

DAPNet: Dual Attention Probabilistic Network for Underwater Image Enhancement

Xueyong Li, Rui Yu, Weidong Zhang, Wenyi Zhao, Guojia Hou, and Zheng Liang

Abstract

Underwater images are prone to being impacted by light attenuation and scattering, leading to color distortion, decreased contrast, visual blur, and various other issues. However, most deep methods based on deep learning techniques are trained on underwater synthetic images, ignoring the ambiguity of reference images. It is easy to suffer from under-enhancement or over-enhancement. To solve these problems, we propose a remarkably efficient and resilient method for enhancing underwater images, called DAPNet. Concretely, we introduce extended information blocks in the encoder to make up for the information loss during the convolution and downsampling processes and then convert the input original underwater image into a low-dimensional representation and use the decoder to perform convolution and upsampling. In the encoding and decoding process, a dual attention module be embedded to focus on the critical position information of the image by the spatial attention and the important channel information of the image via the channel attention mechanism, thereby enhancing the network's ability to detect different channels and position details. Meanwhile, the output features are transformed via adaptive instance normalization to generate samples. Finally, stable enhancement results are obtained from the sample space through monte carlo likelihood estimation. Extensive experiments on three benchmarks show that our DAPNet surpasses the current state-of-the-art methods in terms of enhancement effect. Additionally, our DAPNet also verified its usability in the enhancement of low-light images and the dehazing of images.

Index Terms

Underwater image enhancement, underwater imaging, encoder, decoder, adaptive instance normalization.

This work was supported in part by the Natural Science Foundation of Henan Province under Grant 232300420428, in part by the Teacher Education Curriculum Reform Research of Henan Province under Grants 2024-JSJYYB-029, in part by the National Natural Science Foundation of China under Grant 61901240, in part by the Key Specialized Research and Development Program of Science and Technology of Henan Province under Grant 232102210018, in part by the Key Research and Development Project of Henan Province under Grants 231111220700, in part by the National College Students' Innovation and Entrepreneurship Training Program under Grant 202310467031, and in part by the Teacher Education Curriculum Reform Research of Henan Institute of Science and Technology under Grant 2024JSJY04. (*Corresponding author: Weidong Zhang.*)

Xueyong Li and Rui Yu are with the School of Computer Science, Henan Institute of Science and Technology, Xinxiang, China (e-mail: lixueyonghist@163.com, yuruifancy@163.com).

Weidong Zhang is with the School of Information Engineering and the Institute of Computer Applications, Henan Institute of Science and Technology, Xinxiang, China (e-mail: zwd_wd@163.com).

Wenyi Zhao is with the School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing, China (e-mail: 2013211876@bupt.edu.cn).

Guojia Hou is with the College of Computer Science and Technology, Qingdao University, Shandong, China (e-mail: hgjouc@126.com).

Zheng Liang is with the School of Internet, Anhui University, Hefei, Anhui 230093, China (e-mail: zliang@ahu.edu.cn).

I. INTRODUCTION

Recently, the strategic center of the international community has been placed on the development and utilization of marine resources. Underwater robot research and ocean exploration play a vital role in developing marine resources, and the low visual images will directly affect the navigation of underwater robots and target recognition, etc. Underwater image quality degradation mainly manifests in low contrast, low brightness, color cast, and blur. The detailed reason is that light is absorbed and scattered in water, which is also a major hindrance to computer vision tasks [1–3].

At present, deep learning methods have emerged endlessly in underwater image enhancement. Most deep learning methods take reference images and raw underwater images as input, where the reference images assist network learning and guide the enhancement process. This causes these algorithms to rely on the quality of reference images (*e.g.*, [4–6] synthesized paired underwater images). Reference images are mainly derived from underwater synthetic images, which usually only cover specific scenes and conditions and lack diversity. Computers often cannot fully simulate the actual underwater environment when generating synthetic images. This may lead to a large deviation between the reference image and GroundTruth, making the method not well applicable to various scenarios but increasing the uncertainty of the image enhancement effect. Although some underwater reference image datasets have high visual quality, most existing UIE methods generate clean images through point estimation by sampling points in the latent space to generate unique images. These methods lack diversity and make it difficult to control multiple features of the image accurately. To enhance image quality, the development of a comprehensive methodology becomes imperative.

To address the aforementioned issues, we propose a dual attention probabilistic network called DAPNet. Specifically, two modules are proposed to improve the quality of samples generated by CVAE [7]. First, the encoder and decoder are enhanced with the integration of an extended information module to extract additional contextual features and enrich the result features generated by the encoder. Then, Introducing a dual attention module to make the network pay more attention to channels and spatial relationships to capture key features. Furthermore, the styles of the features are aligned through adaptive intensity normalization to generate a set of samples, and the final enhancement results are obtained from the samples through monte carlo likelihood estimation [8]. Table II shows that our network can minimize the blurriness of synthetic images and achieve better results. The key outcomes of this study are summarized as follows:

- We introduce extended information blocks into the encoder to enable the encoder to learn more complex feature representations, capture critical features in the data, and make up for the problem of partial information loss during the downsampling process.
- We design a detail enhancement strategy, which enhances the correlation between feature channels and space by introducing dual attention, enhances the model's awareness of essential features of channels and spaces, and effectively inhibits the appearance of noise and artifacts.
- We reconstructed the loss and introduced adaptive parameters to reflect data characteristics more accurately, thereby improving the accuracy of the model. Correcting the prior distribution by learning a meaningful

posterior distribution makes the prior and posterior distributions more consistent.

II. RELATED WORK

Researchers have explored many image enhancement techniques to restore degraded underwater images, which can be broadly divided into traditional methods [9–12] and deep learning-based methods [13–16].

Traditional Methods. Traditional methods for image enhancement can be roughly grouped into two types. The first category is non-physical methods, which involve directly adjusting image pixel values to enhance image quality without considering the physical degradation process specific to underwater images. Representative methods include color channel-based methods [17–20], retinex-based methods [21–25], and histogram methods [26–28]. For example, Zhang et al. [29] proposed first to correct the image color to obtain the corrected image globally and locally enhanced images. Accordingly, they introduced a fusion strategy based on weighted wavelet visual perception to obtain underwater images of exceptional quality. These methods have achieved some success in improving image quality. However, they also have some limitations. Non-physical methods are usually based on specific situations, and it is difficult to ensure sound enhancement effects under different underwater imaging conditions. Problems of under-enhancement or over-enhancement may occur, and they lack generalization capabilities. For example, Liu et al. [30] do not consider water depth.

The second category is methods based on physical models, which require prior knowledge and accurate physical parameter estimation. These include Laplace reflection prior [31], dark channel prior [32], etc. For instance, Zhang et al. [33] proposed a weighted fusion strategy that uses the brightness information of the image to achieve a balanced enhancement of brightness and improvement of contrast. Peng et al. [34] developed a method to recover underwater images by using image blur and light absorption to estimate scene depth. This method takes into account the association between the blur intensity of the underwater image, the observation distance, and the concentration of impurities in the water. By analyzing image blur and combining it with a light absorption model, depth information of underwater scenes is accurately and reliably estimated to restore image quality. Li et al. [35] introduced an adaptive background light estimation method and used multicolor components and light attenuation methods to eliminate color casts, effectively enhance image details, and restore image texture. Traditional methods, which rely on physical models, are designed to address the challenges posed by underwater scattering and absorption parameters. These methods leverage prior knowledge to compensate for the loss of information effectively. However, these approaches tend to be sensitive to factors specific to underwater scenes and lack robustness. The accuracy of parameter estimation is highly dependent on the precision of prior knowledge. Therefore, in some complex underwater scenes, once these priors fail, the image recovery effect is often unsatisfactory.

Deep Learning Methods. Methods based on deep learning have developed more and more rapidly this year. Generally speaking, deep learning-based techniques for enhancing underwater images are roughly divided into CNN-based and GAN-based methods.

GAN is an unsupervised learning method that utilizes confrontational training, where the generator and discriminator engage in adversarial interactions to produce images. GAN-based approaches for enhancing and

generating underwater images have been extensively utilized and researched. Li et al. [36] used WaterGAN to develop a training dataset to train the generative network by using RGB-D images in the air and underwater image samples as input. Li et al. [37] introduced a general cross-domain GAN framework to generate high-quality images. Wu et al. [38] used multi-scale fusion connections to integrate the advantages of prior methods through attention feature aggregation to predict the enhancement results. Islam et al. [39] proposed an objective function to perceive image quality to supervise adversarial training. Uplavikar et al. [40] introduced adversarial loss on the encoder and used a similar generative adversarial approach to learn a variety of underwater image types to restore degraded underwater images. Li et al. [41] based on generative adversarial networks and combined with physical models to ameliorate the scene adjustability of the method. The GAN network solves the problem of dependence on paired training data sets very well, but GAN is very unstable and prone to mode collapse and convergence difficulties. The main reason is that when the generator generates images tending to one pattern, the discriminator can only learn simple rules to distinguish real images and generated images, which results in the generator being unable to generate diverse images. Additionally, due to the persistent updating of generator and discriminator parameters, the imbalance in their update speed will also lead to instability in the GAN training process.

CNN is the main branch of underwater image enhancement. Many existing methods for image processing are based on CNN. For example, Wang et al. [42] developed an intelligent protocol based on CNN and reinforcement learning, which incorporates seven methods to enhance underwater images: compensating for attenuation channel effects, adjusting white balance, applying tone mapping, fine-tuning saturation, stretching contrast, performing gamma correction, and employing high-pass fusion. A cascade of image enhancement technologies uses reinforcement learning to configure parameters to achieve image quality restoration intelligently. Qi et al. [43] and Chen et al. [44] introduced semantic information feature maps to improve image enhancement accuracy and enhance network generalization capabilities. Rezende et al. [7] combined Bayesian reasoning and CNN to derive a deep directional generation model. Introducing condition information allows data with specific constraints to be generated under given conditions. This technique finds wide application in the domain of image generation. Sun et al. [45] used convolution and deconvolution encoding and decoding to restore missing details of the image pixel by pixel. It can be seen that using CNN as a feature extractor in the codec can effectively improve the codec performance and image restoration effect. Specifically, the CNN-based codec structure encodes the input data into a low-dimensional feature map with the encoder to remove irrelevant Noise and redundant information. Subsequently, it uses the decoder to decode the feature map into output data to restore the data details and structural information. Since the encoding and decoding structures are symmetrical to each other, we can achieve effective feature extraction and reconstruction. For example, Gangisetty et al. [46] used an encoder to encode degraded images into low-dimensional feature maps. The decoder learns to restore the degraded images to a non-degraded form and uses skip connections to propagate feature maps. Qi et al. [47] proposed a collaborative network and introduced relevant feature-matching units into the codec structure to enhance underwater images. Zheng et al. [48] proposed an introspectively trained variational autoencoder to

synthesize high-definition images by introducing a conditional capsule encoder for conditional synthesis and using a classifier to introduce feature matching into the loss function to promote learning of aligned features. Taking inspiration from these approaches, we employ a CNN as a feature extractor and incorporate CVAE to comprehend the underlying data distribution and generate novel samples.

III. PROPOSED METHOD

We will provide an elaborate introduction to the specifics of DAPNet. As shown in Fig. 1, DAPNet uses a CVAE-based encoding and decoding structure in the feature extraction part. The EIB module is embedded to the encoding and decoding to capture more multi-context feature information, and the DAB module is introduced to strengthen the connection between channels and space. It helps the model better understand and utilize the structure and characteristics of the input data. Furthermore, the average and standard deviation of the features will be calculated, and PadaIN [49] will be used to complete the style feature alignment to obtain the enhanced sample image. Experiments show that the EIB and DAB can significantly increase and improve the image enhancement effect of the model.

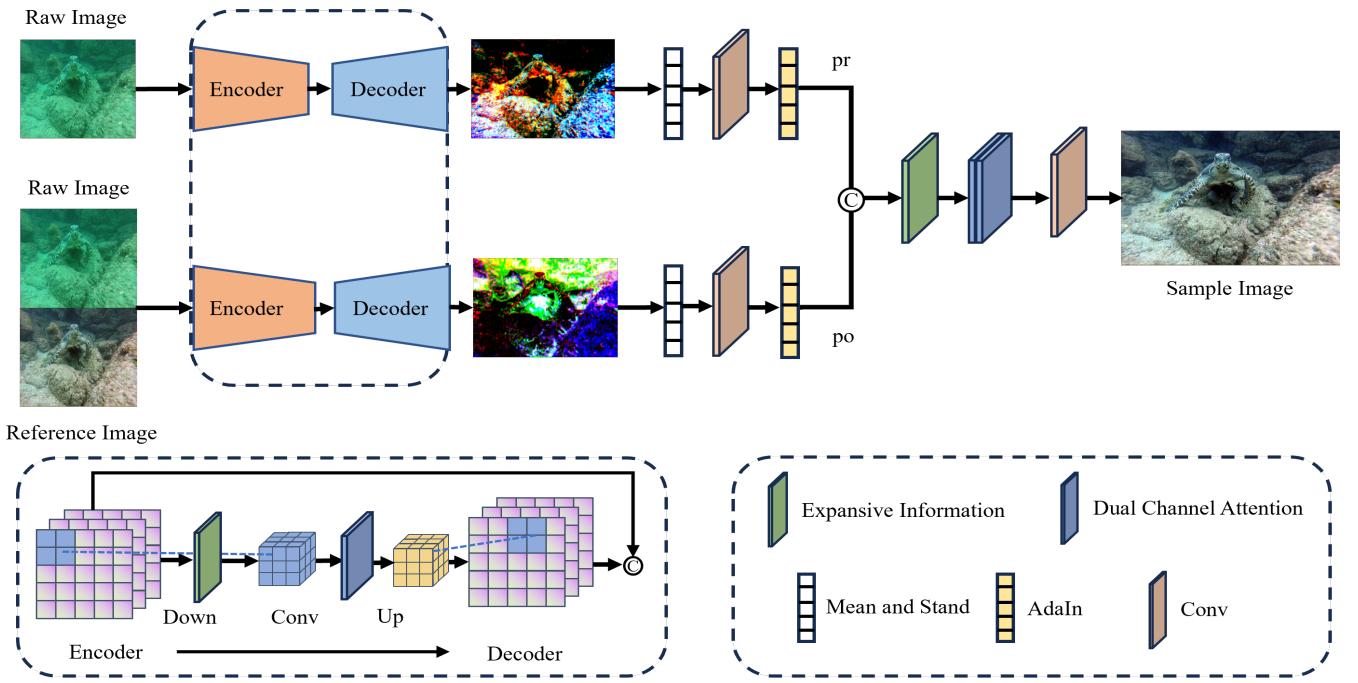


Fig. 1. The outline of our proposed approach. DAPNet adopts a dual-branch structure of prior and posterior distributions. The upper branch receives the original image as input, and the lower branch receives the original image and the labeled image as input. First, the input data is downsampled by an encoder, and the original image is encoded into the average and standard deviation of the latent space. In addition, a latent vector is sampled from this latent distribution and passed to the decoder along with the feature information of the labeled image. The decoder decodes the features of the latent vectors and labels images into reconstructed feature maps through upsampling and deconvolution layers. Next, based on the reconstructed feature map feature information, the standard deviation and mean of the Gaussian distribution are established, multiple potential vectors are sampled from the Gaussian distribution, and random samples are extracted from them. These random samples are injected into PAdaIn [49], feature alignment is performed, and sample images are generated. Finally, the final enhanced image is obtained from the sample space through the monte carlo [8] likelihood estimation method.

DAPNet has two branches. The top branch inputs the original underwater image and is used to infer the prior

distribution of a single original image, and the bottom branch inputs both the initial and the reference images and is used to construct the posterior distribution of images. To enhance the connection between the channels and spaces of features and extract more essential features, we introduce the DAB module to the network and use the output features to establish distributions for the mean and standard deviation. After that, we normalize the extracted features. Since BatchNorm [50] always performs each normalization based on the entire batch, this may cause the model to deviate when processing images with different styles and characteristics, while InstanceNorm [51] can normalize each sample independently, making style transfer more accurate and effective, we use InstanceNorm to normalize image features. Consequently, we use PAdaIn to align the style features to improve image quality. Specifically, it is assumed that the dimension of the feature map in the pr and po blocks is $N \times H \times W$, where H and W denote the vertical and horizontal dimensions of the feature map and N represents the number of channels in the feature map. First, calculate each channel's mean and standard deviation of the input feature map and obtain the mean vector and standard deviation vector with a size of $N \times 1 \times 1$. The 1×1 convolution operation can obtain the mean σ annotation difference μ of each channel from the mean vector. Additionally, use σ and μ to construct the Gaussian distribution of the N-dimensional mean M_{GN} and q and w to construct the Gaussian distribution of the N-dimensional standard deviation S_{GN} . InstanceNorm will process the prior distribution to reduce the difference between batches of samples and obtain a more accurate posterior distribution. In the testing phase, we take a single raw image as input and use random samples from the Pr block to generate predictions, which can be expressed as

$$m_{pr} \sim \mathcal{N}_1 (\mu(x), \sigma^2(x)), \quad (1)$$

$$s_{pr} \sim \mathcal{N}_2 (q(x), w^2(x)), \quad (2)$$

where x represents the input original image, m_{pr} and s_{pr} are two arbitrary samples of the mean and standard deviation distribution in the prior distribution. During the training phase, we learn the posterior distribution using the input image and the corresponding reference image. We find meaningful augmented statistical values through a posterior network and map them to two posterior distributions of mean and standard deviation, which can be expressed as

$$m_{po} \sim \mathcal{N}_1 (\mu(y, x), \sigma^2(y, x)), \quad (3)$$

$$s_{po} \sim \mathcal{N}_2 (q(y, x), w^2(y, x)), \quad (4)$$

where y represents the reference image, x represents the original image, m_{po} and s_{po} are two arbitrary samples of the mean and standard deviation distribution in the posterior distribution, N_1 is the mean of the N-dimensional

Gaussian distribution, and N_2 is the standard deviation of the N-dimensional Gaussian distribution. These two random samples will be injected into the PAdaIN module and can be expressed as

$$f(\text{PAdaIN}(x, y)) = m_{\text{po}} \frac{x - \mu x}{\sigma x} + s_{\text{po}} \times IN(pr), \quad (5)$$

where x represents the input feature map, y represents the feature map of the label image, σx and μx are the mean and standard deviation of the input feature map, subtracting the mean from the value of each channel and dividing by the standard deviation to obtain the standardized value. Subsequently, by scaling the normalized x to the mean and standard deviation of y , IN stands for InstanceNorm. Finally, we get the converted features of PAdaIn [49].

A. Expansive Information Block

The standard convolution operation requires a large amount of calculations and parameters, which partially restricts the efficiency of the model. The depth-separable convolution kernel allows the model to reduce the number of parameters while maintaining significant accuracy. In addition, we introduce the LeakyReLU activation function to increase the expressiveness of the model and avoid the problem of neuronal death. In this way, DAPNet can better handle the complex relationship between inputs and outputs while ensuring the stability of the model. On this basis, we will utilize depthwise separable convolution as the feature extraction block, design a simple and effective module, and embed the module into DAPNet. As shown in Fig. 2, we first perform a 3×3 convolution on the input and keep the image size unchanged and perform a 5×5 and 7×7 convolution on each channel of each output feature. To obtain broader information, we use a 1×1 size convolution kernel to deconvolve pixel by pixel and use the residual connection to prevent problems such as overfitting, gradient disappearance, and gradient explosion. The feature extraction process of EIB can be expressed as

$$y = x + \sum_{m=0}^{h-1} \sum_{n=0}^{w-1} x_{i+m, j+n, c} \times w_{m, n, c, k}, \quad (6)$$

where y is the output feature, h and w represent the width and height of the convolution kernel, c represents the number of input channels, x represents the $i + m$ rows and $j + n$ columns of the input feature map on the c th channel, w represents the weight value of the convolution kernel in the c^{th} input channel of m rows and n columns and the k^{th} output channel.

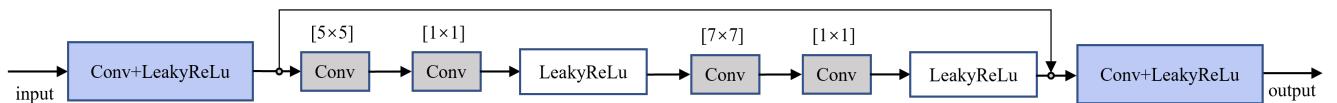


Fig. 2. The module architecture of EIB, which consists of multi-layer depth-separable convolution and leakyReLu. Initially, Features are extracted by 3×3 convolution, and then the features are passed on for further processing. Subsequently, the output and input features are spliced to achieve optimal information flow.

B. Double Attention Block

DAPNet employs an encoder-decoder structure in the feature extraction phase. The standard code structure will reduce the spatial resolution during the down-sampling and up-sampling processes, leading to the loss of image details and texture information. Therefore, we introduced the DAB module to maintain detailed texture. As shown in Fig. 3, the DAB module first passes through a leakyReLU activation function, then uses consecutive 3×3 convolutions for feature extraction, and calculates the weights of space and channels. The DAB module treats different features and pixel areas differently, which allows for processing different types of information while providing additional flexibility. We convolve the input features twice without changing the feature size and the extracted features are added with attention weights to extract significant features further. Finally, residual connections are used to fuse the input features with the features after adding attention. This fusion method can retain the original features' information while enhancing the features' expressive power and texture details.

1) *Channel Attention (CA)*: The channel attention mechanism mainly focuses on the features between channels. First, we obtain global information by using global average pooling. Subsequently, we apply global maximum pooling to extract local features from the feature map, leveraging both global and regional information to obtain a more comprehensive feature representation, which can be expressed as

$$\Lambda_c = F_{\text{avg}}(x_c) + F_{\text{max}}(x_c) = \frac{1}{h \times w} \sum_{i=1}^h \sum_{j=1}^w x_c(i, j), \quad (7)$$

where Λ_c is the feature value of the current channel, F_{avg} is the global average pooling, F_{max} is the global maximum pooling, x_c is the feature value of the c channel at the (i, j) position. After that, we use two convolutional layers and sigmoid and ReLU activation functions to calculate the channel weight. Finally, we multiply the original input feature x and the channel weight C_c to obtain the weight of each channel Feature representation F_{CA} , which can be expressed as

$$C_c = \lambda(\text{Conv}(\varepsilon(\text{Conv}(\Lambda_c)))), \quad (8)$$

$$F_{\text{CA}} = x \cdot C_c. \quad (9)$$

2) *Spatial Attention (SA)*: Spatial attention helps the model process spatial information. Specifically, we first perform mean pooling and maximum pooling on the input features, then fuse the features, and reduce their dimensionality through a convolution. Finally, we output the weight value through the sigmoid activation function, which can be expressed as

$$F_{\text{SA}} = \lambda(\text{conv}(F_{\text{mean}} \oplus F_{\text{max}})), \quad (10)$$

where F_{mean} represents the features after average pooling, F_{max} represents the features after maximum pooling. Additionally, we fuse these two features to synthesize feature information at different scales. The fusion process

is implemented before the convolution operation conv and is weighted using the weight value λ , where the weight value λ is obtained through the Sigmoid activation function. In the end, we multiply the input features with the attention coefficient obtained through the attention mechanism to achieve weighted adjustment of the input features so that the model can pay more attention to important spatial location features in the image.

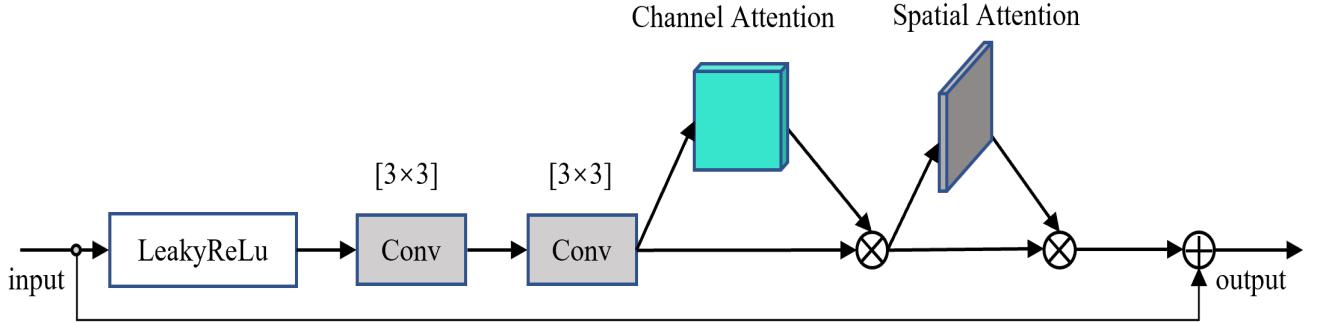


Fig. 3. The module structure of DAB, which consists of multi-layer convolution, channel attention, and spatial attention. Originally, we used two 3×3 convolution kernels to extract detailed features and used a weighted fusion channel and spatial attention mechanism to adjust the weight of the features. Subsequently, the model's ability to express detailed features is enhanced by fusing input features and output features.

C. Loss Function

To make the prior distribution and posterior distribution as consistent as possible, DAPNet is trained by minimizing the variational lower bound. By reconstructing the loss, we penalize the difference between the DAPNet output and the reference image. Our reconstructed loss L_r is expressed as

$$L_r = L_{\text{huber}} + \lambda L_{\text{vgg16}}, \quad (11)$$

where L_{huber} [52] is a smoothing loss that reduces the sensitivity of the mean square error to discrete points to a certain extent while retaining the advantages of the absolute error when the data is imbalanced. λ represents the weight and L_{vgg16} [53] is the perceptual loss. We dynamically adjust hyperparameters to further improve model robustness, which can be expressed as

$$L_\delta(y, f(x)) = \begin{cases} \frac{1}{2}(y - f(x))^2, & \text{if } |y - f(x)| \leq \delta \\ \delta|y - f(x)| - \frac{1}{2}\delta^2, & \text{if } |y - f(x)| > \delta \end{cases}, \quad (12)$$

where δ is a hyperparameter used to control the balance between squared error and absolute error. When the δ prediction value is greater than the squared error. Otherwise, the absolute error is used. y represents the reference image and $f(x)$ is the sample generated by the generative model. We will sum the learning rate, and δ is connected, using a considerable δ value at the beginning of training and gradually reducing the δ value with continuous iteration and the reduction of the learning rate, making DAPNet pay more attention to minor errors. In addition, to ensure the consistency of the prior distribution and the posterior distribution, Kullback-Leibler divergence is used to make the prior and posterior distribution consistent, which can be expressed as

$$L_k = D_{\text{KL}}(\mathcal{N}_1(x) \parallel \mathcal{N}_1(y, x)) + D_{\text{KL}}(\mathcal{N}_2(x) \parallel \mathcal{N}_2(y, x)), \quad (13)$$

where D_{KL} is the degree of dispersion between the two distributions, the final loss function is the weighted sum of the above losses, which is expressed as

$$L = L_r + \beta L_k, \quad (14)$$

where β is the weight. During the testing phase, we will let the network predict n enhancement variables based on the number of samples n (default is 25). Additionally, we will average possible samples through monte carlo [8] likelihood estimation to obtain the final prediction result. Some sample images were generated, as shown in Fig. 4.



Fig. 4. The first image from left to right is the raw image, and the rest are sample images.

IV. EXPERIMENT AND ANALYSIS

We comprehensively assess our designed DAPNet using multiple benchmark datasets and evaluation metrics. First, we provide an overview of the dataset and evaluation metrics to ensure the accuracy and comprehensiveness of the evaluation. Furthermore, we conduct qualitative and quantitative comparisons with other methods using DAPNet. Next, ablation analysis is performed to prove the effect of our designed module. Finally, we conduct application testing to substantiate the effect of DAPNet on high-level computer vision missions.

Comparision Methods. We compare DAPNet with ten UIE methods, including two model-free methods (Retinex [54] and Fusion [55]), four grounded on prior experimental methods (IBLA [56], HistogramPrior [57], MLLE [58], ACDC [18]), and four deep learning-based methods (Water-Net [13], CLUIE-Net [59], LA-Net [60], U-shape [61]). We execute the author's provided source code with recommended parameters to obtain optimal outcomes from different methods.

Benchmark Datasets. We prove the performance of our method by testing it on three datasets: UIEB [13], UCCS [62], and UIQS [62]. The UIEB dataset comprises 950 genuine underwater images, and UCCS serves as an evaluation metric for assessing the color correction capabilities of the methods applied to underwater images. It includes three subsets of 100 images in blue, green, and teal. The UIQS dataset consists of 3630 images, showcasing a wide range of underwater scenes and presenting varying degrees of image quality. The established UIEB data set is used as a training set. This data set is trained with the first 700 raw images and matching label images, and each original underwater image has multiple reference images. The remaining datasets are utilized for testing.

Evaluation Metrics. We used five metrics to rate the quality of underwater images, including peak signal-to-noise ratio PSNR [63], structural similarity SSIM [64], color accuracy DeltaE [65] (adopting the CIE2000 standard), UIQM [66] and natural image quality evaluation NIQE [67]. In the UIEB [13] dataset, SSIM, PSNR,

and DeltaE are calculated based on the label images to ensure a fair comparison with existing methods. A higher PSNR score indicates that the resulting image is closer to the actual image and more similar to the basic real image in structure and texture. DeltaE is used to quantify the difference between two colors, with lower DeltaE values indicating a lower color difference and higher values indicating more pronounced color differences. UIQM metric is used to assess the performance of reference-free predicted images. It encompasses three metrics for underwater image attributes: underwater image color metric (UICM), underwater image sharpness metric (UISM), and underwater image contrast metric (UIConM). A higher UIQM score indicates better image quality. NIQE is also a reference-free measurement that does not require a reference image. Lower NIQE scores are indicative of superior human visual perception. Since the UIQS and UCCS datasets only contain original images, we employ NIQE and UIQM as evaluation metrics to assess the model's performance. The use of these metrics helps to assess the effectiveness of methods and provides comparison and measurement of performance differences between different approaches. By considering the results of these metrics together, we can gain a more comprehensive understanding of the model's performance and reliability in underwater image enhancement.

Implementation Details. Our proposed DAPNet is employed using Pytorch and trained on NVIDIA Tesla V100S PCIE. Adam is used as the optimization algorithm, the mini-batch size is 4, and the learning rate is 1×10^{-4} . We calculated the average standard deviation of PSNR and SSIM for each image to confirm the optimal number of monte carlo [8] likelihood estimation samples. We executed the model five times and obtained the average value for a more robust evaluation. According to the information provided in Table I, it can be observed that the standard deviation of PSNR and SSIM tends to be stable as the number of sampling iterations increases. This finding confirms that DAPNet enhances image diversity. To achieve a better balance between the stability and computational efficiency of DAPNet, this study sets the default number of sampling iterations for MC estimation to 25.

TABLE I
THE EFFECT OF THE NUMBER OF SAMPLES.

Sampling Times	1	5	10	15	20	25	35
PSNR (mean stand deviation)	0.842	0.573	0.340	0.196	0.178	0.164	0.160
SSIM (mean stand deviation)	0.008	0.004	0.004	0.003	0.003	0.002	0.002

A. Color Correction Comparisons on the UCCS Dataset

Qualitative Evaluations: We conducted a comparative analysis of diverse methodologies on the UCCS [62] dataset to assess their color correction performance. According to the comparison in Fig. 5, Retinex [54] and Histogram-Prior [57] methods introduced red color correction, while Fusion [55] failed to correct blue underwater images. MLLE [58] and ACDC [18] effectively remove the blue bias but suffer from local over-enhancement issues. WaterNet [13] successfully removes the blue bias but introduces light yellow tones and

local darkness in the image. IBLA [56] fails to correct the green color, while CLUIENet [59] removes the blue color cast but introduces local yellow tones and has limited green correction capability. LANet [60] and U-shape [61] methods eliminate the blue color but suffer from local over-enhancement problems. In comparison, our method outpaces the contrastive methodologies significantly in contrast enhancement and preserving image details.

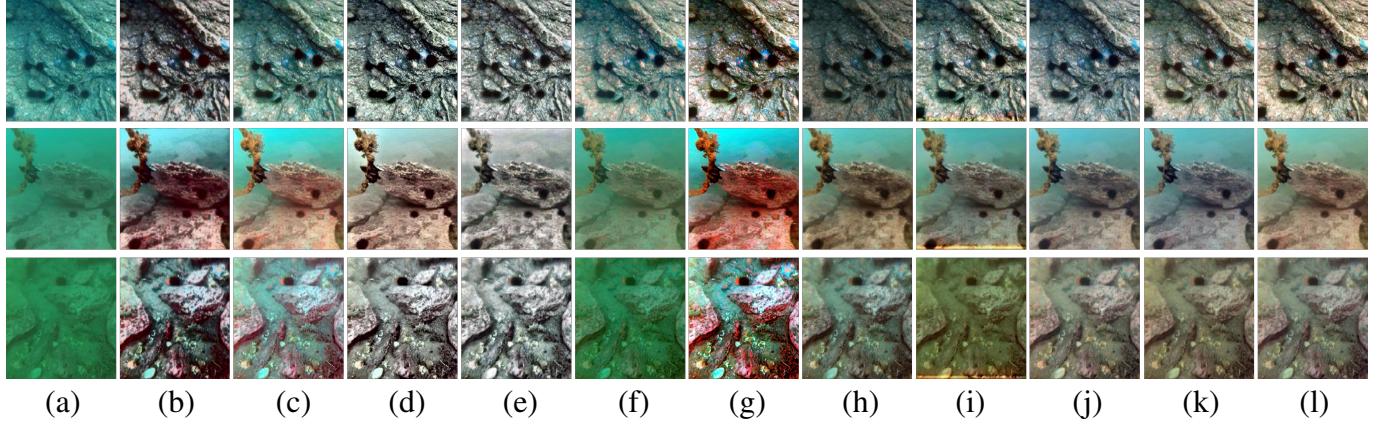


Fig. 5. Comparison of visualization results of UCCS dataset. (a) Raw, (b) Retinex [54], (c) Fusion [55], (d) MLLE [58], (e) ACDC [18], (f) IBLA [56], (g) Histogram-Prior [57], (h) WaterNet[13], (i) CLUIENet [59], (j) LANet [60], (k) U-shape [61], and (l) DAPNet.

B. Visibility Comparisons on the UIQS Dataset

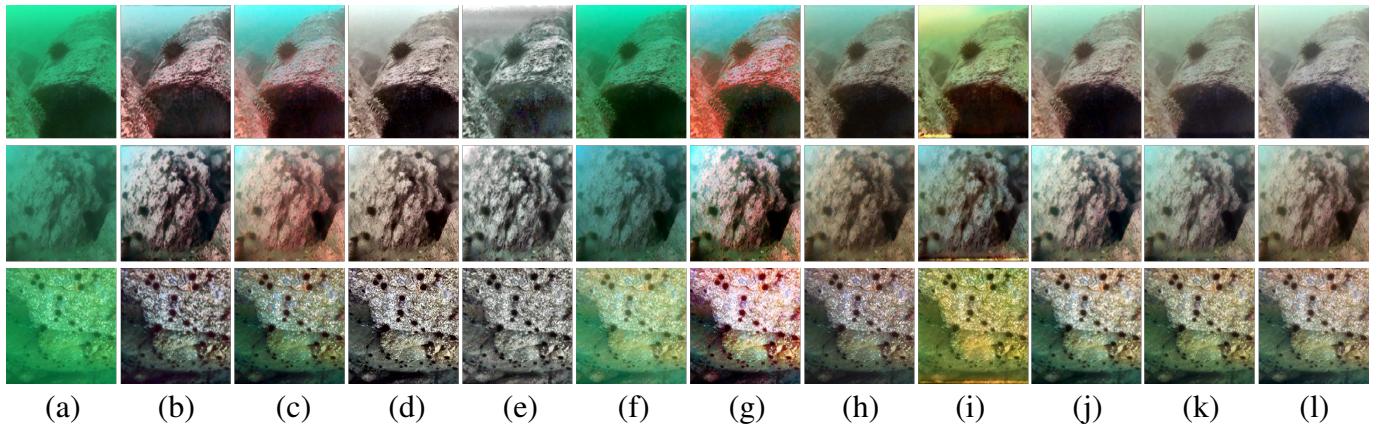


Fig. 6. Comparison of visualization results of UIQS dataset. (a) Raw, (b) Retinex [54], (c) Fusion [55], (d) MLLE [58], (e) ACDC [18], (f) IBLA [56], (g) Histogram-Prior [57], (h) WaterNet[13], (i) CLUIENet [59], (j) LANet [60], (k) U-shape [61], and (l) DAPNet.

Qualitative Evaluation. We evaluated the impact of various methods on the UIQS [62] dataset regarding visibility enhancement. The results of enhancement using different methods are shown in Fig. 6. Considering the confined space, we only exhibit some of the different experimental results. We further conduct a qualitative comparison on the UIQS dataset and rate the enhanced effectiveness of approaches to processing diverse degraded underwater images. From Fig. 6, we can see that Retinex [54], Fusion [55], and Histogram-Prior [57] introduce red while maximizing deep green. IBLA [56] is not ideal for dealing with color casts. MLLE [58]

improves the image's brightness and contrast, but the effect on detail enhancement is not ideal. CLUIENet [59] cannot eliminate severe color casts well, and there is partial under-enhancement. WaterNet [13], LANet [60], and U-shape [61] have a good effect on removing color deviation. However, WaterNet introduces a light yellow color and has an unsatisfactory effect on image contrast and brightness enhancement. LANet and U-shape improve brightness and saturation but on details. It's not ideal for sharpening. In contradistinction, our method productively eliminates color casts, sharpens image details and textures, and does not cause over-enhancement or under-enhancement.

C. Comprehensive Comparisons on the UIEB Dataset

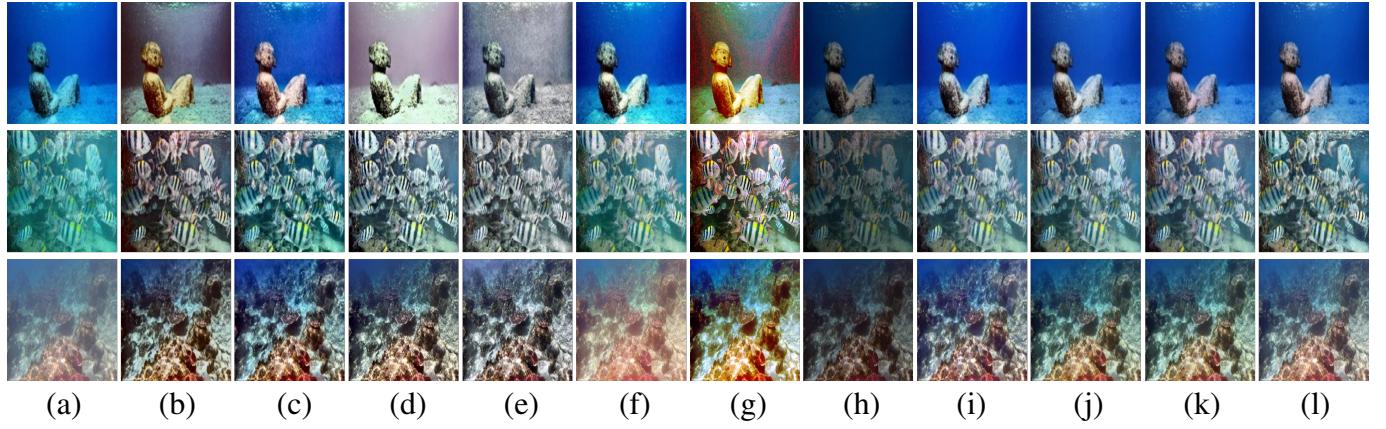


Fig. 7. Qualitative comparisons on UIEB. Comparison of visualization results of UIEB dataset. (a) Raw, (b) Retinex [54], (c) Fusion [55], (d) MLLE [58], (e) ACDC [18], (f) IBLA [56], (g) Histogram-Prior [57], (h) WaterNet[13], (i) CLUIENet [59], (j) LANet [60], (k) U-shape [61], and (l) DAPNet.

Qualitative Evaluation. We selected a set of representative underwater images from the UIEB [13] dataset and categorized them into three groups based on their visual appearance: bluish images, greenish images, and hazy images. Fig. 7 illustrates the enhancement results obtained using various methods. Due to space limitations, we only offer some representative results of different methods. We can see from Fig. 7 that ACDC [18], LANET [56], and U-shape [61] introduce local color distortion, IBLA cannot completely remove the fog effect, and WaterNet [13] introduces darkness. Histogram-Prior [57] introduces unnecessary color distortion. Retinex [54] and Fusion [55] restored hazy images to a certain extent, but on blue and cyan images, Retinex introduced yellow and reddish colors, Fusion did not completely remove cyan, and MLLE [58] and ACDC [18] could not handle blue and cyan images well. MLLE over-increases the image contrast and fails to restore image details. LANet [60] and U-shape effectively restore the visualization effect, introducing an unnatural reddish color to local colors. By way of contrast, our method effectively eliminates unnatural colors without apparent over-enhancement and under-enhancement, significantly improving the visibility of the image.

D. Quantitative Comparisons

We conducted quantitative comparisons on three datasets, and the results are presented in Table II. In the UIEB [13] dataset, we evaluated our method using metrics such as PSNR [63], SSIM [64], DelateE [65], NIQE

TABLE II

QUANTITATIVE COMPARISON OF UIEB [13], UIQS [62], AND UCCS [62] DATASETS. THE BOLD SCORES DENOTE THE BEST RESULTS UNDER EACH CASE.

Methods	UIEB					UIQS		UCCS	
	PSNR ↑	SSIM ↑	DeltaE ↓	NIQE ↓	UIQM ↑	NIQE ↓	UIQM ↑	NIQE ↓	UIQM ↑
Retinex [54]	17.863	0.701	15.13	9.074	1.847	8.234	2.268	10.357	2.217
Fusion [55]	20.870	0.688	11.62	6.747	2.458	8.238	2.405	10.865	2.157
IBLA [56]	19.515	0.740	17.53	8.290	2.395	8.507	2.153	14.516	2.195
Histogram-Prior [57]	15.212	0.605	18.28	6.780	2.157	9.086	2.552	10.119	2.235
MLLE [58]	17.894	0.751	19.93	8.750	2.368	11.276	2.375	11.013	1.903
ACDC [18]	17.217	0.752	19.77	12.701	2.154	13.454	1.862	14.023	2.034
WaterNet [13]	19.158	0.827	9.87	6.879	1.685	10.107	2.865	10.252	3.018
CLUIENet [59]	21.035	0.837	14.15	8.126	3.054	9.471	2.498	9.347	2.657
LANet [60]	20.674	0.660	9.32	6.133	2.865	8.173	2.972	10.031	2.925
U-shape [61]	20.810	0.869	13.74	7.148	2.912	8.384	2.914	8.274	3.249
DAPNet	22.065	0.871	10.68	5.803	2.964	8.032	2.953	8.019	2.864

[67], and UIQM [66]. Our method achieved the best or nearly the best scores in the quantitative assessments. The qualitative and quantitative comparisons on the UIEB dataset, as shown in Table II and Fig. 7, demonstrate our method’s superior texture details, brightness, and enhanced visibility capabilities. Without reference images for the UIQS [62] and UCCS [62] datasets, we assessed our method using NIQE and UIQM metrics. From the quantitative comparisons in Table II, Fig. 6, and Fig. 5, our method achieved the highest or nearly the highest scores in the quantitative evaluations. Furthermore, we observed that our method outperforms the comparison methods to advance the quality of underwater images with varying degradation levels and different color characteristics. Overall, these findings indicate that our method exhibits excellent color correction capabilities. In general, the quantitative comparison demonstrates that our method can effectively enhance image contrast, brightness, and detailed texture across various underwater scenes. Furthermore, our approach exhibits strong color correction capabilities.

E. Ablation Study

To showcase the efficiency of our module., we execute ablation studies on the UIEB [13] dataset. Specifically, we removed the extended information extraction block (-W/O EIB) and dual attention module (-W/O DAB). The experimental outcomes reveals that the whole model received the highest score and also proved the positive impact of each core module on DAPNet. Judging from the results in Fig. 8 and Table III, both quantitative evaluation and visual comparison clearly demonstrate the advantages of the complete model. Specifically, we observed the following: For the -w/o EIB case, the model is not accurate enough in handling details. This means that the model cannot accurately capture subtle changes and details in the image. This can lead to a lack of clarity and detail in the processed image. For the -w/o DAB case, the model cannot capture the

correlation and importance between different positions and spaces in the image. We can also find from Fig. 8 that the model cannot correct the color cast well and introduces darkness. There is an obvious loss of detail and structural information, and unbalanced colors and contrasts appear.

So, we can see that the complete model shows advantages in detail processing, local detail enhancement, and global color consistency. It more accurately captures and enhances image details while maintaining overall color consistency, resulting in more natural and realistic image effects. This illustrates the role and importance of each constraint and component in the algorithm, providing strong support for the algorithm's overall performance.

TABLE III

ABLATION STUDIES OF DIFFERENT MODULES TESTED ON THE UIEB [13] DATASET. THE BOLD SCORES DENOTE THE BEST RESULTS UNDER EACH CASE.

Ablated models	PSNR \uparrow	SSIM \uparrow	DeltaE \downarrow	NIQE \downarrow	UIQM \uparrow
-w/o EIB	15.9347	0.794	14.38	9.144	2.237
-w/o DAB	17.554	0.8131	15.45	9.916	2.304
DAPNet (full model)	22.065	0.871	10.68	5.803	2.964

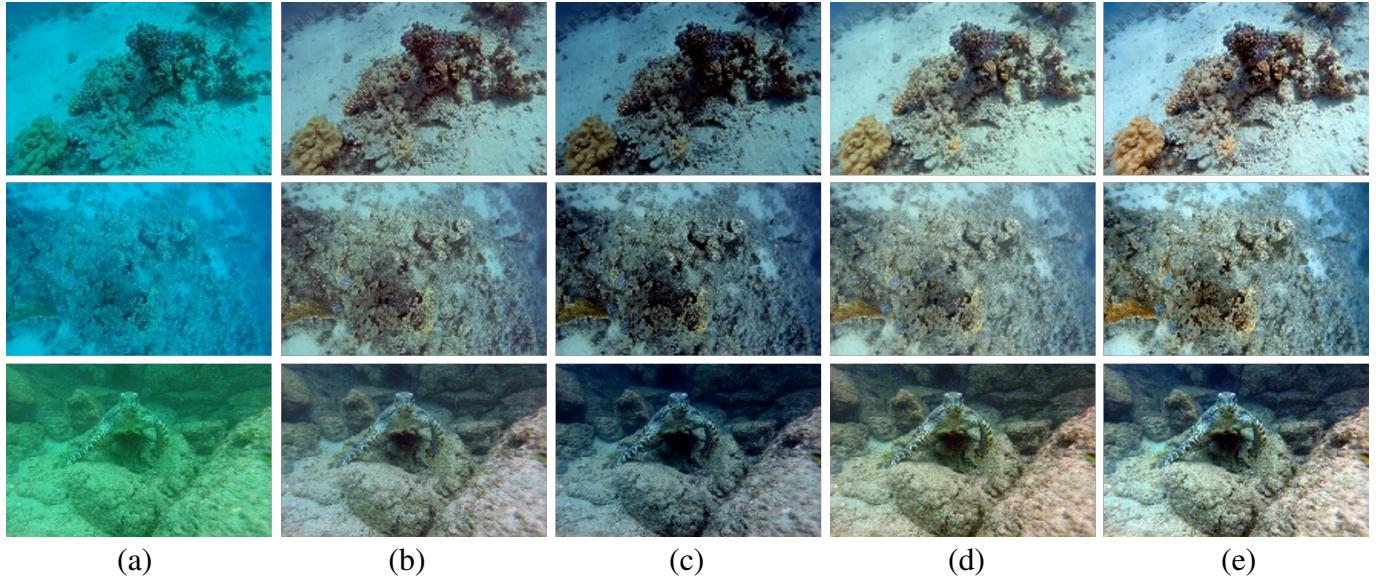


Fig. 8. The ablated results of our method. (a) Raw. (c) -w/o EIB. (b) -w/o DAB. (d) DAPNet (full model). (e) reference.

F. Applications

To further verify the application effect of this algorithm in image processing, we applied key point detection, saliency detection, and underwater image segmentation.

Keypoint Detection: The key point detection method intends to calculate the number of key points in the image. We used the SIFT operator [68] to calculate the number of feature points in the image. As presented in Fig. 9, the number of feature points in the original images in the first and second rows is 110 and 498. After being enhanced by the algorithm of this article, the number of feature points increased to 272 and 952. It

can be clearly seen that the number of feature points increases significantly after being optimized by DAPNet. Further application testing showed that the images processed by DAPNet showed good application effects in the subsequent feature extraction process.

Saliency Detection: Underwater image saliency detection focuses on identifying the most visually appealing regions within an image. To accomplish this, we utilize the graph-based manifold ranking method (GBMR) [69] to observe saliency in both the original underwater images and the optimized underwater images. As shown in Fig. 9, the salient objects in the underwater image enhanced by our method exhibit a more accurate and comprehensive structure.

Color Segmentation: We divide the image into several disjoint uniform regions based on the specific properties of some image regions. To achieve this goal, we employ the superpixel-based Fast Fuzzy C-Means (FCM) method to cluster two raw underwater images and enhance the corresponding images of different regions by our method. As shown in Fig. 9, when FCM is applied in combination with our method, the segmentation results show higher consistency with the enhanced underwater images.

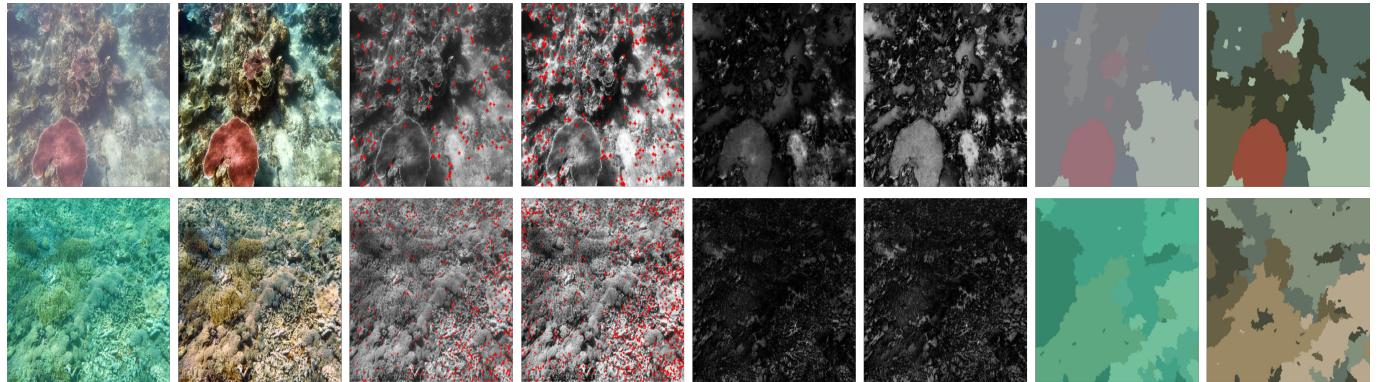


Fig. 9. Apply test results. Keypoint detection, saliency detection, and color segmentation for underwater image segmentation. From left to right are the original image, the image enhanced by our method, and the corresponding application results of the original input and enhanced results.

G. Generalization Performance of Our Method



Fig. 10. Application results for blurry and low-light images. (a) The upper row is the blurred image, and the lower row is the result after enhancement by DAPNet. (b) The upper row is a low-light image, and the lower row is the enhanced result of DAPNet.

We further try to apply it to other types of images that are degraded. Through exploratory experiments, we obtained some encouraging results. DAPNet is able to achieve better enhancement effects, especially for hazy images and low-light images. For haze images, the main problem is that it affects the contrast and color of the picture. As shown in Fig. 10(a), DAPNet can substantially elevate the contrast and color performance of haze images, making the images clearer and more vivid. By reducing the impact of haze, DAPNet effectively restores image details and colors, improving the quality of visual perception. In low-light conditions, the brightness and detail of the image are often severely affected. However, by applying our algorithm, as shown in Fig. 10(b), we are able to improve the brightness and detail presentation of low-light images significantly. DAPNet can enhance weak signals in images and restore details obscured by low light, making images more clearly visible.

In summary, DAPNet can not only productively elevate the quality of underwater images but also achieve satisfactory enhancement effects when processing hazy images and low-light images. This means that DAPNet has broader application potential and can cope with many different types of image degradation problems.

V. CONCLUSION

We have presented a robust and generalizable for enhancing underwater images. We introduce expanded information blocks in the encoder, which effectually address the problem of possible information loss in the encoder's downsampling process and can capture more context. We also introduce dual attention modules to enhance the network's perception of channels and spaces, enriching potential Sample space. Our approach has been extensively evaluated on diverse benchmarks, showcasing its effectiveness and robustness. Our method has good generalization ability to low light and fog images. Although our method has a good effect in enhancing the details and colors of underwater images, there is a lot of room for improvement in operating efficiency. This challenging case is an aspect that we aim to address in our future endeavors.

REFERENCES

- [1] J. Zhou, Q. Liu, Q. Jiang, W. Ren, K.-M. Lam, and W. Zhang, "Underwater camera: Improving visual perception via adaptive dark pixel prior and color correction," *International Journal of Computer Vision*, pp. 1–19, 2023.
- [2] Y. Kang, Q. Jiang, C. Li, W. Ren, H. Liu, and P. Wang, "A perception-aware decomposition and fusion framework for underwater image enhancement," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 3, pp. 988–1002, 2022.
- [3] G. Hou, N. Li, P. Zhuang, K. Li, H. Sun, and C. Li, "Non-uniform illumination underwater image restoration via illumination channel sparsity prior," *IEEE Transactions on Circuits and Systems for Video Technology*, 2023.
- [4] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," pp. 2223–2232, 2017.
- [5] H. Li, J. Li, and W. Wang, "A fusion adversarial underwater image enhancement network with a public test dataset," *arXiv preprint arXiv:1906.06819*, 2019.
- [6] Y. Guo, H. Li, and P. Zhuang, "Underwater image enhancement using a multiscale dense generative adversarial network," *IEEE Journal of Oceanic Engineering*, vol. 45, no. 3, pp. 862–870, 2019.
- [7] D. J. Rezende, S. Mohamed, and D. Wierstra, "Stochastic backpropagation and approximate inference in deep generative models," pp. 1278–1286, 2014.
- [8] K. Sohn, H. Lee, and X. Yan, "Learning structured output representation using deep conditional generative models," *Advances in neural information processing systems*, vol. 28, 2015.
- [9] S. K. Dhara, M. Roy, D. Sen, and P. K. Biswas, "Color cast dependent image dehazing via adaptive airlight refinement and non-linear color balancing," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 5, pp. 2076–2081, 2020.

- [10] J. Zhou, D. Liu, X. Xie, and W. Zhang, "Underwater image restoration by red channel compensation and underwater median dark channel prior," *Applied Optics*, vol. 61, no. 10, pp. 2915–2922, 2022.
- [11] J. Jackson, S. Kun, K. O. Agyekum, A. Oluwasanmi, and P. Suwansrikham, "A fast single-image dehazing algorithm based on dark channel prior and rayleigh scattering," *IEEE Access*, vol. 8, pp. 73 330–73 339, 2020.
- [12] L. Montier, S. Plaszczynski, F. Levrier, M. Tristram, D. Alina, I. Ristorcelli, J.-P. Bernard, and V. Guillet, "Polarization measurement analysis-ii. best estimators of polarization fraction and angle," *Astronomy & Astrophysics*, vol. 574, p. A136, 2015.
- [13] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao, "An underwater image enhancement benchmark dataset and beyond," *IEEE Transactions on Image Processing*, vol. 29, pp. 4376–4389, 2019.
- [14] M. Miao and S. Wang, "Pa-colornet: progressive attention network based on rgb and hsv color spaces to improve the visual quality of underwater images," *Signal, Image and Video Processing*, pp. 1–9, 2023.
- [15] S. Gangisetty and R. R. Rai, "Floodnet: Underwater image restoration based on residual dense learning," *Signal Processing: Image Communication*, vol. 104, p. 116647, 2022.
- [16] Y. Wang, J. Guo, H. Gao, and H. Yue, "Uiec²-net: Cnn-based underwater image enhancement using two color space," *Signal Processing: Image Communication*, vol. 96, p. 116250, 2021.
- [17] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and P. Bekaert, "Color balance and fusion for underwater image enhancement," *IEEE Transactions on image processing*, vol. 27, no. 1, pp. 379–393, 2017.
- [18] W. Zhang, Y. Wang, and C. Li, "Underwater image enhancement by attenuated color channel correction and detail preserved contrast enhancement," *IEEE Journal of Oceanic Engineering*, vol. 47, no. 3, pp. 718–735, 2022.
- [19] W. Zhang, S. Jin, P. Zhuang, Z. Liang, and C. Li, "Underwater image enhancement via piecewise color correction and dual prior optimized contrast enhancement," *IEEE Signal Processing Letters*, vol. 30, pp. 229–233, 2023.
- [20] Z. Liang, W. Zhang, R. Ruan, P. Zhuang, X. Xie, and C. Li, "Underwater image quality improvement via color, detail, and contrast restoration," *IEEE Transactions on Circuits and Systems for Video Technology*, 2023.
- [21] W. Zhang, L. Dong, and W. Xu, "Retinex-inspired color correction and detail preserved fusion for underwater image enhancement," *Computers and Electronics in Agriculture*, vol. 192, p. 106585, 2022.
- [22] P. Zhuang and X. Ding, "Underwater image enhancement using an edge-preserving filtering retinex algorithm," *Multimedia Tools and Applications*, vol. 79, pp. 17 257–17 277, 2020.
- [23] J. Zhou, J. Yao, W. Zhang, and D. Zhang, "Multi-scale retinex-based adaptive gray-scale transformation method for underwater image enhancement," *Multimedia Tools and Applications*, pp. 1–21, 2022.
- [24] S. Zhang, T. Wang, J. Dong, and H. Yu, "Underwater image enhancement via extended multi-scale retinex," *Neurocomputing*, vol. 245, pp. 1–9, 2017.
- [25] P. Zhuang, C. Li, and J. Wu, "Bayesian retinex underwater image enhancement," *Engineering Applications of Artificial Intelligence*, vol. 101, p. 104171, 2021.
- [26] A. S. A. Ghani and N. A. M. Isa, "Underwater image quality enhancement through integrated color model with rayleigh distribution," *Applied soft computing*, vol. 27, pp. 219–230, 2015.
- [27] A. S. A. Ghani, "Image contrast enhancement using an integration of recursive-overlapped contrast limited adaptive histogram specification and dual-image wavelet fusion for the high visibility of deep underwater image," *Ocean Engineering*, vol. 162, pp. 224–238, 2018.
- [28] L. Zhou, Q. Liu, Y. Fan, X. Song, X. Pan, and W. Zhang, "Underwater image enhancement via complementary advantage fusion of global and local contrast," *Computers and Electrical Engineering*, vol. 112, p. 108990, 2023.
- [29] W. Zhang, L. Zhou, P. Zhuang, G. Li, X. Pan, W. Zhao, and C. Li, "Underwater image enhancement via weighted wavelet visual perception fusion," *IEEE Transactions on Circuits and Systems for Video Technology*, 2023.
- [30] R. Liu, Z. Jiang, S. Yang, and X. Fan, "Twin adversarial contrastive learning for underwater image enhancement and beyond," *IEEE Transactions on Image Processing*, vol. 31, pp. 4922–4936, 2022.
- [31] P. Zhuang, J. Wu, F. Porikli, and C. Li, "Underwater image enhancement with hyper-laplacian reflectance priors," *IEEE Transactions on Image Processing*, vol. 31, pp. 5442–5455, 2022.
- [32] W. Song, Y. Wang, D. Huang, A. Liotta, and C. Perra, "Enhancement of underwater images with statistical model of background light and optimization of transmission map," *IEEE Transactions on Broadcasting*, vol. 66, no. 1, pp. 153–169, 2020.
- [33] Z. Liang, W. Zhang, R. Ruan, P. Zhuang, and C. Li, "Gifm: An image restoration method with generalized image formation model for poor visible conditions," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–16, 2022.

- [34] Y.-T. Peng and P. C. Cosman, "Underwater image restoration based on image blurriness and light absorption," *IEEE transactions on image processing*, vol. 26, no. 4, pp. 1579–1594, 2017.
- [35] J. Zhou, Y. Wang, C. Li, and W. Zhang, "Multicolor light attenuation modeling for underwater image restoration," *IEEE Journal of Oceanic Engineering*, 2023.
- [36] J. Li, K. A. Skinner, R. M. Eustice, and M. Johnson-Roberson, "Watergan: Unsupervised generative network to enable real-time color correction of monocular underwater images," *IEEE Robotics and Automation letters*, vol. 3, no. 1, pp. 387–394, 2017.
- [37] F. Li, J. Zheng, Y.-f. Zhang, W. Jia, Q. Wei, and X. He, "Cross-domain learning for underwater image enhancement," *Signal Processing: Image Communication*, vol. 110, p. 116890, 2023.
- [38] J. Wu, X. Liu, Q. Lu, Z. Lin, N. Qin, and Q. Shi, "Fw-gan: Underwater image enhancement using generative adversarial network with multi-scale fusion," *Signal Processing: Image Communication*, vol. 109, p. 116855, 2022.
- [39] M. J. Islam, Y. Xia, and J. Sattar, "Fast underwater image enhancement for improved visual perception," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 3227–3234, 2020.
- [40] P. M. Uplavikar, Z. Wu, and Z. Wang, "All-in-one underwater image enhancement using domain-adversarial learning," in *CVPR workshops*, 2019, pp. 1–8.
- [41] R. Cong, W. Yang, W. Zhang, C. Li, C.-L. Guo, Q. Huang, and S. Kwong, "Pugan: Physical model-guided underwater image enhancement using gan with dual-discriminators," *IEEE Transactions on Image Processing*, 2023.
- [42] H. Wang, S. Sun, and P. Ren, "Meta underwater camera: A smart protocol for underwater image enhancement," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 195, pp. 462–481, 2023.
- [43] Q. Qi, K. Li, H. Zheng, X. Gao, G. Hou, and K. Sun, "Sguie-net: Semantic attention guided underwater image enhancement with multi-scale perception," *IEEE Transactions on Image Processing*, vol. 31, pp. 6816–6830, 2022.
- [44] L. Chen, F. Zhou, S. Wang, J. Dong, N. Li, H. Ma, X. Wang, and H. Zhou, "Swipenet: Object detection in noisy underwater scenes," *Pattern Recognition*, vol. 132, p. 108926, 2022.
- [45] X. Sun, L. Liu, Q. Li, J. Dong, E. Lima, and R. Yin, "Deep pixel-to-pixel network for underwater image enhancement and restoration," *IET Image Processing*, vol. 13, no. 3, pp. 469–474, 2019.
- [46] S. Gangisetty and R. R. Rai, "Underwater image restoration using deep encoder–decoder network with symmetric skip connections," *Signal, Image and Video Processing*, pp. 1–9, 2022.
- [47] Q. Qi, Y. Zhang, F. Tian, Q. J. Wu, K. Li, X. Luan, and D. Song, "Underwater image co-enhancement with correlation feature matching and joint learning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 3, pp. 1133–1147, 2021.
- [48] K. Zheng, Y. Cheng, X. Kang, H. Yao, and T. Tian, "Conditional introspective variational autoencoder for image synthesis," *IEEE Access*, vol. 8, pp. 153 905–153 913, 2020.
- [49] Z. Fu, W. Wang, Y. Huang, X. Ding, and K.-K. Ma, "Uncertainty inspired underwater image enhancement," in *European Conference on Computer Vision*. Springer, 2022, pp. 465–482.
- [50] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," pp. 448–456, 2015.
- [51] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," *arXiv preprint arXiv:1607.08022*, 2016.
- [52] K. Gokcesu and H. Gokcesu, "Generalized huber loss for robust learning and its efficient minimization for a robust statistics," *arXiv preprint arXiv:2108.12627*, 2021.
- [53] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," pp. 694–711, 2016.
- [54] X. Fu, P. Zhuang, Y. Huang, Y. Liao, X.-P. Zhang, and X. Ding, "A retinex-based enhancing approach for single underwater image," pp. 4572–4576, 2014.
- [55] C. Ancuti, C. O. Ancuti, T. Haber, and P. Bekaert, "Enhancing underwater images and videos by fusion," pp. 81–88, 2012.
- [56] B. P. Hanmantte and M. Ingle, "Underwater image restoration based on light absorption," pp. 1–4, 2018.
- [57] C.-Y. Li, J.-C. Guo, R.-M. Cong, Y.-W. Pang, and B. Wang, "Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior," *IEEE Transactions on Image Processing*, vol. 25, no. 12, pp. 5664–5677, 2016.
- [58] W. Zhang, P. Zhuang, H.-H. Sun, G. Li, S. Kwong, and C. Li, "Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement," *IEEE Transactions on Image Processing*, vol. 31, pp. 3997–4010, 2022.
- [59] K. Li, L. Wu, Q. Qi, W. Liu, X. Gao, L. Zhou, and D. Song, "Beyond single reference for training: underwater image enhancement via comparative learning," *IEEE Transactions on Circuits and Systems for Video Technology*, 2022.

- [60] S. Liu, H. Fan, S. Lin, Q. Wang, N. Ding, and Y. Tang, "Adaptive learning attention network for underwater image enhancement," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 5326–5333, 2022.
- [61] L. Peng, C. Zhu, and L. Bian, "U-shape transformer for underwater image enhancement," *IEEE Transactions on Image Processing*, 2023.
- [62] R. Liu, X. Fan, M. Zhu, M. Hou, and Z. Luo, "Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light," *IEEE transactions on circuits and systems for video technology*, vol. 30, no. 12, pp. 4861–4875, 2020.
- [63] I. Avcibas, B. Sankur, and K. Sayood, "Statistical evaluation of image quality measures," 2001.
- [64] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [65] G. Sharma, W. Wu, and E. N. Dalal, "The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations," *Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur*, vol. 30, no. 1, pp. 21–30, 2005.
- [66] K. Panetta, C. Gao, and S. Agaian, "Human-visual-system-inspired underwater image quality measures," *IEEE Journal of Oceanic Engineering*, vol. 41, no. 3, pp. 541–551, 2015.
- [67] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a completely blind image quality analyzer," *IEEE Signal processing letters*, vol. 20, no. 3, pp. 209–212, 2012.
- [68] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, pp. 91–110, 2004.
- [69] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," pp. 3166–3173, 2013.