attention mechanism, we trained the proposed models with different depths of the attention model, (i.e. 2, 3, 4, 5). The model with deeper layers implies that the model exploits more multi-level feature domains. However, the number of parameters is too high to need more computation resources, which may hamper the performance of our model.

V. CONCLUSION

This study presents a groundbreaking attention mechanism aimed at capturing intricate local and global details within the context of underwater imagery. By acknowledging the unique traits of pivotal parameters in enhancing underwater images, we have developed a dual regression framework that is meticulously designed to identify and leverage features critical for the estimation of ambient light and transmission. Our innovative dual regression model has set new benchmarks within the domain of enhancement of underwater imagery. The comparative analysis underscores the superiority of our Dual Regression model, which incorporates U-Attention, in harnessing the interplay between features across various levels and scales. The U-Attention mechanism, proficient in extracting both local and global features from high to low levels, provides a novel approach to feature representation research, showing great promise for the modeling of image data. Nonetheless, our method may sometimes show noticeable discontinuities. Looking ahead, we plan to delve into the potential of attention mechanisms to mitigate this challenge in our future endeavors.

REFERENCES

[1] D. Akkaynak and T. Treibitz. “A revised underwater image formation model,” 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, pp6723-6732, 2018.

[2] D. Akkaynak and T. Treibitz. “Sea-thru: A method for removing water from underwater images,” 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, pp1682-1691, 2019.

[3] C. Ancuti, C. O. Ancuti, T. Haber, and P. Bekaert. “Enhancing underwater images and videos by fusion,” 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, pp81-88, 2012.

[4] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and P. Bekaert. “Color balance and fusion for underwater image enhancement,” IEEE Trans Image Process, vol. 27, no.1, pp379-393, 2017.

[5] 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, pp2822-2837, 2021.
[6] P. Drews, E. R. Nascimento, S. Botelho, and M. Campos. “Underwater depth estimation and image restoration based on single images,” in IEEE Computer Graphics and Applications, vol. 36, no.2, pp24-35, 2016.
[7] C. Fabbri, M. J. Islam, and J. Sattar. “Enhancing underwater imagery using generative adversarial networks,” 2018 IEEE International Conference on Robotics and Automation (ICRA), pp7159-7165, 2018.
[8] A. Galdran, D. Pardo, A. Picon, and A. Alvarez-Gila. “Automatic redchannel underwater image restoration,” Journal of Visual Communication and Image Representation, vol 26, pp132-145, 2015.
[9] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. “Generative adversarial nets,” Proceedings of the 27th International Conference on Neural Information Processing Systems, vol. 2, pp2672-2680, 2014.
[10] Ren Guo-Xi, “Research on a convolutional neural network method for modulation waveform classification,” IAENG International Journal of Computer Science, vol. 50, no.3, pp875-882, 2023.
[11] Y. Guo, H. Li, and P. Zhuang. “Underwater image enhancement using a multiscale dense generative adversarial network,” in IEEE Journal of Oceanic Engineering, vol. 45, no.3, pp862-870, 2020.
[12] K. He, J. Sun, and X. Tang. “Single image haze removal using dark channel prior,” in IEEE transactions on pattern analysis and machine intelligence, vol. 33, no.12, pp2341-2353, 2011.
[13] K. He, X. Zhang, S. Ren, and J. Sun. “Deep residual learning for image recognition,” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp770-778, 2016.
[14] M. Hou, R. Liu, X. Fan, and Z. Luo. “Joint residual learning for underwater image enhancement,” 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, pp4043-4047, 2018.
[15] Q. Hou, M. -M. Cheng, X. Hu, A. Borji, Z. Tu and P. H. S. Torr, “Deeply Supervised Salient Object Detection with Short Connections,” in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 4, pp815-828, 2019.
[16] D. Huang, Y. Wang, W. Song, J. Sequeira, and S. Mavromatis. “Shallowwater image enhancement using relative global histogram stretching based on adaptive parameter acquisition,” In 2018 International conference on multimedia modeling , pp453-465, 2018.
[17] M. J. Islam, Y. Xia, and J. Sattar. “Fast underwater image enhancement for improved visual perception,” in IEEE Robotics and Automation Letters, vol. 5, no.2, pp3227-3234, 2020.
[18] J. S. Jaffe. “Computer modeling and the design of optimal underwater imaging systems,” in IEEE Journal of Oceanic Engineering, vol. 15, no.2, pp101-111, 1990.
[19] A. Jamadandi and U. Mudenagudi. “Exemplar-based underwater image enhancement augmented by wavelet corrected transforms,” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp11-17, 2019.
[20] J. Kim, J. K. Lee, and K. M. Lee. “Accurate image super-resolution using very deep convolutional networks,” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp1646-1654, 2016.
[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton. “Imagenet classification with deep convolutional neural networks,” Communications of the ACM, vol. 60, no.6, pp84-90, 2017.
[22] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang. “Deep laplacian pyramid networks for fast and accurate super-resolution,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, pp5835-5843, 2017.
[23] C. Li, S. Anwar, J. Hou, R. Cong, C. Guo, and W. Ren. “Underwater image enhancement via medium transmission-guided multi-color space embedding,” in IEEE Transactions on Image Processing, vol. 30, pp4985-5000, 2021.
[24] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao. “An underwater image enhancement benchmark dataset and beyond,” in IEEE Transactions on Image Processing, vol. 29, pp4376-4389, 2019.
[25] C. Li, J. Guo, and C. Guo. “Emerging from water: Underwater image color correction based on weakly supervised color transfer,” in IEEE Signal Processing Letters, vol. 25, no. 3, pp323-327, 2018.
[26] 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,” in IEEE Robotics and Automation Letters, vol. 3, no. 1, pp387-394, 2017.
[27] Z. Liang, X. Ding, Y. Wang, X. Yan, and X. Fu. “Gudcp: Generalization of underwater dark channel prior for underwater image restoration,” in IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no.7, pp4879-4884, 2022.
[28] P. Liu, G. Wang, H. Qi, C. Zhang, H. Zheng, and Z. Yu. “Underwater image enhancement with a deep residual framework,” in IEEE Access, vol. 7, pp94614-94629, 2019.
[29] X. Liu, Z. Gao, and B. M. Chen. “Mlfcgan: Multilevel feature fusionbased conditional gan for underwater image color correction,” in IEEE Geoscience and Remote Sensing Letters, vol. 17, no.9, pp1488-1492, 2019.
[30] Y. Liu, J. Shang, L. Pan, A. Wang, and M. Wang. “A unified variational model for single image dehazing,” in IEEE Access, vol. 7, pp15722-15736, 2019.
[31] Y.-T. Peng and P. C. Cosman. “Underwater image restoration based on image blurriness and light absorption,” in IEEE transactions on image processing, vol. 26, no.4, pp1579-1594, 2017.
[32] O. Ronneberger, P. Fischer, and T. Brox. “U-net: Convolutional networks for biomedical image segmentation,” Medical Image Computing and Computer-Assisted Intervention--MICCAI 2015, pp234-241, 2015.
[33] K. Simonyan and A. Zisserman. “Very deep convolutional networks for large-scale image recognition,” 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings, 2015.
[34] 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,” Advances in Multimedia Information Processing-PCM 2018, pp678-688, 2018.
[35] X. Sun, L. Liu, Q. Li, J. Dong, E. Lima, and R. Yin, “Deep pixel-to-pixel network for underwater image enhancement and restoration,” IET Image Processing, vol. 13, no.3, pp469-474, 2019.
[36] Kang Tan, Linna Li, Qiongdan Huang, “Image manipulation detection using the attention mechanism and faster r-cnn,” IAENG International Journal of Computer Science, vol. 50, no.4, pp1261-1268, 2023.
[37] P. M. Uplavikar, Z. Wu, and Z. Wang. “All-in-one underwater image enhancement using domain-adversarial learning,” In CVPR Workshops, pp1-8, 2019.
[38] K. Wang, L. Shen, Y. Lin, M. Li, and Q. Zhao. “Joint iterative color correction and dehazing for underwater image enhancement,” IEEE Robotics and Automation Letters, vol. 6, no.3, pp5121-5128, 2021.
[39] W. Wang, F. Chang, T. Ji, and X. Wu. “A fast single-image dehazing method based on a physical model and gray projection,” IEEE Access, vol. 6, pp5641-5653, 2018.
[40] Y. Wang, J. Guo, H. Gao, and H. Yue. “Uiec 2-net: Cnn-based underwater image enhancement using two color space,” Signal Processing: Image Communication, vol. 96, pp116250, 2021.
[41] H.-H. Yang, K.-C. Huang, and W.-T. Chen. “Laffnet: A lightweight adaptive feature fusion network for underwater image enhancement,” 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, pp685-692. IEEE, 2021.
[42] M. Yang, K. Hu, Y. Du, Z. Wei, Z. Sheng, and J. Hu. “Underwater image enhancement based on conditional generative adversarial network,” Signal Processing: Image Communication, vol. 81, pp115723, 2020.
[43] K. Zheng, ZG. **a, Y. Zhang, X. Xu, Y. Fu. “Speech emotion recognition based on multi-level residual convolutional neural networks,” Engineering Letters, vol. 28, no.2, pp559-565, 2020.
[44] Y. Zhou, Q. Wu, K. Yan, L. Feng, and W. Xiang. “Underwater image restoration using color-line model,” in IEEE Transactions on Circuits and Systems for Video Technology, vol. 29, no.3, pp907-911, 2018.
[45] P. Zhuang, C. Li, and J. Wu. “Bayesian retinex underwater image enhancement,” Engineering Applications of Artificial Intelligence, vol. 101, pp104171, 2021.

MA Chi, he received his Ph.D. in Dongling School of Economics and Management at the University of Science and Technology Beijing. Since 2010, he has been working as Associate Professor in School of Software Engineering at University of Science and Technology Liaoning. From 2019 to now, he has worked in the school of computer science and engineering at Huizhou University. His research interests include pattern recognition and data mining.

Wang Shaofan, he graduated from Zaozhuang University. Since 2019, he has been studying for master degree in the field of computer science and technology at School of Computer Science and Software Engineering, University of Science and Technology Liaoning. Since 2023, he has worked as an engineer at the the Derkee Information Co., Ltd. His research direction is deep learning and data mining.

Hu Hui, she received her Master's degree in computer applied technology with Nanjing Tech University, Nanjing, China. Since 2001, she has been working as lecturer in department of computing and information systems at Yancheng Institute of Technology. From 2011, she has worked in the school of computer science and engineering at Huizhou university. Her research interests include artificial intelligence, computer vision, image processing, pattern recognition.