


# Adaptive Learning Attention Network for Underwater Image Enhancement

Shiben Liu , Huijie Fan , Sen Lin , *Member, IEEE*, Qiang Wang , Naida Ding , and Yandong Tang 

**Abstract**—Underwater images suffer from color casts and low illumination due to the scattering and absorption of light as it propagates in water. These problems can interfere with underwater vision tasks, such as recognition and detection. We propose an adaptive learning attention network for underwater image enhancement based on supervised learning, named LANet, to solve these degradation issues. First, a multiscale fusion module is proposed to combine different spatial information. Second, we design a novel parallel attention module(PAM) to focus on the illuminated features and more significant color information coupled with the pixel and channel attention. Then, an adaptive learning module(ALM) can retain the shallow information and adaptively learn important feature information. Further, we utilize a multinomial loss function that is formed by mean absolute error and perceptual loss. Finally, we introduce an asynchronous training mode to promote the network's performance of multinomial loss function. Qualitative analysis and quantitative evaluations show the excellent performance of our method on different underwater datasets. The code is available at: <https://github.com/LiuShiBen/LANet>.

**Index Terms**—Attention mechanism, computer vision for automation, deep learning methods, underwater image enhancement.

## I. INTRODUCTION

CURRENTLY, with the rapid development of artificial intelligence, Autonomous Underwater Vehicles(AUVs) [1] are widely used in underwater scene analysis [2], oceanic biological monitoring and human-robot collaboration [3]. Underwater vehicles that use high-end cameras are restricted to obtaining underwater images with more robust perception by the poor underwater imaging environment and lighting conditions [4]. So this critically affects the performance of vision-based tasks in water. An available solution to restore potentially clean images is

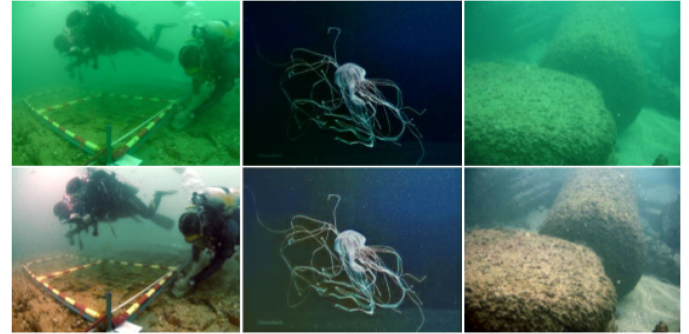


Fig. 1. Real-world underwater image enhancement by LANet. Top row: degraded underwater images; bottom row: the corresponding results of our model.

tremendous for improving the visual quality of images captured in water and accurately understanding the underwater world.

The influence of underwater image degradation on the underwater formation model is mainly in following aspects: (1) Light scattering in water affects the quality of underwater images with low illumination and blur, including forward and back scattering. Forward scattering (the reflected light randomly deviates from the propagation path and is received by the camera) is generally negligible. Back scattering (part of the light is reflected by water or suspended particles before it reaches the target and then received by the camera) is dominant in the scattering process. (2) In water, light attenuation depends on the light wavelength and imaging range, leading to color casts. The absorption of red light is the strongest, and blue-green light is the weakest. Fig. 1 shows the visual effects of degraded underwater images affected by color casts, low illumination, and detail blur. The results demonstrate that our method can obtain clear underwater images.

Considering the uneven distribution of illumination features and unbalanced color information across the underwater image. Thus, we design a PAM including channel attention(CA) and pixel attention(PA) in parallel, which better pays attention to illumination features and color information. The ability of network to extract features is relatively weak with the increase of layers [5]. We propose an adaptive learning module(ALM), it can effectively learn representative features and transfer shallow information to deeper network layers. Finally, we put forward an adaptive learning attention network LANet to obtain underwater images with bright color and high illumination through supervised learning.

Manuscript received November 8, 2021; accepted February 23, 2022. Date of publication March 7, 2022; date of current version March 15, 2022. This letter was recommended for publication by Associate Editor Blair Thornton and Editor Cesar Cadena Lerma upon evaluation of the reviewers' comments. This work was supported in part by the National Natural Science Foundation of China under Grants 61991413, U20A20200, and 62073205, and in part by the Youth Innovation Promotion Association of the Chinese Academy of Sciences under Grant 2019203. (Corresponding author: Huijie Fan.)

Shiben Liu, Huijie Fan, Sen Lin, Naida Ding, and Yandong Tang are with the State Key Laboratory of Robotics and Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China, and with the Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang 110016, China (e-mail: liushiben310@163.com; fanhuijie@sia.cn; lin\_sen6@126.com; dingnaida@sia.cn; ytang@sia.cn).

Qiang Wang is with the Key Laboratory of Manufacturing Industrial Integrated, Shenyang University, Shenyang 110016, China, and with the State Key Laboratory of Robotics and Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China (e-mail: wangqiang@sia.cn).

Digital Object Identifier 10.1109/LRA.2022.3156176

The main contributions of this paper are as follows:

- 1) We formulate an end-to-end fully-convolutional network LANet for underwater image enhancement. We propose a PAM, which integrates pixel attention and channel attention by using the parallel connection. This module focuses on the illumination features and more significant color information.
- 2) ALM is designed to adaptive learn the most representative features and transfer shallow information into deep layers. In addition, we introduce an asynchronous training mode to improve network's performance.
- 3) We compare it with state-of-the-art methods. The results demonstrate that LANet performs incredibly well in regions with serious color casts and low illumination. In addition, the enhanced images prominently heighten the performance of object detection and salient object detection.

## II. RELATED WORK

Due to the powerful extraction ability of deep learning technology, it has achieved extremely competitive performance in low-level vision tasks. In this section, according to the different network models, it can be divided into CNN-based [6], [7] and GAN-based methods [2], [8], [9] for underwater image enhancement. Finally, we introduce the status of underwater datasets.

### A. Underwater Image Enhancement Methods

1) *GAN-Based Methods*: GANs can produce better output through the game between the generative model and discriminant model. In [10], the residual multiscale dense block was introduced into the generator to extract more details and make full use of features. Chen *et al.* [11] proposed multibranch discriminator that can adaptively improve the underwater visual quality. There are often significant differences between synthetic and real-world datasets, affecting algorithm performance. Li *et al.* [12] introduced a weakly supervised color migration method to correct color distortion. Uplavikar *et al.* [13] proposed a model that can reverse learning the content image features using the learned domain agnostic features to generate enhanced underwater images. Some researchers [14], [15] employed the image formation model to estimate the clear underwater images. However, the color restoration of underwater images are incomplete in vision.

2) *CNN-Based Methods*: CNNs can effectively extract different feature expressions from low-level details to high-level semantics under supervision information and then use these discriminative features to achieve different tasks. Hou *et al.* [16] developed a novel framework to jointly performing residual learning on transmission and image domains for underwater scene enhancement. Liu *et al.* [17] introduced a unified paradigm for transmission-based image enhancement tasks, which alternately refines the transmission map to restore the potential transparent scene layer. For low-illumination and turbid underwater images, Sun *et al.* [18] designed an enhanced model based on an asymmetric encoder-decoder to learn the mapping from low-quality

to high-quality underwater images. Jamadandi *et al.* [19] used wavelet pooling and nonpooling as network components to perform progressive whitening and coloring transformation of underwater images. More convolutional layers can learn better semantic features. Liu *et al.* [20] applied VDSR [21] to the field of underwater image enhancement based on the idea of residual learning and used edge difference loss (EDL) to improve the detail enhancement capabilities of the deep learning model.

It is worth noting that more learn-based methods ignore the relevance of illumination features and color information in underwater image. In addition, shallow network can not effectively learn the semantic information of underwater images. To solve these questions, we propose an adaptive learning attention network of removing color casts and improving illumination.

### B. Datasets

Under the influence of complex underwater environment, it is difficult to obtain ground truth underwater images. Thus, some researchers synthesize datasets through networks or collect datasets with underwater robots. Fabbri *et al.* [22] adopted CycleGAN [23] to synthesize a large number of pairs of images. Li *et al.* [24] synthesized ten types of underwater images using the NYU-v2 indoor dataset [25]. Li *et al.* [26] utilized many classical algorithms to process degraded images and then manually filtered out clear images. Islam *et al.* [8] established the EUVP dataset, which contained a large number of paired and unpaired underwater images with poor perceptual quality. Qi *et al.* [27] established an underwater image coenhancement dataset (UICoD) with the ground truth images. UICoD includes 1,272 images from 46 different video sequences. The SQUID dataset [28] contains 57 underwater stereo pairs taken from four different dive sites in Israel. The above datasets are utilized for network training, improving the quality of underwater images, and further laying the foundation for data-driven underwater image enhancement.

## III. METHOD

We proposed LANet to improve illumination and remove color casts in underwater images. Fig. 2 shows the framework of LANet is divided into five parts. (1) Multiscale fusion module(MFM) is employed to strengthen the effective connection of spatial information between three branches. (2) PAM can effectively integrate color information and illumination feature. (3) ALM adaptively learns feature weights and transfers shallow information to the deeper network structure. (4) Group structure(GS) increases the depth and expressiveness of the LANet. (5) Multinomial loss functions are constructed to improve the network's performance.

### A. Multiscale Fusion Module

To better extract the global features of the input image, we design three branches to learn the receptive field of the input image by the convolution kernel sizes are  $3 \times 3$ ,  $5 \times 5$ , and  $7 \times 7$ , respectively. The large convolution kernel size may better extract global features in the image. The output of three

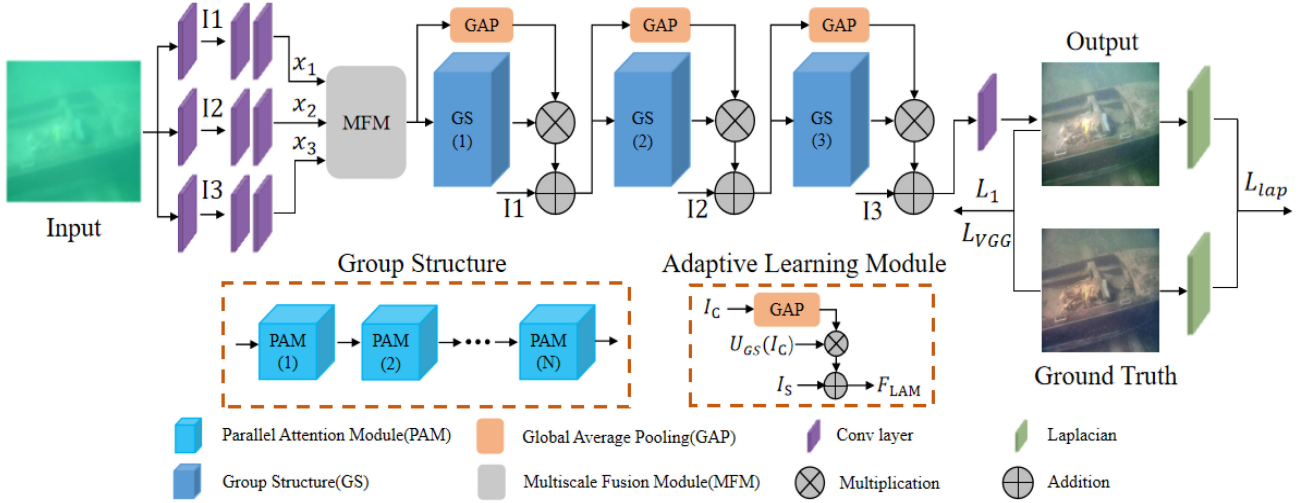


Fig. 2. Framework of LANet.  $x_1$ ,  $x_2$  and  $x_3$  denote the outputs of the three branches.  $L_1$ ,  $L_{VGG}$  and  $L_{lap}$  indicate mean absolute error loss, perceptual loss and laplace loss.

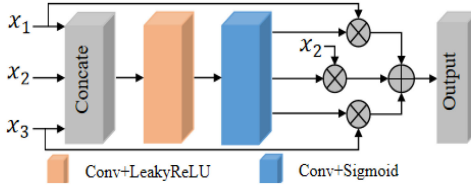


Fig. 3. Multiscale fusion module. Conv denotes convolutional operation.

branch are  $x_1$ ,  $x_2$  and  $x_3$ . In addition, we design a multiscale fusion module (MFM) for merging the global features between branches with different convolution kernel sizes. The multiscale fusion module is shown in Fig. 3.  $x_1$ ,  $x_2$  and  $x_3$  denote the outputs of the three branches and are used as the inputs of the multiscale fusion module.

### B. Parallel Attention Module

1) *Pixel Attention (PA)*: We design PA that primarily pays attention to pixel values. The higher the pixel values within a specific range indicate that images have high illumination and bright color. Firstly, we use two convolution layers to focus on pixel features, the shape of the features change from  $C \times H \times W$  to  $1 \times H \times W$ . Finally, The pixel features are multiplied by the input features of PA. PA can be seen in Fig. 4. The number of channels for PA output feature is  $C$ , defined as 64.

2) *Channel Attention (CA)*: CA mainly assigns a weight value to each channel, balancing color information and adjusting uneven illumination features between different channels. CA can be seen in Fig. 4. Firstly, we take features into a channel conversion using global average pooling (GAP), the shape of features change from  $C \times H \times W$  to  $C \times 1 \times 1$ . Then, we employ two convolution layers to learn weight values. Finally, the weight values are multiplied by the input features of CA.

Many underwater enhancement methods [15], [18] neglect the relevance of illumination features and color information. Thus, we design a PAM including CA and PA in parallel to focus on

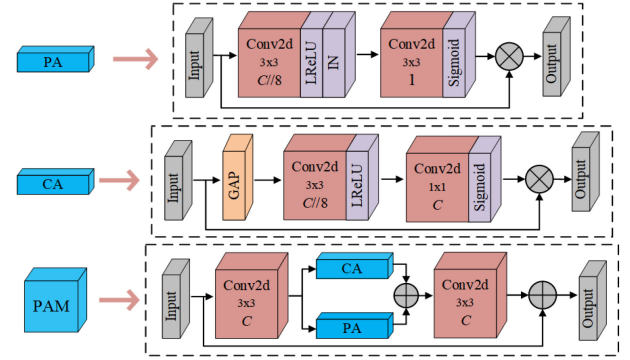


Fig. 4. Illustration of the parallel attention module. PA is pixel attention, CA denotes channel attention, PAM indicates parallel attention module.

illumination features and color information. This can be seen in Fig. 4. Firstly, we use a convolution layer to extract features. Then, parallel PA and CA are utilized to learn corresponding features. The learned features are added in pixels. Furthermore, a convolution layer is added to balance illumination features and color information. Finally, we introduce local residual connections that focus on practical information. Ablation study (Sec. IV-D) show that the parallel connection of PA and CA is better than the serial connection in the PAM. This is shown in Fig. 6.

### C. Group Structure

The group structure is composed of 20 parallel attention modules (PAMs). And the continuous PAM increases the depth and performance of the LANet. At the end of LANet, using a convolution layer reduces the dimension to obtain underwater image with bright color and high illumination.

### D. Adaptive Learning Module

Excessively increasing the depth of the network will lead to the weakening of the expression ability. We design an ALM,



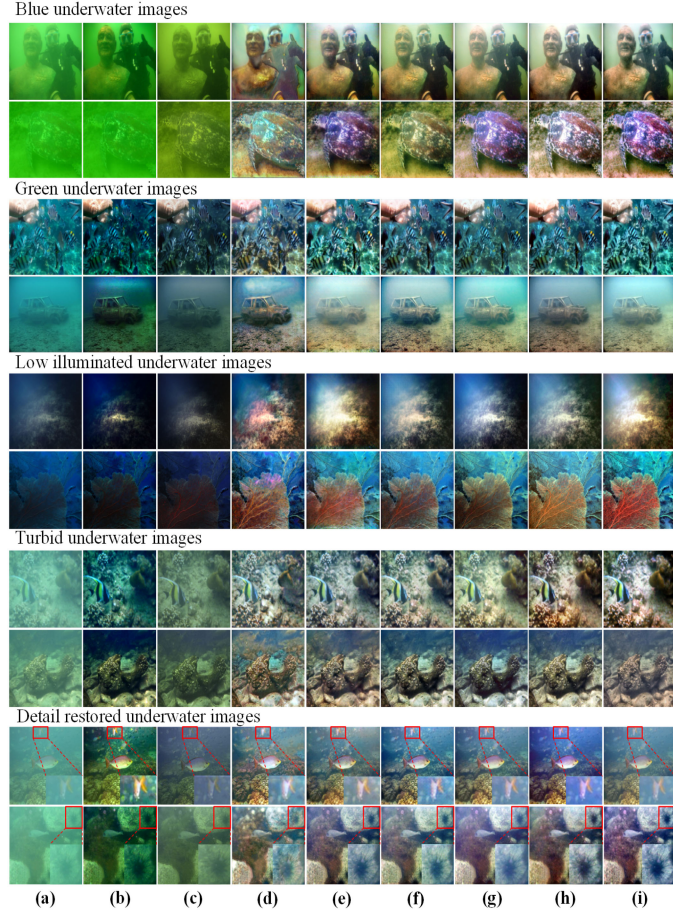


Fig. 5. Experimental comparison on underwater images. a) raw underwater image; b) UDCP; c) Fusion; d) UResnet; e) FUNIE-GAN; f) UGAN; g) WaterNet; h) LANet; i) Ground Truth.

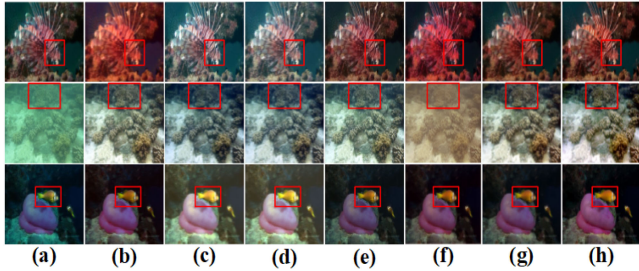


Fig. 6. Ablation study. a) raw underwater image; b) N2group; c) Ngroup; d) Npooling; e) Sgroup; f) Nasn; g) LANet; h) Ground Truth.

which can adaptive learn the most representative features and transfer shallow information into deep layers. First, we utilize the global average pooling(GAP) to adaptively obtain  $1 \times C$  feature vector, which demonstrate the importance of different channel characteristics. Then, feature vector is multiplied by the output feature of group structure(GS) to get representative features. In addition, we add shallow information and representative features for residual learning. The ALM is shown in Fig. 2. The ALM can be described by:

$$F_{ALM} = U_G(I_c) \otimes U_{GS}(I_c) \oplus I_s \quad (1)$$

where  $F_{ALM}$  is the output of ALM.  $U_G$  and  $U_{GS}$  perform GAP and GS operations, respectively.  $I_c$  is the input characteristics of the current GS.  $\oplus$  denotes addition.  $\otimes$  is multiplication.  $I_s$  indicates shallow information.

### E. Loss Function

1) *Laplace Loss*: To reduce edge information loss, we employ the Laplace operator as the edge detection operator so that the details of the generated underwater images are improved. Laplace template can be described as:

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

The generated images from LANet are denoted as  $I_g$ , and the corresponding ground truth images are named  $I_l$ . Simultaneously, the convolution operation is performed on the generated image  $I_g$  and ground truth images  $I_l$  using the Laplacian template. Laplace loss  $L_{lap}$  is defined as:

$$L_{lap} = \frac{1}{N} \sum_{i=1}^N (lap(I_g^i) - lap(I_l^i))^2 \quad (2)$$

where  $N$  denotes total pixels in the image.  $i$  indicates the position of the pixel.

2) *Feature Reconstruction Loss*:  $L_1$  loss as the content loss is adopted to maintain consistency between pixels [29]. The LANet generated images and corresponding ground truth images are calculated using the  $L_1$  loss. It is presented by:

$$L_1 = \frac{1}{N} \sum_{i=1}^N |I_g^i - I_l^i| \quad (3)$$

The feature reconstruction of underwater images is carried out through the perceptual loss based on the VGG network [5], which is pretrained on the ImageNet dataset [30]. The perceptual loss function is defined as follows:

$$L_{VGG} = \frac{1}{CWH} \sum_{c=1}^C \sum_{w=1}^W \sum_{h=1}^H (V(I_{c,w,h}^c) - V(I_{c,w,h}^g))^2 \quad (4)$$

$C$ ,  $W$ , and  $H$  represent the image's channel, width, and height.  $V$  denotes the nonlinear conversion through the VGG-16 network. The  $L_1$  loss and the perceptual loss function are linearly superimposed to obtain feature reconstruction loss  $L$ . It is described as:

$$L = L_1 + \lambda L_{VGG} \quad (5)$$

where  $\lambda$  are constant, set to 0.5.

3) *Asynchronous Training Mode*: We adopt asynchronous training mode [20] to make the network more robust and generate better underwater images. In ablation study(Sec.IV-D), we verify the advantages of asynchronous training mode. The specific steps are as follows. First,  $L_{lap}$  is utilized to calculate the gradient, and backpropagation is performed to update the network weight. Second, the  $L$  loss is leveraged to calculate the gradient and backpropagation to update the network weight.

TABLE I  
THE NUMBER OF TESING AND TRAINING IMAGES

	EUVP	Underwater ImageNet	UIEBD
Training images	2000	2000	790
Testing images	300	400	100

TABLE II  
EVALUATION VALUE OF UCIQE AND UIQM

model	UCIQE $\uparrow$	UIQM $\uparrow$
Raw	0.4143	2.7229
UDCP	0.5245	3.8734
Fusion	0.4707	4.2559
UResnet	0.5778	4.7818
FUnIE-GAN	0.5877	4.8528
UGAN	0.5941	4.8848
Water-Net	0.5899	4.6740
LANet	<b>0.6092</b>	<b>4.9748</b>

Therefore, each batch is trained twice, and the network weights are updated twice.

#### IV. EXPERIMENTS

To verify the effectiveness of LANet, we first introduce the implementation details of the experiment. Then, we provide qualitative and quantitative comparisons with representative methods on different datasets. These methods include UDCP [31], fusion [32], UResnet [20], FUnIE-GAN [8], UGAN [22] and Water-Net [26]. Finally, we perform a series of ablation studies to verify each component of LANet. In addition, we present the enhanced images by our method in the practical underwater application and discuss the speed of our methods.

##### A. Implementation Details

We independently conduct experiments on these datasets. Each dataset is divided into training dataset and testing dataset, and all methods are compared under the same experiment settings. These datasets include EUVP [8], underwater ImageNet [22] and UIEBD [26]. The details are as follows: (1) In the EUVP dataset, we randomly select 2,000 pairs of images for training and 300 pairs of images for testing from underwater scenes. (2) The underwater ImageNet dataset generates 6128 pairs of underwater images by using CycleGAN. For a random set of 2,000 and 400 pairs of images for training and testing, respectively. (3) The UIEBD dataset includes 890 pairs of underwater images. We randomly set 790 pairs of images for training and use the rest of the image for testing. The number of tesing and training images can be seen in Table I. We train our proposed LANet using Adam and set the learning rate to 0.0001. The batch size set 4, epoch is 40. We use PyTorch as the deep learning framework on an Inter(R) W-2223 CPU, 32 GB RAM, and GeForce RTX 3090 GPU.

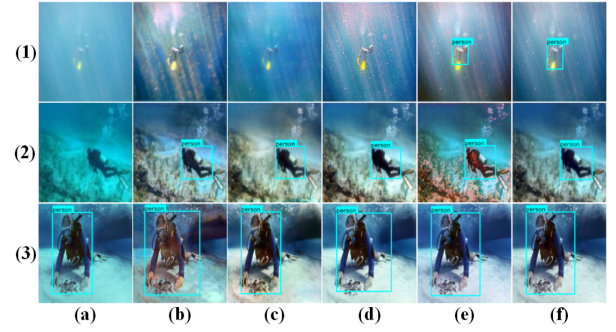


Fig. 7. Object detection on underwater images. a) raw underwater image; b) UResnet; c) FUnIE-GAN; d) UGAN; e) Water-Net; f) LANet.

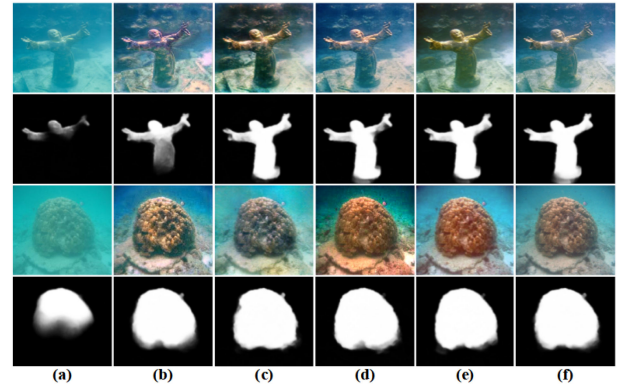


Fig. 8. Salient object detection on underwater images. a) raw underwater image; b) UResnet; c) FUnIE-GAN; d) UGAN; e) Water-Net; f) LANet.

##### B. Qualitative Analysis

To verify the effectiveness of LANet in removing color deviation and promoting low illumination with different underwater environments. We choose five categories of experimental comparison, including green underwater images, blue underwater images, low illuminated underwater images, turbid underwater images, and detail restored underwater images, each category presented two images in Fig. 5. UDCP method has a poor effect in reducing color casts due to the lack of color correction processing. Fusion method solves the problem of blue underwater images. However, it can not improve low illumination effectively. UResnet and UGAN methods can deal with a slight color casts but can not remove the severe color casts. FUnIE-GAN, Water-Net, and LANet methods remove color casts well. However, Water-Net method retains a layer of fog on the surface of underwater images.

Due to the complexity of the underwater environment, underwater images exhibit low illumination. UDCP and Fusion methods processing effect is poor at low illuminated regions. UResnet may produce additional colors. FUnIE-GAN method generates illumination is too bright, resulting in unclear details. UGAN and Water-Net methods incompletely improve low illumination.

From Fig. 5, the UDCP and Fusion have poor performance in removing turbidity and color restoration. The method of UResnet can get better recovery, but it introduces additional

TABLE III  
IMAGE QUALITY ASSESSMENT IN TERMS OF PSNR,SSIM AND MSE

Method	EUVP			Underwater ImageNet			UIEBD		
	PSNR↑	SSIM↑	MSE↓	PSNR↑	SSIM↑	MSE↓	PSNR↑	SSIM↑	MSE↓
Raw	20.88	0.6762	625	17.78	0.4598	1196	18.18	0.6489	1531
UDCP	17.09	0.6464	1,608	16.41	0.5166	1783	13.97	0.6154	3157
Fusion	13.81	0.3707	2,851	12.96	0.4692	3498	13.04	0.3871	3427
UResnet	25.81	<b>0.8780</b>	243	24.47	0.8389	266	19.32	0.7891	857
FUnIE-GAN	24.60	0.8557	305	23.30	0.8097	334	23.56	0.8689	345
UGAN	24.73	0.8562	290	24.49	0.8356	259	23.03	0.8591	402
Water-Net	23.78	0.8000	328	23.32	0.8064	340	<b>24.08</b>	0.8882	397
LANet	<b>25.82</b>	0.8668	<b>238</b>	<b>24.69</b>	<b>0.8509</b>	<b>252</b>	24.06	<b>0.9085</b>	<b>319</b>

TABLE IV  
ABLATION STUDY

Method	EUVP			Underwater ImageNet			UIEBD		
	PSNR↑	SSIM↑	MSE↓	PSNR↑	SSIM↑	MSE↓	PSNR↑	SSIM↑	MSE↓
Raw	20.88	0.6762	625	17.78	0.4598	1196	18.18	0.6489	1531
N2group	19.41	0.6214	841	24.34	0.8404	272	23.09	0.8808	397
Ngroup	18.50	0.5992	1143	19.23	0.6785	1092	20.63	0.7900	684
Npooling	18.70	0.6140	1075	19.21	0.6769	1042	21.48	0.8122	533
Sgroup	25.21	0.8519	248	24.24	0.8381	277	22.59	0.8578	466
Natm	25.26	0.8011	266	24.14	0.7993	284	23.60	0.8910	357
LANet	<b>25.82</b>	<b>0.8668</b>	<b>238</b>	<b>24.69</b>	<b>0.8509</b>	<b>252</b>	<b>24.06</b>	<b>0.9085</b>	<b>319</b>

TABLE V  
CONFIDENCE VALUE OF OBJECT DETECTION

images	raw	UResnet	FUnIE-GAN	UGAN	Water-Net	LANet
1	No	No	No	No	0.68	<b>0.74</b>
2	No	0.71	<b>0.80</b>	0.72	0.65	0.76
3	0.79	0.81	0.64	80	0.83	<b>0.87</b>

background noise. The remaining methods can achieve satisfactory restoration results, while our proposed LANet is more visually significant.

We perform a set of detailed recovery experiments. The lower right corner is an enlarged view of the red box in Fig. 5. It can be seen that the traditional methods have poor recovery ability regardless of long-distance or short-distance detail restoration. In complex scenes, UResnet and FUnIE-GAN methods can produce color casts questions, resulting in poor detail recovery. UGAN, Water-Net, and LANet methods have sound effects in detail recovery. From the comparison results in Fig. 5, our proposed LANet effectively removes the color casts and turbidity, and improves the brightness and saturation in underwater images.

### C. Quantitative Evaluation

First, we choose underwater color image quality evaluation (UCIQE) [33] and underwater image quality evaluation metric (UIQM) [34] as no-reference image quality evaluation. UIQM measures underwater image metrics by colorfulness (UICM), sharpness (UISM) and contrast (UIConM). UCIQE evaluates underwater image quality by color density, saturation, and contrast.

UCIQE metric is defined as follows:

$$V_{UCIQE} = c_1 \times \sigma_c + c_2 \times \sigma_{conl} + c_3 \times \sigma_s \quad (6)$$

where  $\sigma_c$  is the standard deviation of chromaticity;  $\sigma_{conl}$  represents the contrast of brightness; and  $\sigma_s$  is the average value of saturation. We set  $c_1 = 0.4680$ ,  $c_2 = 0.2745$  and  $c_3 = 0.2576$  according to [33].

We employ UCIQE and UIQM to evaluate the recovered underwater images in Fig. 5. The evaluation metrics are shown in Table II. Bold indicates the optimal value. Our proposed LANet achieves better performance on both UCIQE and UIQM metrics. This indicates that the overall effect of LANet is more natural, the illumination is higher, and the color recovery is better than other methods.

Then, we verify the effectiveness of the LANet on different datasets. We train the model on EUVP, Underwater ImageNet and UIEBD datasets, respectively. We use the peak signal-to-noise ratio (PSNR) [35], structural similarity (SSIM) [36], and mean square error (MSE) to quantitatively analyze the testing results on the EUVP, underwater ImageNet, and UIEBD datasets. The quantitative results of distinct methods with different datasets are shown in Table III. Our proposed LANet achieves the better evaluation metrics than other methods in the different datasets. The results demonstrate that LANet has stronger adaptability, produces less noise and is visually closer to the underwater ground truth images.

### D. Ablation Study

To verify the effectiveness of the proposed modules in the LANet, we do several ablation studies, and the results are shown



TABLE VI  
EXPERIMENTAL RESULTS IN TERMS OF PARAMETERS, FLOPS, AND FPS

	UDCP	Fusion	UResnet	FUnIE-GAN	UGAN	Water-Net	LANet
Parameters(M)	-	-	0.60	17	57.20	1.09	5.15
Flops(G)	-	-	77.81	10.69	19.81	450.32	355.37
Fps	1.35	1.09	89.21	246.72	154.91	16.39	18.42

in Table IV. the details are as follows: (1) N2group denotes only one group structure and one ALM, (2) Ngroup denotes without group structure and ALM, (3) Npooling denotes only no ALM, and (4) Sgroup denotes that PA and CA use a serial connection in the PAM. (5) Natm is without asynchronous training mode. We compared the proposed LANet with N2group, Ngroup, Npooling, Sgroup and Natm as an ablation experiment.

The ablation experiment is presented in Fig. 6. First and third images show that Ngroup and Npooling are poor in color recovery and overexposure. Second and third images demonstrate that adding one group structure and one LAC can effectively avoid the overexposure problem and highlight image details. However, color recovery supersaturation due to shallow networks cannot learn image features well. Therefore, we chose three group structures and the corresponding LAC, giving enhanced underwater images vivid color and significant visual effects. Finally, to prove the advantages of parallel connections in PAM, we add a set of serial connection experiments. First and third images indicate that serial connections are worse than parallel connections in illumination and color recovery. In addition, Nasn enhanced underwater image is red as a whole in first image, and produce a layer fog in second image. It indicates that asynchronous training mode makes the LANet more robust and generates better underwater images.

In Table IV, there are main reasons show that N2group, Ngroup, and Npooling have lower than Raw for EUVP. First, Euvp dataset has large internal differences, and the network may incomplete learn its characteristics. Second, N2group and Ngroup have fewer parameters can't better learn the mapping relationship. Npooling employs more convolution layers, resulting in deep network depth and unable to effectively learn image features. The results demonstrate that each module plays an essential role in our network.

### E. Application

The improvement of underwater image quality can promote the application of underwater robots in high-level visual perception. We verify the proposed method that improves the accuracy of object detection and enhances detail recovery of salient object detection. We leverage YOLOv4 [37] to detect the table of people in enhanced underwater image, as shown in Fig. 7. We utilize the confidence value, which indicates the probability that the detection box overlaps with the original label box, as the quantitative metric. The greater the confidence value, the more accurate the detected results are. In Fig. 7, there are no boxes in the images, indicating that the confidence value is less than 0.4. Confidence value of object detection can be seen as Table V, "No" means confidence value is less than 0.4. LANet

can detect all boxes, and have higher confidence values than other methods. It demonstrates LANet can improve the accuracy of object detection.

The purpose of salient object detection is to highlight the salient object area in image. We utilize the attentive saliency network (ASNet) [38] to identify and segment the salient object in the scene, which proves that enhanced methods can recover the details of the underwater image well. The salient object detection is presented in Fig. 8. In contrast, the enhanced images can clearly show the complete outline, highlighting that LANet has a clear detail recovery function. Experimental results demonstrate that LANet has a dominant visual effect and practicability.

### F. Discussion of Speed Aspect

We calculate parameters, flops and fps of different methods. Table VI shows the experimental results on parameters, flops, and fps. Our proposed LANet utilizes more multiplication of PA and CA, resulting in higher flops. However, it is faster than the Water-Net, even ten times faster than traditional methods (UDCP and Fusion). Then, the fps of LANet is 18.42, and it basically meets the real-time requirements of underwater robots. From Table III and Table VI, LANet has certain competitiveness in balancing between speed and quantitative metrics compared with other methods.

## V. CONCLUSION

We proposed a novel adaptive learning attention network underwater image enhancement method. This method incorporates the color information and illumination features by parallel channel attention and pixel attention. Besides, we employ adaptive learning module by global average pooling and residual learning to highlight the feature performance ability of the network. Our method performs very prominently in removing color casts and low illumination, and also immensely restores the detailed information of the underwater image. Experiments demonstrated that our proposed LANet has significant advantages on different underwater datasets, including qualitative analysis and quantitative evaluations. Furthermore, our method also achieves good results in object detection and salient object detection of underwater images.

## REFERENCES

- [1] F. Shkurti *et al.*, "Multi-domain monitoring of marine environments using a heterogeneous robot team," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2012, pp. 1747–1753.
- [2] J. Li, K. A. Skinner, R. M. Eustice, and M. Johnson-Roberson, "WaterGAN: Unsupervised generative network to enable real-time color correction of monocular underwater images," *IEEE Robot. Automat. Lett.*, vol. 3, no. 1, pp. 387–394, Jan. 2018.

- [3] M. J. Islam, M. Ho, and J. Sattar, "Understanding human motion and gestures for underwater human-robot collaboration," *J. Field Robot.*, vol. 36, no. 5, pp. 851–873, 2019.
- [4] R. Schettini and S. Corchs, "Underwater image processing: State of the art of restoration and image enhancement methods," *EURASIP J. Adv. Signal Process.*, vol. 2010, pp. 1–14, 2010.
- [5] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Int. Conf. Learn. Representations*, May 2015, *arXiv:1409.1556*.
- [6] Y. Wang, J. Zhang, Y. Cao, and Z. Wang, "A deep CNN method for underwater image enhancement," in *Proc. IEEE Int. Conf. Image Process.*, 2017, pp. 1382–1386.
- [7] K. Cao, Y.-T. Peng, and P. C. Cosman, "Underwater image restoration using deep networks to estimate background light and scene depth," in *Proc. IEEE Southwest Symp. Image Anal. Interpretation*, 2018, pp. 1–4.
- [8] M. J. Islam, Y. Xia, and J. Sattar, "Fast underwater image enhancement for improved visual perception," *IEEE Robot. Automat. Lett.*, vol. 5, no. 2, pp. 3227–3234, Apr. 2020.
- [9] X. Ye, H. Xu, X. Ji, and R. Xu, "Underwater image enhancement using stacked generative adversarial networks," in *Pacific Rim Conf. Multimedia*, 2018, pp. 514–524.
- [10] Y. Guo, H. Li, and P. Zhuang, "Underwater image enhancement using a multiscale dense generative adversarial network," *IEEE J. Ocean. Eng.*, vol. 45, no. 3, pp. 862–870, Jul. 2020.
- [11] X. Chen, J. Yu, S. Kong, Z. Wu, X. Fang, and L. Wen, "Towards real-time advancement of underwater visual quality with GAN," *IEEE Trans. Ind. Electron.*, vol. 66, no. 12, pp. 9350–9359, Dec. 2019.
- [12] C. Li, J. Guo, and C. Guo, "Emerging from water: Underwater image color correction based on weakly supervised color transfer," *IEEE Signal Process. Lett.*, vol. 25, no. 3, pp. 323–327, Mar. 2018.
- [13] P. M. Uplavikar, Z. Wu, and Z. Wang, "All-in-one underwater image enhancement using domain-adversarial learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2019, pp. 1–8.
- [14] X. Liu, Z. Gao, and B. M. Chen, "IPMGAN: Integrating physical model and generative adversarial network for underwater image enhancement," *Neurocomputing*, vol. 453, pp. 538–551, 2021.
- [15] L. Hong, X. Wang, Z. Xiao, G. Zhang, and J. Liu, "WSUIE: Weakly supervised underwater image enhancement for improved visual perception," *IEEE Robot. Automat. Lett.*, vol. 6, no. 4, pp. 8237–8244, Oct. 2021.
- [16] M. Hou, R. Liu, X. Fan, and Z. Luo, "Joint residual learning for underwater image enhancement," in *Proc. 25th IEEE Int. Conf. Image Process.*, 2018, pp. 4043–4047.
- [17] R. Liu, M. Hou, J. Liu, X. Fan, and Z. Luo, "Compounded layer-prior unrolling: A unified transmission-based image enhancement framework," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2019, pp. 538–543.
- [18] 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 Process.*, vol. 13, no. 3, pp. 469–474, 2019.
- [19] A. Jamadandi and U. Mudénagudi, "Exemplar-based underwater image enhancement augmented by wavelet corrected transforms," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2019, pp. 11–17.
- [20] P. Liu, G. Wang, H. Qi, C. Zhang, H. Zheng, and Z. Yu, "Underwater image enhancement with a deep residual framework," *IEEE Access*, vol. 7, pp. 94614–94629, 2019.
- [21] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1646–1654.
- [22] C. Fabbri, M. J. Islam, and J. Sattar, "Enhancing underwater imagery using generative adversarial networks," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2018, pp. 7159–7165.
- [23] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2223–2232.
- [24] C. Li, S. Anwar, and F. Porikli, "Underwater scene prior inspired deep underwater image and video enhancement," *Pattern Recognit.*, vol. 98, 2020, Art. no. 107038.
- [25] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from RGBD images," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 746–760.
- [26] C. Li *et al.*, "An underwater image enhancement benchmark dataset and beyond," *IEEE Trans. Image Process.*, vol. 29, pp. 4376–4389, Nov. 2019.
- [27] Q. Qi *et al.*, "Underwater image co-enhancement with correlation feature matching and joint learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 3, pp. 1133–1147, Mar. 2022, doi: [10.1109/TCSVT.2021.3074197](https://doi.org/10.1109/TCSVT.2021.3074197).
- [28] D. Berman, D. Levy, S. Avidan, and T. Treibitz, "Underwater single image color restoration using haze-lines and a new quantitative dataset," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 8, pp. 2822–2837, Aug. 2020.
- [29] K. Wang, L. Shen, Y. Lin, M. Li, and Q. Zhao, "Joint iterative color correction and Dehazing for underwater image enhancement," *IEEE Robot. Automat. Lett.*, vol. 6, no. 3, pp. 5121–5128, Jul. 2021.
- [30] 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.*, 2009, pp. 248–255.
- [31] P. L. Drews, E. R. Nascimento, S. S. Botelho, and M. F. M. Campos, "Underwater depth estimation and image restoration based on single images," *IEEE Comput. Graph. Appl.*, vol. 36, no. 2, pp. 24–35, Mar./Apr. 2016.
- [32] 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, pp. 379–393, Jan. 2018.
- [33] M. Yang and A. Sowmya, "An underwater color image quality evaluation metric," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 6062–6071, Dec. 2015.
- [34] Y.-T. Peng and P. C. Cosman, "Underwater image restoration based on image blurriness and light absorption," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1579–1594, Apr. 2017.
- [35] I. Avcibas, B. Sankur, and K. Sayood, "Statistical evaluation of image quality measures," *J. Electron. Imag.*, vol. 11, no. 2, pp. 206–223, 2002.
- [36] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [37] A. Bochkovskiy, C.-Y. Wang, and H.-Y. Mark Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
- [38] W. Wang, J. Shen, X. Dong, and A. Borji, "Salient object detection driven by fixation prediction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 1711–1720.