Article

NUAM-Net: A Novel Underwater Image Enhancement Attention Mechanism Network

Zhang Wen [ID], Yikang Zhao, Feng Gao [ID], Hao Su, Yuan Rao *[ID] and Junyu Dong *

Faculty of Information Science and Engineering, Ocean University of China, Qingdao 266005, China;
wenzhang@stu.ouc.edu.cn (Z.W.); zhaoyikang2022@163.com (Y.Z.); gaofeng@ouc.edu.cn (F.G.);
suhao@ouc.edu.cn (H.S.)
* Correspondence: raoyuan@stu.ouc.edu.cn (Y.R.); dongjunyu@ouc.edu.cn (J.D.)

Abstract: Vision-based underwater exploration is crucial for marine research. However, the degradation of underwater images due to light attenuation and scattering poses a significant challenge. This results in the poor visual quality of underwater images and impedes the development of vision-based underwater exploration systems. Recent popular learning-based Underwater Image Enhancement (UIE) methods address this challenge by training enhancement networks with annotated image pairs, where the label image is manually selected from the reference images of existing UIE methods since the groundtruth of underwater images do not exist. Nevertheless, these methods encounter uncertainty issues stemming from ambiguous multiple-candidate references. Moreover, they often suffer from local perception and color perception limitations, which hinder the effective mitigation of wide-range underwater degradation. This paper proposes a novel NUAM-Net (Novel Underwater Image Enhancement Attention Mechanism Network) that addresses these limitations. NUAM-Net leverages a probabilistic training framework, measuring enhancement uncertainty to learn the UIE mapping from a set of ambiguous reference images. By extracting features from both the RGB and LAB color spaces, our method fully exploits the fine-grained color degradation clues of underwater images. Additionally, we enhance underwater feature extraction by incorporating a novel Adaptive Underwater Image Enhancement Module (AUEM) that incorporates both local and long-range receptive fields. Experimental results on the well-known UIEBD benchmark demonstrate that our method significantly outperforms popular UIE methods in terms of PSNR while maintaining a favorable Mean Opinion Score. The ablation study also validates the effectiveness of our proposed method.

Keywords: underwater image enhancement; deep learning; probabilistic network; LAB color space; conditional variational autoencoder; interactive channel crossing attention

1. Introduction

Underwater visual quality degrades due to wavelength-dependent light scattering and absorption under the water, resulting in low-visibility, low-contrast, and color-cast issues in underwater images [1]. This limits the accuracy of vision-based underwater systems and tasks, e.g., underwater tracking [2,3], robot navigation [4,5], and ecological monitoring [6,7]. Researching advanced Underwater Image Enhancement (UIE) techniques [8–11], which improve the visual quality of degraded underwater images and benefit vision-based underwater systems, is of great significance for the development of marine engineering.

Recently, deep learning-based image enhancement methods have made significant advancements by training models with well-collected image pairs to learn the mapping from the low quality images to the reference images. However, it is impractical for underwater image enhancement tasks to obtain groundtruth clear images since the irreversibility of underwater imaging progresses with highly complex degradation. To address this challenge, popular methods [10,12–14] propose generating reference images to approximate groundtruth images to train UIE models. For instance, Ref. [10] utilizes 12 state-of-the-art

UIE algorithms to generate a set of enhanced images, manually selecting the best image as the reference image. Leveraging these pairs of underwater images and their well-enhanced reference images, deep learning-based UIE approaches have achieved impressive performance in improving the visual quality of underwater images [11]. Nevertheless, the reference image can not perfectly approximate the groundtruth and is susceptible to various influences, including subjective human preferences during the selection process and variations in algorithm parameters. These lead to insufficient UIE learning for the uncertainty issue of the ambiguous label, i.e., multiple potential solutions exist for the same degraded underwater image. As shown in Figure 1, using a single reference image as the label to train a UIE model is sometimes insufficient since lacking the true clear image and multiple candidate references can lead to ambiguity in selecting the best one.
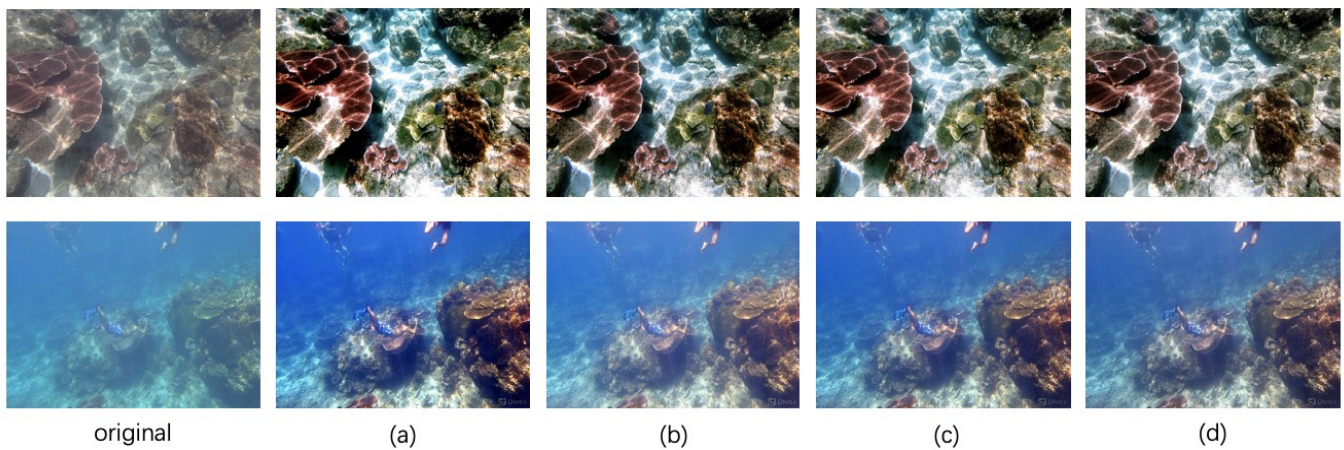

Figure 1. Illustration of uncertainty issue in UIE learning. We show examples of UIEBD datasets, i.e., the original image, (**a**) selected reference, (**b**) contrast adjustment result, (**c**) saturation adjustment result, and (**d**) gamma correction result. Multiple potential solutions can be ambiguous in reference selection since different people might choose different labels as the reference.

To address the uncertainty issue, we follow PUIE-Net [8] to tackle the uncertainty problem as the probabilistic sampling approximation problem. Let $\mathbf{x}$ and $\mathbf{y}$ denote the degraded underwater image observation and the clear enhanced image, respectively. Considering that $\mathbf{z}$ represents the uncertainty arising from different people choosing reference images generated by different algorithms as training labels for deep learning networks, UIE aims to model the clear image distribution from $\mathbf{x}$ with uncertain reference $\mathbf{z}$, i.e., $p(\mathbf{y} \mid \mathbf{z}, \mathbf{x})$. For a given $\mathbf{x}$, we can assume that $\mathbf{z}$ follows a distribution $p(\mathbf{z} \mid \mathbf{x})$ because the uncertainty of $\mathbf{z}$ is generated by the process of $\mathbf{x}$. Once the sampling size S is large enough, $\mathbf{z}$ approximately follows a normal distribution and the UIE model can be approximated as [8]

$$p(\mathbf{y} \mid \mathbf{x}) \approx \frac{1}{S} \sum_{s=1}^{S} p\left(\mathbf{y} \mid \mathbf{z}^{(s)}, \mathbf{x}\right), \mathbf{z}^{(s)} \sim p(\mathbf{z} \mid \mathbf{x}) \tag{1}$$

Motivated by this theoretical method, which is different from most existing methods, PUIE-Net proposed a probabilistic training framework that randomly samples one of the multiple candidate references instead of the selected "best" reference for training the UIE model, avoiding the uncertainty issue.

Although PUIE-Net achieves encouraging enhancement results, it does not perform well in challenging scenarios. We argue that there are two perception limitations to PUIE-Net: **(1) local perception limitation**—PUIE-Net adopts the U-shaped SE-ResNet50 architecture with a limited local receptive field as the feature extractor, which makes it hard to model the long-range dependencies as well as global perception for dealing with large-scale underwater degradation; **(2) color perception limitation**—most existing methods, as well

as PUIE-Net, exclusively extract features from the RGB color space, which is not always enough to capture fine-grained underwater color degradation clues.

In this paper, we propose a novel NUAM-Net to address these limitations. Firstly, we proposed a novel Adaptive Underwater Image Enhancement Module (AUEM) that leverages three parallel mechanisms—Large-Kernel Attention (LKA), Simple Gate (SG), and Channel Attention (CA)—to model the long-range spatial and channel interaction with both local and long-range receptive fields to avoid the local perception limitation. Secondly, we enrich the color perception by extracting features from both RGB space and a wider and more accurate color-represented LAB space, to highlight fine-grained underwater color degradation clues and address the color perception limitation. Built on the probabilistic training framework, our NUAM-Net achieves significant PSNR improvements on the popular UIEBD benchmark compared to state-of-the-art UIE methods.

In conclusion, our contributions are summarised as follows:

— Based on a probabilistic training framework, we propose a novel NUAM-Net that extracts features from both RGB and LAB color spaces and that models long-range spatial-channel interaction with both local and long-range receptive fields, avoiding the uncertainty issue in UIE learning as well as the local and color perception limitations introduced by PUIE-Net;
— We conduct comprehensive experiments on the well-known UIEBD benchmark, and the highly competitive PSNR and SSIM results against state-of-the-art UIE methods demonstrate the effectiveness of our method. The ablation study also illustrates the gains of the proposed components.

2. Related Work

In this section, we briefly introduce the previous works regarding model-free UIE methods, prior-based UIE methods, learning-based methods, and the attention mechanism.

2.1. Model-Free UIE Methods

Model-free techniques typically refine underwater images by directly adjusting pixel luminance without relying on specific physical models, such as using Contrast-Limited Adaptive Histogram Equalization (CLAHE) [15], White Balancing (WB) [16], and Retinex [17]. Ref. [18] introduced a fusion-based Underwater Image Enhancement (UIE) method where the inputs and weights are determined solely from the degraded images. Ref. [19] improved upon this method by incorporating white balancing techniques and an innovative multiscale fusion strategy to achieve better enhancement results. Fu and colleagues [20] proposed a Retinex-based UIE method designed for enhancing individual underwater images. Gao and associates [21] developed an underwater image enhancement technique inspired by the functionality of fish retinas, aiming to address issues such as color bias, unevenness, and content blur in images. While these model-free techniques are efficient and straightforward to implement, their disregard for the complex mechanisms of underwater imaging can sometimes lead to unstable outcomes and fail to achieve the desired image enhancement effects.

2.2. Prior-Based UIE Methods

Prior-based methods focus on estimating the parameters of underwater imaging models through prior hypotheses, and then use these physical models to enhance the quality of underwater images. Chiang and colleagues [22] proposed a method that utilizes dehazing technology to enhance underwater images. Galdran et al. [23] adapted the Dark Channel Prior (DCP) [24] by using information from the red channel to infer the depth map of underwater images. Li et al. [25] introduced a dehazing method tailored to the characteristics of underwater environments and proposed a contrast enhancement technique based on the principles of the minimum information loss and the prior knowledge of a histogram distribution. Berman et al. [26] considered the spectral profiles of different water types and additionally estimated two global parameters: the attenuation ratios

between the blue–red and blue–green channels. Akkaynak et al. [27] developed the Sea-thru method, which is based on an improved physical imaging model and uses RGBD images as input to estimate scattering from the darkest pixel and its known depth map, and then estimates the attenuation coefficient of varying illumination across the scene. While these methods are effective in specific contexts, they may not be sufficiently robust in handling more complex scenarios due to the challenge faced by parameterized physical models in perfectly capturing the complexity and diversity of underwater environments.

2.3. Deep Learning Methods

Learning-based UIE methods stand out from model-free and prior-based approaches by leveraging the powerful feature extraction capabilities of deep neural networks and non-linear mapping functions, driven by data, to enhance underwater images. Li et al. [28] were the first to unsupervisedly use generative adversarial networks to create synthetic underwater images, which were then employed to train an enhancement network. Li et al. [29] proposed a method that requires only weak supervision, reducing the need for paired data. Guo et al. [30] employed a multi-scale dense generative adversarial network for underwater image enhancement. Li et al. [31] developed a lightweight UIE model that incorporates underwater scene priors. Li et al. [10] curated a comprehensive real-world UIE dataset, UIEB, with reference images manually selected from several existing UIE methods, and proposed a gated fusion network for image enhancement based on this dataset. Jamadandi et al. [32] suggested enhancing underwater images using networks combined with wavelet transform corrections. Addressing the diverse degradation characteristics of underwater images, Uplavikar et al. [33] trained a deep neural network to extract domain-invariant features from given images, with the domain defined by the Jerlov water type. Li et al. [11] introduced Ucolor, a UIE network based on medium transmission-guided multi-color space embedding. Kar et al. [34] proposed a zero-shot restoration method for underwater and dehazed images, leveraging theoretically derived degradation properties. However, many current learning-based methods [35–42] rely on end-to-end training with annotated image pairs, leading to uncertainties due to the ambiguity of multiple potential reference images. To address this issue, PUIE-Net [8] approached the uncertainty problem as a probabilistic sampling approximation and introduced a probabilistic training framework for UIE. Our research builds upon this probabilistic training framework and introduces NUAM-Net, which models local and long-range dependencies with enhanced color perception to capture detailed underwater color degradation cues, overcoming the limitations of local perception and color understanding in PUIE-Net.

2.4. Attention Mechanism

In deep learning, attention mechanisms have become a key technique [43,44] and are acclaimed for their ability to enhance a model's focus on key elements within input data [45–47]. Channel Attention (CA) meticulously examines the dynamics of cross-channel feature activation [48], highlighting the relational importance of different features, while spatial attention evaluates the significance of the layout of information space [49], optimizing the model's perceptual field. Large-Kernel Attention (LKA) [50], by combining Depthwise Convolution, Depthwise Dilation Convolution, and Pointwise Convolution, effectively captures long-distance relationships within features, improving adaptability. In this paper, we propose a novel adaptive underwater enhancement module that takes advantage of the local and long-range receptive fields of CA and LKA to model the long-range spatial and channel interaction; it also leverages an extra Simple Gate (SG) to fully explore the complementary information between the CA and LKA. This module shows significant gains in our ablation study.

3. Method

In this section, we elaborate our method.

3.1. Probabilistic Training Framework and Multi-Label Training

To avoid the uncertainty issue, we have adopted the probabilistic training framework [8] to perform a multi-label training strategy for UIE learning. In multi-label training, the dataset we use contains four different labels, as shown in Figure 1. During the training phase, each time an image is input, we randomly select one of the four labels as the input label for training. The selection method is as follows:

$$l = label_i, i \in \{0, 1, 2, 3\} \tag{2}$$

In the formula, l is the label that serves as input during the network training process. $label_i$ is one of the four labels that we randomly select, where 0 represents the label in the UIEB dataset, 1 represents the label obtained through contrast adjustment, 2 is the label obtained through saturation adjustment, and 3 is the label obtained through gamma correction.

3.2. Network Architecture

Figure 2 illustrates the architecture of the NUAM-Net network. The network architecture consists of two branches, each including a feature extractor based on U-Net. Specifically, the upper branch aims to extract segmentation features from a single original underwater image, while the lower branch aims to construct UIE segmentation features using the input underwater image and its multiple labels. In the upper branch, we concatenate the original image and its conversion to the LAB color space along the channel dimension as input. In NUAM-Net, to enhance the parameters and extraction capabilities of the feature extractor, we replace the convolutional extractor with SE-ResNet50 (as shown in Table 1). Due to the lack of certain prior knowledge during the feature extraction process, we introduce the LAB color space of images to integrate prior information of the image. The LAB color space can better separate the color information and brightness information of the image, which is beneficial for the reconstruction of underwater images.

$$f = F_{extractor}(input \oplus input_{LAB}) \tag{3}$$

f is the feature extracted by the feature extractor, $F_{extractor}$ is the operation in the Feature Extractor, $input$ and $input_{LAB}$ are the RGB picture and LAB picture, and $\oplus$ operation is a concatenation operation performed on the channel dimensions of multiple features, merging them into a single feature.
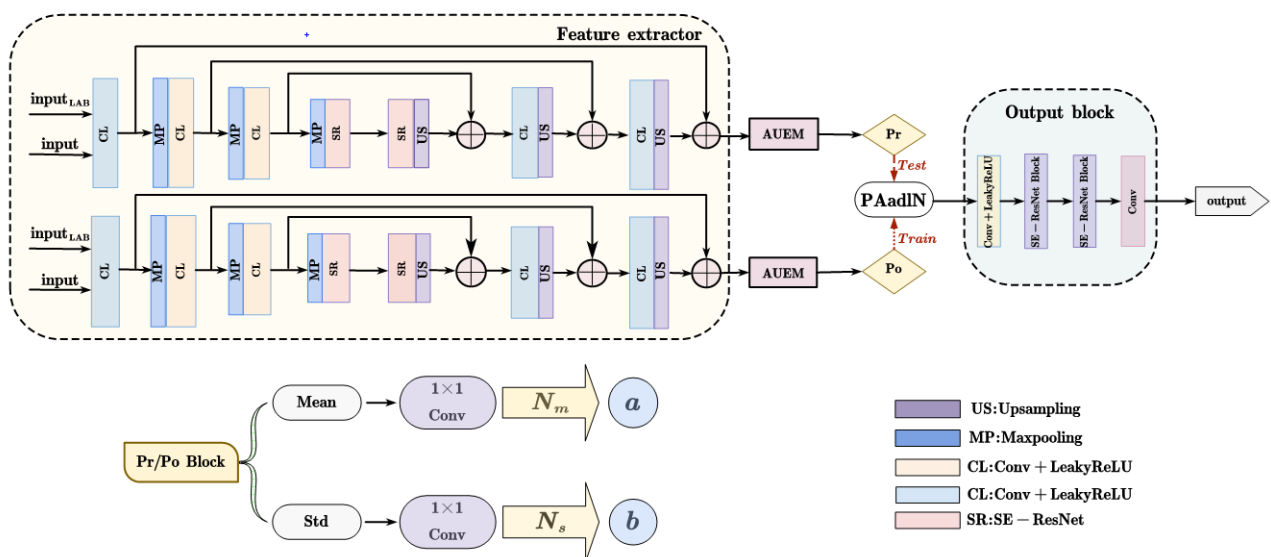

Figure 2. The network architecture of NUAM-Net. It consists of the feature extractor, PAdaIN, AUEM, and the output blocks. The extractor's architecture is similar to the U-Net.

Table 1. Structure of SE-ResNet50, consisting of three main blocks.

Output Size	Layers Used in SE-ResNet50	Number
112×112	conv, 7×7, 64, stride 2	1
	maxpool, 3×3, stride 2	1
56×56	conv, 1×1, 64 conv, 3×3, 64 conv, 1×1, 256 fc, [16, 256]	3
28×28	conv, 1×1, 128 conv, 3×3, 128 conv, 1×1, 512 fc, [32, 512]	4
14×14	conv, 1×1, 256 conv, 3×3, 256 conv, 1×1, 1024 fc, [64, 1024]	6
7×7	conv, 1×1, 512 conv, 3×3, 512 conv, 1×1, 2048 fc, [128, 2048]	3
1×1	global average pool, 100-d fc, softmax	1

The core part of this network lies in the feature enhancement transfer module after feature extraction. To obtain stronger features, we designed a probability enhancement module called AUEM, which takes the features extracted by the feature extractor as input. The output features are the result of concatenating the enhanced features with the original features along the channel dimension.

$$f_{enhance} = F_{AUEM}(ReLU(Conv(f_{in}))) \tag{4}$$

f_{in} is the input feature, $f_{enhance}$ is the feature enhanced by AUEM module, F_{AUEM} is the enhancement operation, $ReLU$ is the activation function, and $Conv$ is the convolution operation.

Next, we need to construct image enhancement style features based on the features extracted from a large sample. In an image, the style can be described by the mean and variance of the extracted features across each channel. This is mainly because they reflect the statistical characteristics of the color distribution and brightness distribution of the image, which, to a large extent, determine the appearance and feel of the image, such as whether it is bright, colorful, high contrast, etc. These statistical features provide important clues for image processing and analysis. During the training phase, we calculate the variance and standard deviation for the target image and the original image across each channel dimension.

$$\mu_l(c) = \frac{1}{H \times W} \sum_{x \in (0,W), y \in (0,H), c \in (0,C)} f_l(x, y, c) \tag{5}$$

$$\sigma_l^2(c) = \frac{1}{H \times W} \sum_{x \in (0,W), y \in (0,H), c \in (0,C)} (f_l(x, y, c) - \mu_c)^2 \tag{6}$$

$$\mu_{in}(c) = \frac{1}{H \times W} \sum_{x \in (0,W), y \in (0,H), c \in (0,C)} f_{in}(x, y, c) \tag{7}$$

$$\sigma_{in}^2(c) = \frac{1}{H \times W} \sum_{x \in (0,W), y \in (0,H), c \in (0,C)} (f_{in}(x, y, c) - \mu_c)^2 \tag{8}$$

where $\sigma_l(c)$ and $\mu_l(c)$ are the variance and mean of the label features across each channel, $\sigma_{in}(c)$ and $\mu_{in}(c)$ are the variance and mean of the image to be processed across each channel, H is the height of the image, W is the width (in pixels), C denotes the number of channels, and f_l and f_{in} are the feature vectors after extraction and enhancement, respectively.

After obtaining these features of mean and variance, we randomly sample from these features to construct the normal distribution functions for these means and variances. We perform the random sampling operation through convolutions, and then construct the functions based on the sampling results.

$$M_l = Normal(\mu_{Conv(\mu_l(c_i))}, \sigma_{Conv(\mu_l(c_i))}) \tag{9}$$

$$V_l = Normal(\mu_{Conv(\sigma_l(c_i))}, \sigma_{Conv(\sigma_l(c_i))}) \tag{10}$$

$$M_{in} = Normal(\mu_{Conv(\mu_{in}(c_i))}, \sigma_{Conv(\mu_{in}(c_i))}) \tag{11}$$

$$V_{in} = Normal(\mu_{Conv(\sigma_{in}(c_i))}, \sigma_{Conv(\sigma_{in}(c_i))}) \tag{12}$$

where V_l and M_l represent the normal distribution functions for the variance and mean of the label features, respectively, and V_{in} and M_{in} represent the normal distribution functions for the variance and mean of the input features, respectively.

The distributions of the mean and variance of the target image are used as the style parameters for image style transfer in the PAdaIN module (detailed in Section 3.3), and we specifically characterize this term in the loss function by using the KL divergence to describe the difference between the two distributions (elaborated in Section 3.5 in the description of KL divergence in the loss function). The purpose is to complete the transformation of the image style during the feature extraction process.

3.3. PAdaIN

In this paper, we treat the underwater enhancement problem as the domain style-transfer problem, and we therefore adopt Adaptive Instance Normalization with posterior distribution (*PAdaIN*) [8]:

$$PAdaIN(\mathbf{x}) = b(\frac{\mathbf{x} - \mu(\mathbf{x})}{\sigma(\mathbf{x})}) + a \tag{13}$$

Here, x represents the features of the content image, μ and σ denote the mean and standard deviation operations, respectively. b and a are two random samples drawn from the posterior distribution of the mean and standard deviation. Specifically, the posterior distribution can be learned through CVAEs [51]. A conditional variational autoencoder (CVAE), which combines raw data and their corresponding categories as inputs to the encoder, can be used to generate the data for specified categories.

$$a \sim \mathcal{N}_{m}\left(\mu(\mathbf{x}), \sigma^2(\mathbf{x})\right) \tag{14}$$

$$b \sim \mathcal{N}_{s}\left(m(\mathbf{x}), v^2(\mathbf{x})\right) \tag{15}$$

$\mathcal{N}_m$ and $\mathcal{N}_s$ represent the Gaussian distributions of mean and standard. The variables a and b are drawn randomly from the distributions of mean and standard deviation, respectively. $\mu(x)$ and $\sigma(x)$ represent the mean and standard deviation of the mean of the input image. $m(x)$ and $v(x)$ represent the mean and standard deviation of the standard deviation of the input image.

3.4. Adaptive Underwater Image Enhancement Module

AUEM consists of two parts; the architecture is shown in Figure 3. Firstly, the features of LAB color space images are concatenated with those of the original images. Subsequently, they undergo a convolution to adjust the feature dimensions, followed by the AIEM

(Adaptive Illumination Enhancement Module) [52], which consists of two components: Hierarchical Information Extraction (HIE) and IMAconv.
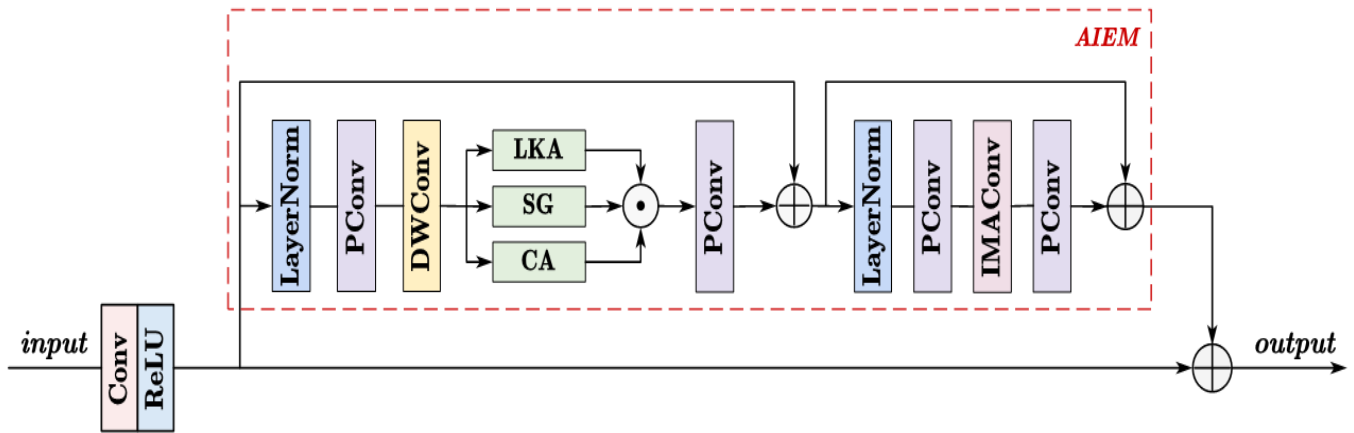

Figure 3. The overview of the AUEM. It consisted of a conv block and AIEM block. In the AIEM block, we try to combine and enhance the probabilistic feature. AIEM includes PConv, DWConv, LKA, SG, and IMAConv, which are five types of convolution blocks.

HIE employs three parallel operations: LKA, SG, and CA for feature extraction. Large-Kernel Attention, which is shown in Figure 4a, decomposes the feature extraction into three types of convolution: Depthwise Convolution (DW-conv), Depthwise Dilated Convolution (DW-D-Conv), and Point Convolution. DW-Conv is a 55 convolution, and DW-D-Conv is a 55 convolution with a dilation rate of 3. Point convolution is a 1×1 convolution. DW-Conv processes local structural information, DW-D-Conv is used to capture long-range dependencies, and Point Convolution is used for inter-channel interaction. The Simple Gate (SG), which is shown in Figure 4b, divides the features along the channel dimension into two parts, decomposing $f \in \mathbb{R}^{C \times H \times W}$ into $f_1 \in \mathbb{R}^{\frac{C}{2} \times H \times W}$ and $f_2 \in \mathbb{R}^{\frac{C}{2} \times H \times W}$. Then, these two features undergo a wise multiplication (the values at corresponding positions in two features are multiplied) operation. The channel attention-processed feature, which is shown in Figure 4c, $f \in \mathbb{R}^{C \times H \times W}$ passes through a channel attention module to obtain $f_1 \in \mathbb{R}^{1 \times H \times W}$. Then, it passes through a 1×1 convolution, followed by a ReLU activation function, and, subsequently, through another 1×1 convolution with a Sigmoid activation function to produce feature $f_2 \in \mathbb{R}^{1 \times H \times W}$. Finally, a wise multiplication operation is performed between f_2 and the original feature f to obtain the final feature.
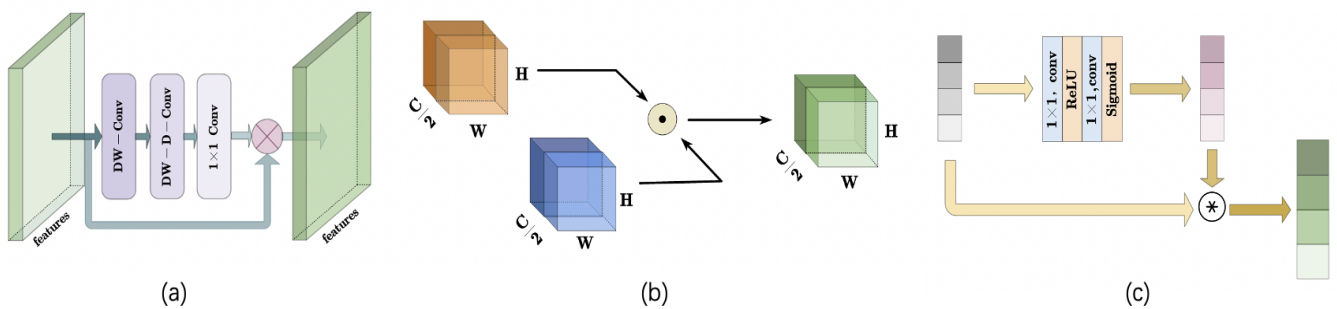

Figure 4. Structures of LKA, SG, and CA used in our AUEM module. (**a**) Large-Kernel Attention (LKA), (**b**) Simple Gate (SG), and (**c**) Channel Attention (CA).

The research motivation of IMAConv is to integrate information from different feature spaces and channels. As shown in Figure 5, features are divided into S branches (dividing the original feature into S parts along the channel dimension), each consisting of three concatenated convolutions. $Conv - 3$ is the dynamic convolution block, x_i is the divided

feature, and $\bar{x}_i$ is the original feature without x_i. $C_n(.)$ is the mapping function to combine each feature. The formulas of C_n are as follows:

$$C_n\left(x_f^i\right) = \begin{cases} A_1 x_f^i \left(1 - x_f^i\right) + x_f^i, n = 1 \\ A_{n-1} C_{n-1}\left(x_f^i\right) \left(1 - C_{n-1}\left(x_f^i\right)\right) + C_{n-1}\left(x_f^i\right), n > 1 \end{cases} \quad (16)$$

$Conv - 3$ employs the concept of dynamic convolution to assign weights to these three convolutional kernels. Dynamic convolution is the dynamic aggregation of multiple parallel convolution cores based on attention. Attention dynamically adjusts the weight of each convolution kernel based on the input, resulting in an adaptive dynamic convolution. After passing through AIEM, the enhanced features are concatenated with the features before inputting to the AIEM module. The output can be represented as

$$Out = F_{AIEM}(Convolution(input)) \oplus input \quad (17)$$

input is the feature extracted by the extractor, *Convolution* is the convolution operation to adjust the feature dimension, (i.e., the convolution operation to fuse the old feature to obtain the new dimension features), F_{AIEM} is the AIEM module operation, and $\oplus$ is the concatenation.
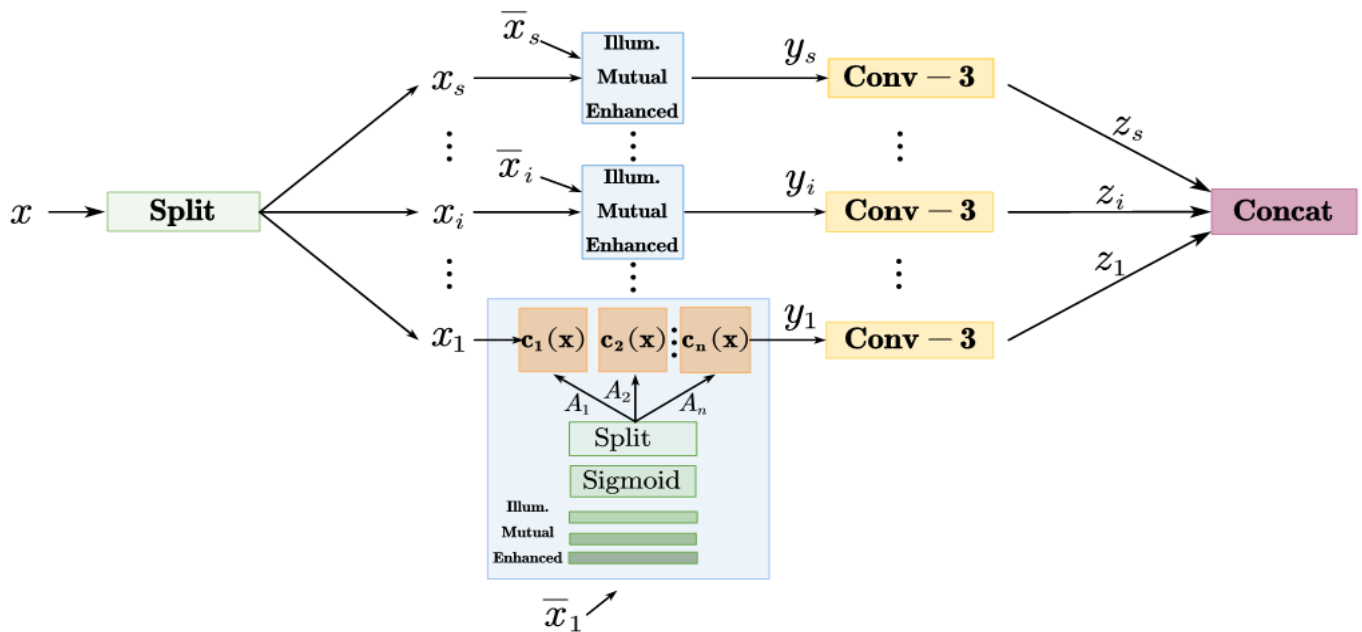

Figure 5. Structures of IMAConv used in our AUEM module.

3.5. Loss Function

In the supervising stage, we utilize the Mean Squared Error (MSE) as the loss function to quantify the discrepancies between the original and the output images, defined as

$$L_{mse} = \frac{1}{C \times H \times W} \sum_{x\epsilon[0,H]\ y\epsilon[0,W]\ c\epsilon[0,C]} \left[x_{label}(x,y,c) - x_{pro}(x,y,c)\right]^2 \quad (18)$$

where x_{label} represents the random label picture and x_{pro} is the network's output.

Additionally, to enhance the human perceptual quality of the processed images, we integrate a perceptual loss function, utilizing a pre-trained Vgg16 network as the perceptual evaluator [23]:

$$L_{per} = \frac{1}{N} \sum_{i=1}^{N} \left(F_{vgg16}(x_{label}) - F_{vgg16}(x_{pro\ i})\right)^2 \quad (19)$$

with N being the batch size and F_{vgg16} being the Vgg16 network equipped with pre-trained weights.

In addition to minimizing the enhancement loss, Kullback–Leibler (KL) divergences are utilized to assimilate the posterior distributions and the prior distributions. This process involves measuring the discrepancy between the posterior and prior distributions, ensuring that the enhanced image aligns well with both the desired enhancement characteristics and the prior knowledge captured by the distributions. By minimizing the KL divergences, the network learns to generate enhanced images that not only match the desired visual attributes but also adhere to the underlying statistical properties encoded in the prior distributions.

$$L_a = D_{KL}(\mathcal{N}_a(\mathbf{x}) \| \mathcal{N}_a(\mathbf{y}, \mathbf{x})) \tag{20}$$

$$L_d = D_{KL}(\mathcal{N}_d(\mathbf{x}) \| \mathcal{N}_d(\mathbf{y}, \mathbf{x})) \tag{21}$$

D_{KL} refers to the KL divergence between two distributions. $\|$ is used to denote the Kullback–Leibler divergence (KL divergence) between two probability distributions.

Finally, to align the processed images closely with their labels, we combined three parts as our model loss function.

$$L = L_{mse} + \beta(L_{per} + L_a + L_d) \tag{22}$$

This formulation aims not only to minimize the direct errors between images but also to improve their realism, their visual appeal to the human eye, and the distribution of the features, while preserving image detail and quality. β represents the weight; in our model, we choose $\beta = 0.1$.

4. Training Configuration

We utilized an extended multi-label dataset UIEBD (Underwater Image Enhancement Benchmark Dataset, which is a multi-label underwater image enhancement dataset) [8]. Some training datasets pictures are shown in Figure 6. A challenge encountered prior to training the probability network was that existing UIE datasets typically provide a single reference map for each degraded underwater image. To facilitate the application of the probability network, we augmented existing UIE datasets by generating multiple reference images. The new dataset we adopted is based on UIEBD [10], a real-world UIE dataset comprising 890 underwater images along with corresponding reference maps.

In the original UIEB (Underwater Image Enhancement Benchmark, a dataset that was proposed in 2020) [10], the authors employed 12 state-of-the-art enhancement algorithms to generate potential groundtruth. Volunteers were then asked to subjectively select the best image among pairwise comparisons of the original underwater image and the 12 enhanced images, with the chosen image serving as the final reference. Ambiguity was addressed in UIEBD through contrast and saturation adjustments as well as gamma correction, given that distortions in underwater images primarily manifest in aspects such as contrast, saturation, brightness, and color.

It is important to note that our aim was to generate ambiguous labels rather than to significantly alter the original labels. Contrast and saturation adjustments were performed using a simple linear transformation formula, where $y = \alpha(x - m) + x$, with x and y representing the input and output, respectively, and m denoting the mean value for each channel. α stands for the adjustment coefficient, which remains consistent for all pixels in contrast adjustment and is determined by each pixel itself in saturation adjustment.

To produce a more reliable reference image, we initially created two adjusted versions for each method (i.e., over-adjustment and under-adjustment), then selected the better one as the potential label. Consequently, we obtained four reference images (including the original label) for each original underwater image, reflecting the uncertainty inherent in the groundtruth recording process.
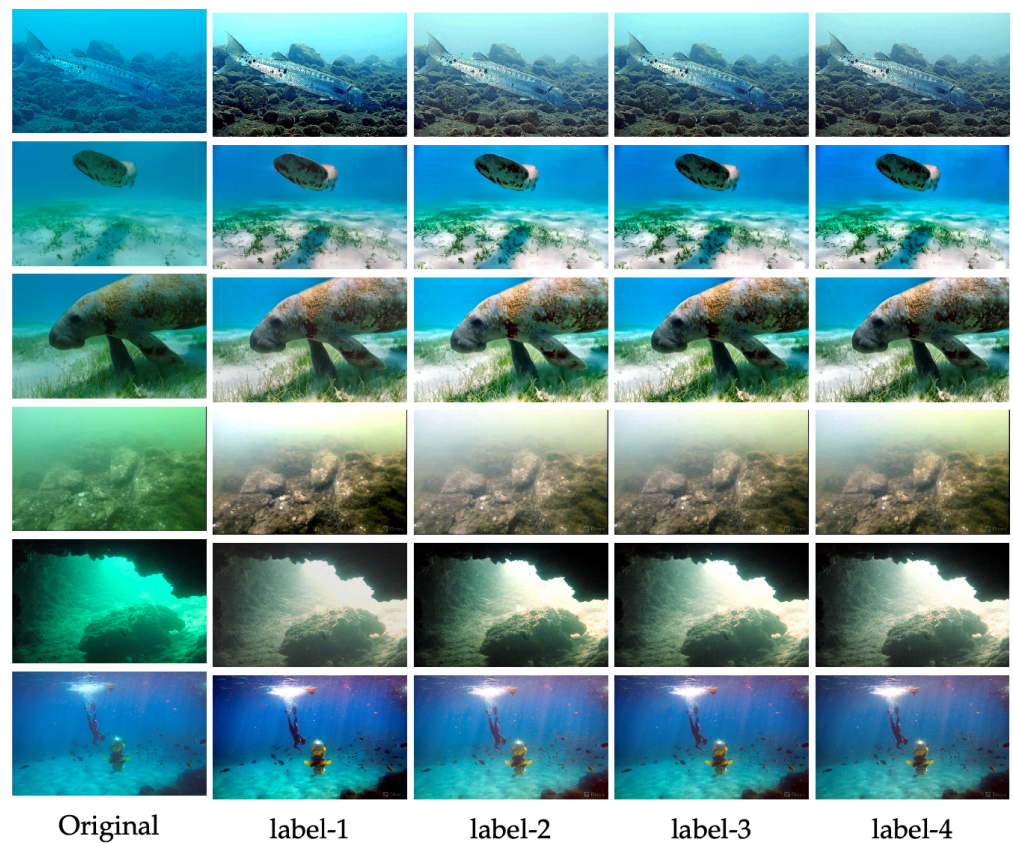
| | | | | |
| Original | label-1 | label-2 | label-3 | label-4 |

Figure 6. Examples of the extended UIEBD dataset, including 4 labels. Label-1 denotes the manually selected label in the original UIEBD dataset, label-2 is the contrast adjustment result, label-3 is the saturation adjustment result, and label-4 is the gamma correction result.

5. Experiments

In this section, we conducted comprehensive experimental evaluation of the proposed method. At first, we describe the implementation details and the validation dataset. Secondly, we introduce the evaluation criteria and compare our method with eight state-of-the-art UIE methods on the UIEBD dataset in terms of both qualitative and quantitative evaluations. Finally, we evaluate the effectiveness of the key components in our proposed method through the ablation study.

5.1. Implementation Details

Our method was implemented in Pytorch and the NUAM-Net model was trained on an NVIDIA RTX 4090 GPU (Santa Clara, CA, USA) with ADAM optimizer, where the learning rate was 1×10^{-4}, the number of training epochs was 500, the batch size was 1, and the image was resized to a resolution of 256 × 256. During the training, we performed random rotations and horizontal–vertical flips for data augmentation.

5.2. Datasets

We validate our method on the popular UIEBD benchmark, and we followed a previous work [8] in utilizing the first 700 original samples for training and the remaining 190 images for testing.

5.3. Performance Criteria

To evaluate the enhancement performance of our method, we employ SSIM (Structural Similarity) [53], PSNR (Peak Signal-to-Noise Ratio) [54], and MOS (Mean Opinion Score) [55] metrics. SSIM and PSNR are full-reference metrics computed based on the manually selected well-enhanced reference image (label image) in UIEBD to ensure a

fair comparison with existing methods. Additionally, we conduct subjective testing to understand user preferences for the results generated by each UIE method. We use MOS to quantify subjective evaluations. We invited 10 participants (5 males and 5 females) to participate in the subjective testing. Original and enhanced underwater images were simultaneously displayed on the screen. Subjective ratings for each image were assessed on a three-level scale according to the following criteria: 3 (excellent), 2 (fair), 1 (poor). Evaluation metrics include color distortion, contrast enhancement, naturalness preservation, brightness improvement, and artifact suppression.

5.4. Comparison Methods

We compared NUAM-Net with eight UIE methods, including two model-free methods (GC, Retinex), one popular prior-based method (DCP), three state-of-the-art deep learning methods (Deep-SESR, Water-Net, Ucolor), and two advanced probabilistic network-based methods (PUIE-MC, PUIE-MP). We report the results of all compared methods using the original implementations provided by their authors in the same experiments to ensure fairness of comparison.

5.5. Results

Table 2 summarizes the quantitative comparison results on the UIEBD dataset. It can be seen that our NUAM-Net achieves highly competitive performance and outperforms other methods in PSNR by a significant margin. Specifically, prior-based methods obtain relatively poor results because they heavily rely on prior knowledge-driven approximate imaging models, limiting their generalization ability to more complex scenarios. We found that the performance of deep learning methods is significantly better than handcrafted methods and our NUAM-Net achieves the best results, showing the effectiveness of the proposed method. We further present qualitative comparison results in Figure 7. It can be observed that, although most methods can enhance contrast to some extent, serious visual defects still exist due to undesirable color adjustments or artifacts. For example, GC and Retinex exhibit unnatural color saturation and blurred image details. Prior-based methods can improve contrast, but color is severely degraded in these cases. Water-Net and Ucolor often produce low-quality results. Due to the enriched color perception and long-range interaction, our method performs well in all these cases and produces consistently cleaner visual quality and more natural fine textures than the state-of-the-art PUIE-Net.

Table 2. Quantitative results on the UIEBD test dataset. We report the metrics of PSNR, SSIM, and MOS values for evaluation. Higher values indicate better performance. The best results are highlighted in red.

Method	PSNR	SSIM	MOS
GC	17.79	0.79	2.1
Retinex	15.47	0.75	1.8
DCP	15.05	0.72	1.6
Deep SESR	17.10	0.63	1.8
Water-Net	21.19	0.84	2.3
Ucolor	21.55	0.85	2.5
PUIE-MC	21.68	0.86	2.4
PUIE-MP	21.66	0.86	2.4
NUAM-Net *	22.38	0.87	2.9

* is our network.
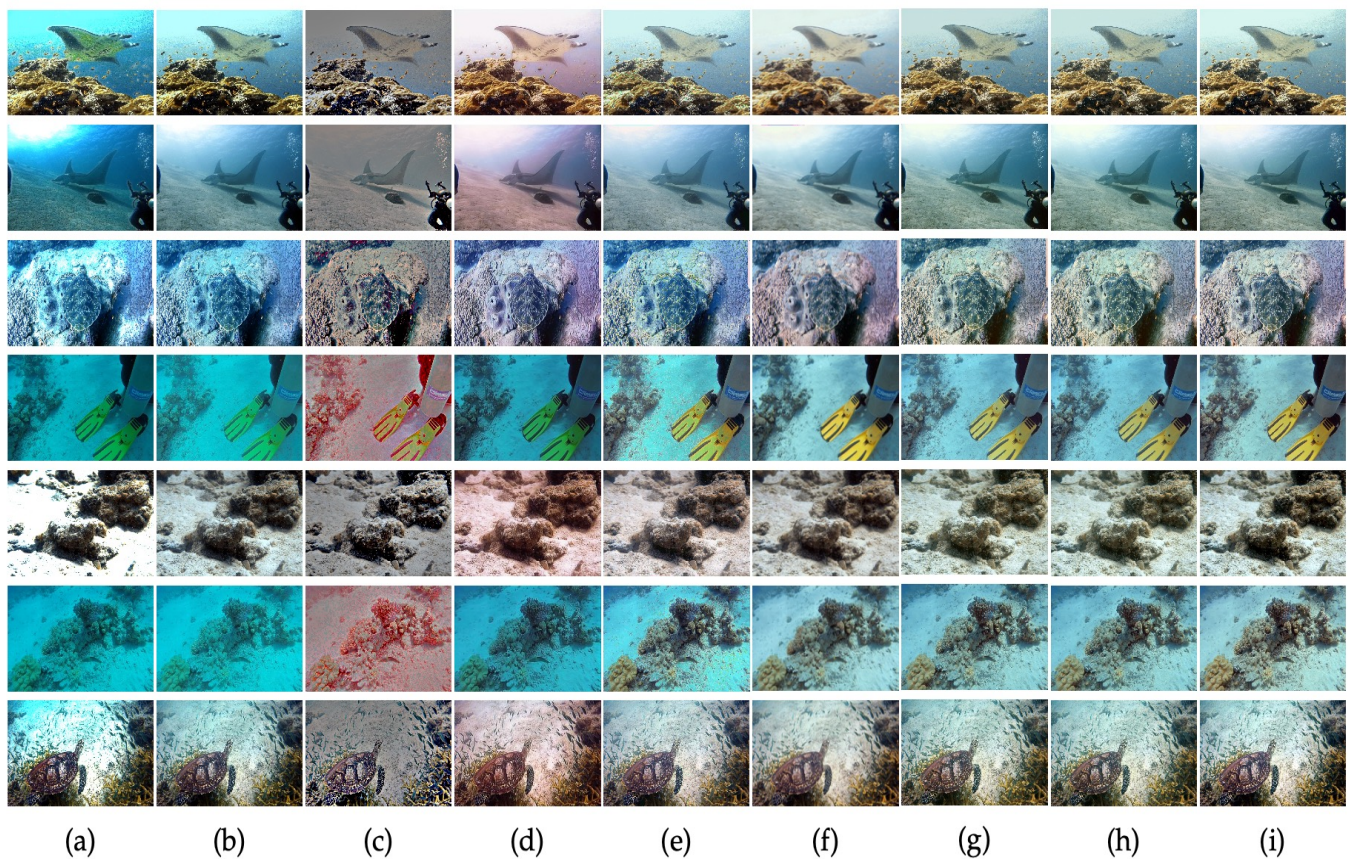
Figure 7. Qualitative results of the UIEBD test dataset. (**a**) DCP, (**b**) GC, (**c**) Retinex, (**d**) SESR, (**e**) Water-Net, (**f**) Ucolor, (**g**) PUIE-MC, (**h**) PUIE-MP, (**i**) Ours.

5.6. Ablation

We conducted ablation experiments on our network, evaluating the performance of some variants of the proposed method with the backbone, backbone+LAB, and backbone+LAB+ AUEM. As shown in Table 3, enriching the color perception by extracting features from the RGB space and wider color-represented LAB space leads to reasonable improvements in the PSNR metric. With the well-designed AUEM, modeling the long-range spatial and channel interactions from both local and long-range receptive fields, the backbone+LAB+AUEM variant further promotes the PSNR score by 0.57. The qualitative results in Figure 8 also show the gains of our proposed method.

We compare the features extracted by the our network and the backbone network. In Figure 9, the higher the number of green flashing dots in the renderings, the higher the number of features that have been extracted. It is obvious that our network can extract more features.

Table 3. Ablation experiments on the UIEBD test dataset.

Network	PSNR	SSIM
Backbone	21.68	0.86
Backbone+LAB	21.81	0.86
Backbone+LAB+AUEM	22.38	0.87

We prove that our network structure has some advantages. We compare two modules: (1) AUEM without LKA, SG, and CA and (2) AUEM. The results are shown in Table 4. Furthermore, we set an extra experiment to prove that our network's structure has advantages (to some extent). We replace the AUEM module by the same capacity convolution

blocks, and we compare NUAM-Net and conv blocks Net. The results are shown in Table 5. We believe that the advantage of our network structure lies in its ability to fuse multi-scale spatial and channel features, as well as the additional color domain that can provide more information.
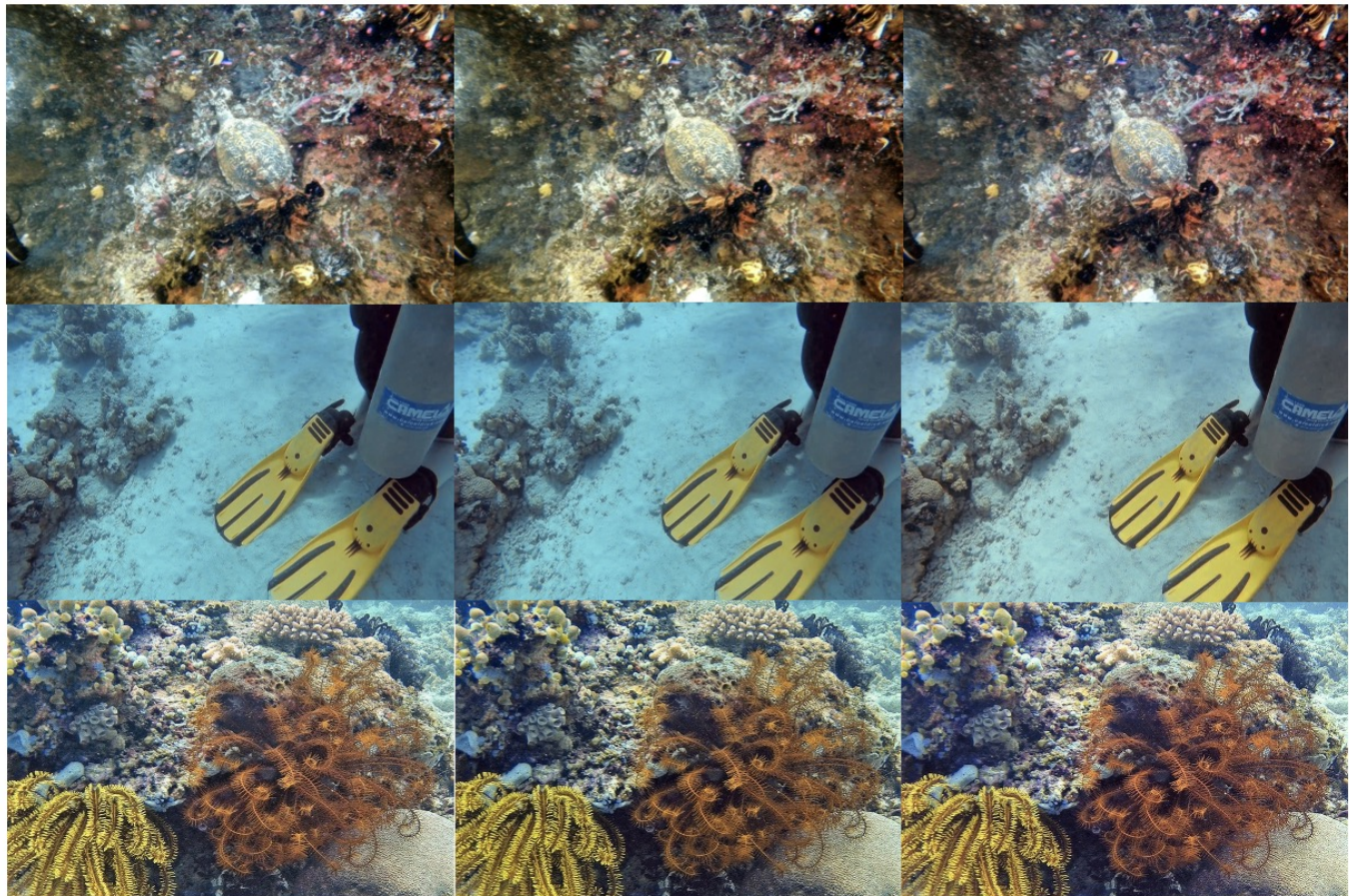
backbone   backbone + LAB   backbone + LAB+ AUEM

Figure 8. Enhancement examples of our ablation studies. We show the enhanced images of backbone, backbone+LAB, and backbone+LAB+AUEM on a subset of the UIEBD test data. It is evident from the image that our network demonstrates significant improvement in enhancement effectiveness.

Table 4. Ablation experiments on AUEM module.

Network	PSNR	SSIM
Backbone+LAB+AUEM (w/o LKA+SG+CA)	22.12	0.86
Backbone+LAB+AUEM	22.38	0.87

Table 5. Replacement experiment on AUEM module.

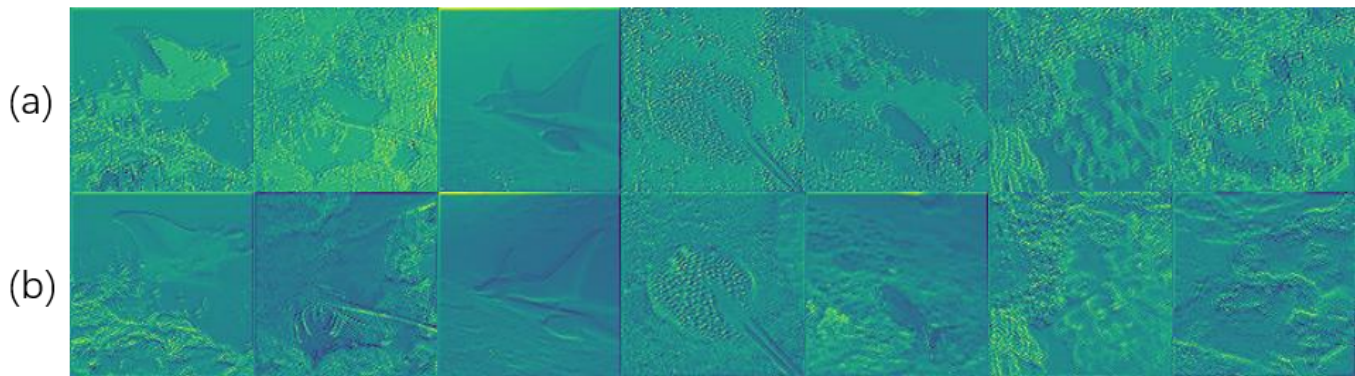| Network	PSNR	SSIM
Backbone+LAB+Conv	22.07	0.86
Backbone+LAB+AUEM	22.38	0.87

Figure 9. Pictures show two extracted results of backbone and our network. (**a**) represents the feature extracted by our network and (**b**) represents the feature extracted by backbone network.

6. Discussion

Based on our experimental results, the performance of our network is remarkably outstanding. We attribute this success primarily to the incorporation of AUEM and the physical priors embedded in LAB color space images. Through ablation experiments, it becomes evident that the most influential factor is AUEM. This module significantly expands the network's receptive field and enhances channel-wise and spatial interactions to a considerable extent. As a result, it plays a vital role in achieving such impressive performance. This finding holds significant implications for addressing the enhancement tasks of underwater images in current probabilistic networks, serving as valuable inspiration for future research in this domain.

7. Conclusions

In this paper, aiming to address the local perception and color perception limitations of current UIE methods, we proposed NUAM-Net for underwater image enhancement. Specifically, our NUAM-Net models the long-range spatial and channel interactions with a novel AUEM module, enabling both local and long-range receptive fields for large-scale degradation perception. Moreover, NUAM-Net extracts features from RGB and an extra LAB color space to fully utilize the fine-grained color degradation clues of underwater images. Based on the probabilistic training framework, our NUAM-Net achieves highly competitive results on the popular UIEBD benchmark compared to the state-of-the-art model-free, prior-based, and learning-based UIE methods. In the future, we plan to extend our method to vision-based underwater systems, such as underwater visual SLAM and visual 3D reconstruction.

Author Contributions: Methodology, F.G.; Software, Z.W.; Validation, Z.W.; Formal analysis, Z.W.; Investigation, Z.W. and F.G.; Resources, H.S.; Data curation, Z.W. and Y.Z.; Writing—original draft, Z.W.; Writing—review & editing, Y.R. and J.D.; Visualization, Z.W.; Supervision, F.G., H.S., Y.R. and J.D.; Project administration, J.D.; Funding acquisition, H.S. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Jian, M.; Liu, X.; Luo, H.; Lu, X.; Yu, H.; Dong, J. Underwater image processing and analysis: A review. *Signal Process. Image Commun.* **2021**, *91*, 116088. [CrossRef]
2. Ghafoor, H.; Noh, Y. An overview of next-generation underwater target detection and tracking: An integrated underwater architecture. *IEEE Access* **2019**, *7*, 98841–98853. [CrossRef]
3. Heshmati-Alamdari, S.; Nikou, A.; Dimarogonas, D.V. Robust trajectory tracking control for underactuated autonomous underwater vehicles in uncertain environments. *IEEE Trans. Autom. Sci. Eng.* **2020**, *18*, 1288–1301. [CrossRef]
4. Wu, Y.; Ta, X.; Xiao, R.; Wei, Y.; An, D.; Li, D. Survey of underwater robot positioning navigation. *Appl. Ocean. Res.* **2019**, *90*, 101845. [CrossRef]
5. Chutia, S.; Kakoty, N.M.; Deka, D. A review of underwater robotics, navigation, sensing techniques and applications. In Proceedings of the 2017 3rd International Conference on Advances in Robotics, New Delhi, India, 28 June–2 July 2017; pp. 1–6.
6. Suo, F.; Huang, K.; Ling, G.; Li, Y.; Xiang, J. Fish keypoints detection for ecology monitoring based on underwater visual intelligence. In Proceedings of the 2020 16th International Conference on Control, Automation, Robotics and Vision (ICARCV), Shenzhen, China, 13–15 December 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 542–547.
7. Aguzzi, J.; Iveša, N.; Gelli, M.; Costa, C.; Gavrilovic, A.; Cukrov, N.; Cukrov, M.; Cukrov, N.; Omanovic, D.; Štifanić, M.; et al. Ecological video monitoring of Marine Protected Areas by underwater cabled surveillance cameras. *Mar. Policy* **2020**, *119*, 104052. [CrossRef]
8. Fu, Z.; Wang, W.; Huang, Y.; Ding, X.; Ma, K.K. Uncertainty inspired underwater image enhancement. In *Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 465–482.
9. Islam, M.J.; Luo, P.; Sattar, J. Simultaneous Enhancement and Super-Resolution of Underwater Imagery for Improved Visual Perception. *arXiv* **2020**, arXiv:2002.01155.
10. Li, C.; Guo, C.; Ren, W.; Cong, R.; Hou, J.; Kwong, S.; Tao, D. An Underwater Image Enhancement Benchmark Dataset and Beyond. *arXiv* **2019**, arXiv:1901.05495.
11. Li, C.; Anwar, S.; Hou, J.; Cong, R.; Guo, C.; Ren, W. Underwater image enhancement via medium transmission-guided multi-color space embedding. *IEEE Trans. Image Process.* **2021**, *30*, 4985–5000. [CrossRef]
12. Yang, M.; Hu, J.; Li, C.; Rohde, G.; Du, Y.; Hu, K. An In-Depth Survey of Underwater Image Enhancement and Restoration. *IEEE Access* **2019**, *7*, 123638–123657. [CrossRef]
13. Liu, R.; Fan, X.; Zhu, M.; Hou, M.; Luo, Z. Real-World Underwater Enhancement: Challenges, Benchmarks, and Solutions Under Natural Light. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *30*, 4861–4875. [CrossRef]
14. Fabbri, C.; Jahidul Islam, M.; Sattar, J. Enhancing Underwater Imagery using Generative Adversarial Networks. *arXiv* **2018**, arXiv:1801.04011.
15. Pizer, S.M.; Johnston, R.E.; Ericksen, J.P.; Yankaskas, B.C.; Muller, K.E. Contrast-Limited Adaptive Histogram Equalization: Speed and Effectiveness. In Proceedings of the First Conference on Visualization in Biomedical Computing, Atlanta, Georgia, 22–25 May 1990.
16. Liu, Y.C.; Chan, W.H.; Chen, Y.Q. Automatic White Balance for Digital Still Camera. *IEEE Trans. Consum. Electron.* **1995**, *41*, 460–466.
17. Rahman, Z.; Jobson, D.J.; Woodell, G.A. Multi-scale retinex for color image enhancement. In Proceedings of the 3rd IEEE International Conference on Image Processing, Lausanne, Switzerland, 16–19 September 1996.
18. Ancuti, C.; Ancuti, C.O.; Haber, T.; Bekaert, P. Enhancing underwater images and videos by fusion. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 81–88.
19. Ancuti, C.O.; Ancuti, C.; Vleeschouwer, C.D.; Bekaert, P. Color Balance and Fusion for Underwater Image Enhancement. *IEEE Trans. Image Process.* **2017**, *27*, 379–393. [CrossRef] [PubMed]
20. Fu, X.; Zhuang, P.; Huang, Y.; Liao, Y.; Zhang, X.P.; Ding, X. A retinex-based enhancing approach for single underwater image. In Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP), Paris, France, 27–30 October 2014.
21. Gao, S.B.; Zhang, M.; Zhao, Q.; Zhang, X.S.; Li, Y.J. Underwater image enhancement using adaptive retinal mechanisms. *IEEE Trans. Image Process.* **2019**, *28*, 5580–5595. [CrossRef]
22. Chiang, J.Y.; Chen, Y.C. Underwater Image Enhancement by Wavelength Compensation and Dehazing. *IEEE Trans. Image Process.* **2012**, *21*, 1756–1769. [CrossRef]
23. Galdran, A.; Pardo, D.; Picón, A.; Alvarez-Gila, A. Automatic Red-Channel underwater image restoration. *J. Vis. Commun. Image Represent.* **2015**, *26*, 132–145. [CrossRef]
24. He, K.; Sun, J.; Tang, X. Single Image Haze Removal Using Dark Channel Prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 2341–2353.
25. Li, C.Y.; Guo, J.C.; Cong, R.M.; Pang, Y.W.; Wang, B. Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior. *IEEE Trans. Image Process.* **2016**, *25*, 5664–5677. [CrossRef]
26. Berman, D.; Treibitz, T.; Avidan, S. Diving into haze-lines: Color restoration of underwater images. In Proceedings of the British Machine Vision Conference 2017, BMVC 2017, London, UK, 4–7 September 2017; BMVA Press: Durham, UK, 2017.

27. Akkaynak, D.; Treibitz, T. Sea-thru: A method for removing water from underwater images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 1682–1691.

28. Li, J.; Skinner, K.A.; Eustice, R.M.; Johnson-Roberson, M. WaterGAN: Unsupervised generative network to enable real-time color correction of monocular underwater images. *IEEE Robot. Autom. Lett.* **2017**, *3*, 387–394. [CrossRef]

29. Li, C.; Guo, J.; Guo, C. Emerging from water: Underwater image color correction based on weakly supervised color transfer. *IEEE Signal Process. Lett.* **2018**, *25*, 323–327. [CrossRef]

30. Guo, Y.; Li, H.; Zhuang, P. Underwater image enhancement using a multiscale dense generative adversarial network. *IEEE J. Ocean. Eng.* **2019**, *45*, 862–870. [CrossRef]

31. Li, C.; Anwar, S. Underwater Scene Prior Inspired Deep Underwater Image and Video Enhancement. *Pattern Recognit.* **2019**, *98*, 107038. [CrossRef]

32. Jamadandi, A.; Mudenagudi, U. Exemplar-based underwater image enhancement augmented by wavelet corrected transforms. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 11–17.

33. Uplavikar, P.M.; Wu, Z.; Wang, Z. All-in-One Underwater Image Enhancement Using Domain-Adversarial Learning. *arXiv* **2019**, arXiv:1905.13342.

34. Kar, A.; Dhara, S.K.; Sen, D.; Biswas, P.K. Zero-shot single image restoration through controlled perturbation of koschmieder's model. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 16205–16215.

35. Yang, H.H.; Huang, K.C.; Chen, W.T. Laffnet: A lightweight adaptive feature fusion network for underwater image enhancement. In Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 30 May–5 June 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 685–692.

36. Hill, M.L.; Kender, J.R.; Natsev, A.I.; Smith, J.R.; Xie, L. Using Near-Duplicate Video Frames to Analyze, Classify, Track, and Visualize Evolution and Fitness of Videos. US Patent 8,798,400, 5 August 2014.

37. Jiang, Q.; Zhang, Y.; Bao, F.; Zhao, X.; Zhang, C.; Liu, P. Two-step domain adaptation for underwater image enhancement. *Pattern Recognit.* **2022**, *122*, 108324. [CrossRef]

38. Xue, X.; Hao, Z.; Ma, L.; Wang, Y.; Liu, R. Joint luminance and chrominance learning for underwater image enhancement. *IEEE Signal Process. Lett.* **2021**, *28*, 818–822. [CrossRef]

39. Panetta, K.; Kezebou, L.; Oludare, V.; Agaian, S. Comprehensive underwater object tracking benchmark dataset and underwater image enhancement with GAN. *IEEE J. Ocean. Eng.* **2021**, *47*, 59–75. [CrossRef]

40. Huo, F.; Li, B.; Zhu, X. Efficient wavelet boost learning-based multi-stage progressive refinement network for underwater image enhancement. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 1944–1952.

41. Qi, Q.; Zhang, Y.; Tian, F.; Wu, Q.J.; Li, K.; Luan, X.; Song, D. Underwater image co-enhancement with correlation feature matching and joint learning. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *32*, 1133–1147. [CrossRef]

42. Jiang, N.; Chen, W.; Lin, Y.; Zhao, T.; Lin, C.W. Underwater image enhancement with lightweight cascaded network. *IEEE Trans. Multimed.* **2021**, *24*, 4301–4313. [CrossRef]

43. Bahdanau, D.; Cho, K.; Bengio, Y. Neural machine translation by jointly learning to align and translate. *arXiv* **2014**, arXiv:1409.0473.

44. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.

45. Liu, X.; Ma, Y.; Shi, Z.; Chen, J. Griddehazenet: Attention-based multi-scale network for image dehazing. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 7314–7323.

46. Wang, W.; Wang, A.; Ai, Q.; Liu, C.; Liu, J. AAGAN: Enhanced single image dehazing with attention-to-attention generative adversarial network. *IEEE Access* **2019**, *7*, 173485–173498. [CrossRef]

47. Jiang, X.; Lu, L.; Zhu, M.; Hao, Z.; Gao, W. Haze relevant feature attention network for single image dehazing. *IEEE Access* **2021**, *9*, 106476–106488. [CrossRef]

48. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.

49. Zhu, X.; Cheng, D.; Zhang, Z.; Lin, S.; Dai, J. An empirical study of spatial attention mechanisms in deep networks. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 6688–6697.

50. Guo, M.H.; Lu, C.Z.; Liu, Z.N.; Cheng, M.M.; Hu, S.M. Visual attention network. *Comput. Vis. Media* **2023**, *9*, 733–752. [CrossRef]

51. Debbagh, M. Learning structured output representations from attributes using deep conditional generative models. *arXiv* **2023**, arXiv:2305.00980.

52. Zou, W.; Gao, H.; Ye, T.; Chen, L.; Yang, W.; Huang, S.; Chen, H.; Chen, S. VQCNIR: Clearer Night Image Restoration with Vector-Quantized Codebook. *arXiv* **2023**, arXiv:2312.08606.

53. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]

54. Welstead, S.T. *Fractal and Wavelet Image Compression Techniques*; Spie Press: Bellingham, WA, USA, 1999; Volume 40.
55. Streijl, R.C.; Winkler, S.; Hands, D.S. Mean opinion score (MOS) revisited: Methods and applications, limitations and alternatives. *Multimed. Syst.* **2016**, *22*, 213–227. [CrossRef]