Underwater motion deblurring based on cascaded attention mechanism

Tengyue Li,[1,2] Shenghui Rong,[1,*] Long Chen,[2] Huiyu Zhou,[2] Bo He[1], Member, IEEE
[1]Ocean University of China, School of Information Science and Engineering, Qingdao 266100, China
[2]University of Leicester, School of Computing and Mathematical Sciences, Leicester LE17RH, United Kingdom
[*]Corresponding author: Shenghui Rong (rsh@ouc.edu.cn)

Abstract. The images captured in the underwater scene frequently suffer from blur effects due to the insufficient light and the relative motion between the captured scenes and the imaging system, which severely hinders the visual-based exploration and investigation in the ocean. In this paper, we propose a feature pyramid attention network (FPAN) to remove the motion blur and restore the blurry underwater images. FPAN incorporates the cascaded attention modules into the feature pyramid network (FPN) that enables it to learn more discriminative information. To facilitate the training of FPAN, we construct a weighted loss function, which consists of a content loss, an adversarial loss, and a perceptual loss. The cascaded attention module and the weighted loss function enable our proposed FPAN to generate more realistic high-quality images from the blurry underwater images. In addition, to deal with the lack of publicly available datasets in underwater image deblurring, we built two specific underwater deblurring datasets, namely Underwater Convolutional Deblurring Dataset (UCDD) and Underwater Multi-frame Averaging Deblurring Dataset (UMADD), to train and examine different deep learning-based networks. Finally, we conduct sea trial experiments on our autonomous underwater vehicle (AUV). Experimental results on two underwater deblurring datasets demonstrate our proposed method achieves satisfactory results, which validates the potential practical values of our proposed method in real-world applications.

Keywords: cascaded attention modules, FPAN, motion blur, underwater deblurring datasets.

*****Shenghui Rong, e-mail: rsh@ouc.edu.cn

1 Introduction

The ocean provides considerable storage of resources including food, oil, and national gas. The usage of advanced technologies in unmanned submersibles (e.g., an autonomous underwater vehicle, AUV) allow people to collect videos to perform the visual ocean exploration in the undersea world [1][58]. However, the images captured in the underwater scenes frequently suffer from blur effects due to the insufficient light and the relative motion between the captured scenes and the imaging system, which dramatically degrades the image visibility and affects the performance of the ocean tasks [2][59]. Thus, removing the motion blur to improve the image quality is of great significance.

Motion blur leads to degradation of image quality, which is generally caused by camera shaking or fast object motions [3]. Techniques have been developed on the image deblurring researches for decades, however, most of them focus on providing solutions for blurry images captured on land. These methods either try to estimate blur kernels and image priors [4-7] or to train a deep neural network to generate clear images directly from the blurry observations [8-12][61]. For the conventional optimization-based deblurring approaches, they are treated as a deconvolution process and the clear images can be obtained by using priors. Fergus *et al.* [5] estimated blur from camera shake using Gaussian scale mixture priors. Krishnan *et al.* [13] regularized the blurry images using a normalized sparsity method. Based on the dark channel prior of blurry images, Pan *et al.* [14] introduced a linear approximation of the minimum operator to compute the dark channel prior, which could be directly extended to non-uniform deblurring in practice. The conventional optimization-based deblurring algorithms promote the development of image deblurring techniques to a certain extent, but the performance on the blurry images with fewer corresponding features are unsatisfactory because these priors are usually designed under limited observations or restricted assumptions [3]. Additionally, some researchers attempted to combine conventional optimization-based methods with deep learning techniques, such as convolutional neural networks (CNNs), to estimate the blur kernels [15-18]. The majority of these blur kernel estimation approaches utilize a conventional optimization-based method in an iterative way, which has shown significant improvement over traditional deconvolution-based deblurring algorithms. However, they are commonly computationally expensive since they repeat a crucial step of a conventional optimization-based method for many times. Compared with conventional optimization-based approaches, deep learning-based image deblurring approaches usually obtain excellent deblurring performance and achieve real-time

processing speed. To train a deep neural network, one can conduct abundant experiments to collect slight or severe blurry images in diverse scenes on land. Different from the construction of land-based deblurring datasets, underwater images often suffer from low visibility (resulting in blur effects), this is because light is scattered and absorbed when traveling through the water [1]. Moreover, acquiring clear images in the underwater scenes is difficult. Thus, it is a challenging task to construct an appropriate underwater deblurring dataset for the motion blur removal task. Besides professional image deblurring algorithms, image restoration methods [1, 2, 19, 40, 41, 42, 58, 59, 64-66] are also able to remove the blur in underwater images. They mainly consider imaging models where the light is attenuated in water body. These methods have a significant effect on color restoration, as well as improving the image sharpness by removing the slight blur in the image. However, they show limited ability in restoring the severe blurry underwater images.

In this paper, we propose a deep learning method based on the cascaded attention mechanism, namely feature pyramid attention network (FPAN), to translate blurry images into clear ones. We also collect and provide two large-scale underwater deblurring datasets for training the underwater image deblurring networks. Both datasets contain clear and blurry images. Meanwhile, the blurry underwater images collected by our AUV-based imaging system are processed using the proposed method to verify the network performance. We compare the proposed method with three conventional methods [1, 13, 19] and two state-of-the-art methods [8, 11], and the experimental results show that our proposed method is more satisfactory.

Our contributions are summarized as follows: 1) We propose a deep learning network, which combines the cascaded attention module and the feature pyramid network (FPN) to remove motion blur and restore the brightness and sharpness of underwater images. 2) We collect and

release two large-scale underwater deblurring datasets for researchers to advance the development of underwater image deblurring.; 3) We conduct experiments on two underwater deblurring datasets, and evaluate the proposed method using real-world experiments on our AUV platform. The experiments show our proposed method achieves satisfactory results.

The rest of this paper is organized as follows. Sec. 2 briefly reviews the related works. Sec. 3 presents the details of the proposed network. Sec. 4 demonstrates the experimental results using the images from the validation sets and the sea trial dataset. Sec.5 presents the post-processing. Sec. 6 concludes this paper.

2 Related Work

In recent years, deep learning techniques had achieved great success in image transformation tasks, which provide an end-to-end solution to translate the distorted images into the clear ones. Previous works [20-22] estimated rigid or non-rigid transformations between two images for tasks such as motion estimation or matching using siamese networks. These networks usually need ground truth clear images, but the ground truth clear images are unknown in many application scenes. Later, the spatial transformer was proposed as a trainable module in classification networks to estimate the parametric transformations [23]. To handle articulations, the method of non-parametric transformations was used in the form of shape representation [24]. Although similar methods in [23, 24] with a convolutional variant can solve specific parametric transformation problems, there are several application scenes too complex to be representable by a small number of bases [20]. Recently, based on the concept of spatial transformer and mapping relationship, Nah *et al*. [8] proposed a multi-scale convolutional neural network (CNN) to restore the degraded images and the network could restore the blurry images in three different levels. Following this, Tao *et al*. [11] extended the multi-scale CNN with the long short-term memory

(LSTM) to produce a scale-recurrent CNN for blind image deblurring that generated promising deblurred results. Kupyn *et al*. [9] inherited the generative adversarial network (GAN) from [25] to construct the DeblurGAN with the gradient penalty and the perceptual loss, that enable the DeblurGAN achieve satisfactory results. Built on the success of DeblurGAN, Kupyn *et al*. [26] proposed DeblurGAN-v2, which was another substantial push on GAN-based motion deblurring framework. The end-to-end deep learning-based methods mentioned above show excellent performance in restoring the blurry images with fewer artifacts than the conventional optimization-based methods [13]. In addition, deep learning-based methods do not need to estimate the blur kernel.

Except removing the image blur in an end-to-end way, deep learning-based methods can also be used as a core step to estimate the blur kernel. Schuler *et al*. [27] designed the deep network architectures for blur kernel estimation by imitating the alternating minimization steps in the conventional optimization-based methods. For studying the spectral property of blurry images, Chakrabarti *et al*. [28] applied a deep CNN to predicting the Fourier coefficients, and the estimated blur kernel was obtained with the coefficients in a projection way. In [29], CNN is used to predict the parametric blur kernels for motion blurry images. Although these CNN-estimated blur kernel methods give another solution for removing the blur in an image, they are not efficient enough since they repeat a step for many times.

The training of deep neural networks is frequently a time-consuming task, and a commonly used network architecture (e.g. encoder-decoder) is usually able to solve many image translation issues, but the results are not impressive enough. In recent years, the attention mechanism is widely utilized for efficiently training a deep network in computer vision tasks, which helps generate satisfactory results [30-32]. The principle of the attention mechanism is that the

importance of different features can be weighed by learning an intermediate attention map and then applying the element-wise product on the attention map and the source feature map [33]. For the task of underwater image processing, the weak textures and features that are crucial in an image can be learned by an attention-based network such as the underwater object located in the low visibility environment and suffering from motion blur.

In this paper, we carry out the research of underwater image deblurring in an end-to-end way. We aim to remove the underwater image blur induced by low visibility, object motion, and camera shaking. As the blur in the images is caused by multiple factors, and the object features are not conspicuous in these images, it is important to propose a network, which can learn more robust features from the training data. The architecture of our network is inspired by Lin *et al*. [34], Mei *et al*. [30], and Kupyn *et al*. [26]. The FPN was proposed by Lin *et al*. [34] for object detection task, and achieved satisfactory results. It is a kind of structure containing a bottom-up and a top-down pathway. The bottom-up pathway is a common convolutional network for feature extraction. The spatial resolution is down-sampled in this pathway and semantic context information is extracted and compressed in this process. As for the top-down pathway, FPN reconstructs spatial resolution from the semantically rich layers. The lateral connections are constructed between the bottom-up and top-down pathways in FPN, which supplement high-resolution details and help localize objects. Inspired by this, Kupyn *et al*. [26] first introduced the idea of FPN to the field of image restoration and enhancement. Later, Zhang *et al*. [35] proposed an attention mechanism in the deep network framework to train GAN, which shows excellent performance. Mei *et al*. [30] trained a model to address the problem of image denoising and image super-resolution using FPN and the attention mechanism. Our network inherited from the structure of FPN. We incorporate those priors and propose a FPAN. Different from the previous

work using one attention module to connect network, we propose an cascaded attention network architecture, which allows our network to learn more details. The network architecture will be introduced in the next section.

3 Methodology

Conventional methods formulate the image deblurring task as a deconvolution problem when the blur kernel is spatially invariant [16]. Let $I_b(x)$ be the blurry image, $I_c(x)$ be the latent clear image, K be the blur kernel, N be the additive white Gaussian noise. The model can be defined as

$$I_b(x) = K * I_c(x) + N \quad (1)$$

Different from the conventional image deblurring methods, the deep learning-based methods provide a simple and direct mapping relationship between the blurry image $I_b(x)$ and the latent clear image $I_c(x)$, which can be expressed as

$$I_c(x) = f(I_b(x), \theta) \quad (2)$$

where f is the complex deep CNN transfers, the blurry image to the latent clear image. θ is the parameter of the deep CNN. Existing deep learning frameworks, especially GAN [36-38], achieved great success in the field of image translation tasks. There are two competing networks in a standard GAN, namely the generator network and the discriminative network. The images generated by the generative network are put into the discriminative network, and the discriminative network judges whether the output results are realistic images or not. However, it requires a large-scale dataset for training, hence, we determine to construct datasets for training our GAN to achieve the mapping function f, and we can easily obtain the latent clear image in an end-to-end way.

3.1 Network Architecture

The pipeline of our proposed network is illustrated in Fig. 1, which includes multiple layers connected in sequences.

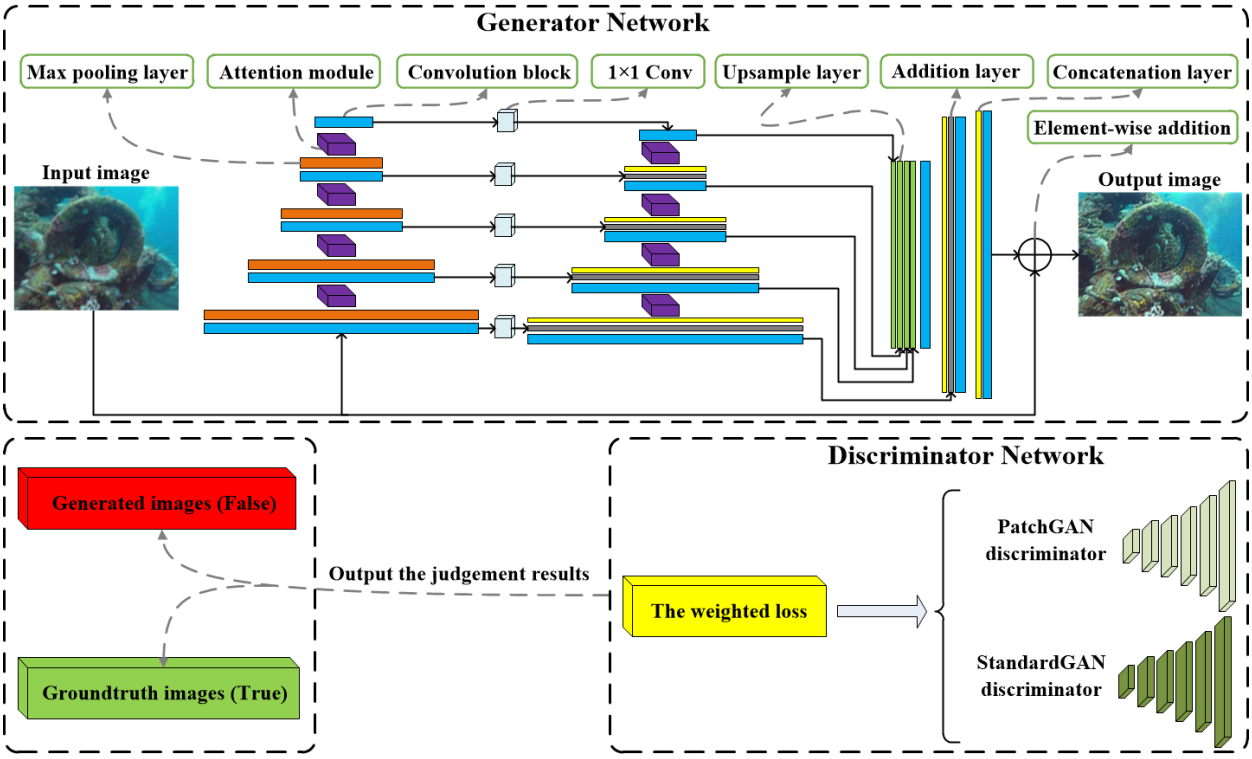

Fig. 1 The architecture of our proposed network.

The generator network. Considering the dark scene and the inconspicuous features of the underwater images, the designed network should learn adequate complex information from the images. Based on this, we choose the FPN backbone of our generator network. Our proposed network takes a three-channel RGB (red, green, and blue channels) image as the input and outputs five feature maps with different scales. The bottom-up pathway for feature extraction is a 3-kernel-2-stride-1-padding convolutional network, and the channels are set to 3, 64, 128, 256, and 512, respectively. The features are transferred to the top-down pathway through the lateral connections and reconstructed spatial resolution from the semantically rich layers. The channel numbers in the top-down pathway are the same with that in the bottom-up pathway. To restore the original image resolution, two up-sampling layers and

convolutional layers are added to reconstruct the spatial resolution. Then a skip connection is used to learn the residual between the input image and the output image of the convolutional layer, and the final output image is obtained after the element-wise addition module.

The cascaded attention mechanism. Although the FPN architecture alone can remove the blur, its performance in actual applications is limited. Therefore, we add the convolutional block attention modules (CBAMs) in our generator network. As is shown in Fig. 2, the CBAMs consists of the channel attention module and the spatial attention module. The channel attention module exploits the inter-channel relationships and focuses on "what" is meaningful given an intermediate feature map *F*, and it can be defined as

$$M_c(F) = \sigma\left(MLP\left(AvgPool\left(F\right)\right)\right) + MLP\left(MaxPool\left(F\right)\right) \quad (3)$$

where $M_c(F)$ is the output channel attention map, *MLP* is the multi-layer perceptron with one hidden layer, σ is the sigmoid function. The spatial attention module utilizes the inter-spatial relationship and concentrates on "where" is an informative area in an image, which can be defined as

$$M_s(F) = \sigma\left(\mathrm{Conv}\left(\left[AvgPool\left(F\right); MaxPool\left(F\right)\right]\right)\right) \quad (4)$$

where $M_s(F)$ is the output spatial attention map, *Conv* is a 7×7 size convolution operation. The attention mechanism is simple but effective for feed-forward convolutional neural networks. It sequentially infers attention maps along two separate dimensions, channel and space when given an intermediate feature map. The attention maps are then multiplied to the input feature map for adaptive feature refinement [39], and it can be expressed as

$$F_{\mathrm{out}} = \left\{M_s\left[M_c\left(F\right)\otimes F\right]\right\}\otimes\left[M_c\left(F\right)\otimes F\right] \quad (5)$$

where F_{out} is the final output feature map from the CBAM module, $\otimes$ is the elenment-wise multiplication. As the attention mechanism has an advantage in helping learn more textures and features information,

thus, we add eight CBAMs in the FPN architecture to form a cascaded attention network. One CBAM is located in the meddle of two different layers, therefore, the next layer of the neural network will learn the information that has been processed by the attention mechanism module in the previous layer in such a designed serial network. Thus Eq. (5) can be rewritten as

$$F_{\text{out}}^{i} = \left\{ M_{s}^{i}\left[M_{c}^{i}\left(F^{i} \right) \otimes F^{i} \right] \right\} \otimes \left[M_{c}^{i}\left(F^{i} \right) \otimes F^{i} \right] \quad (6)$$

where i is the index of the i-th CBAM, $i \in \{1, 2, 3, 4, 5, 6, 7, 8\}$. Moreover, the convolution blocks and the addition layers with the same number of channels are connected by a 1×1 convolution layer, which allows the information processed by the attention mechanism to be used more fully. Finally, the generator network produces a deblurred image $G_{\theta}^{F_{out}^{i}}$. Since we aim to restore the blurry underwater images and overcome the challenges introduced by the dark underwater scene, the introduction of the attention mechanism like the CBAM can meet the needs of the FPN to refine more image details. Taking the underwater camera mounted on our AUV platform as an example, an AUV's speed and the turbulence in the sea will directly affect the blur degree of the captured visual data. In this situation, common CNN shows limited ability in removing the blur and refining the details. By using the attention mechanism, the proposed network can produce clear and bright images.

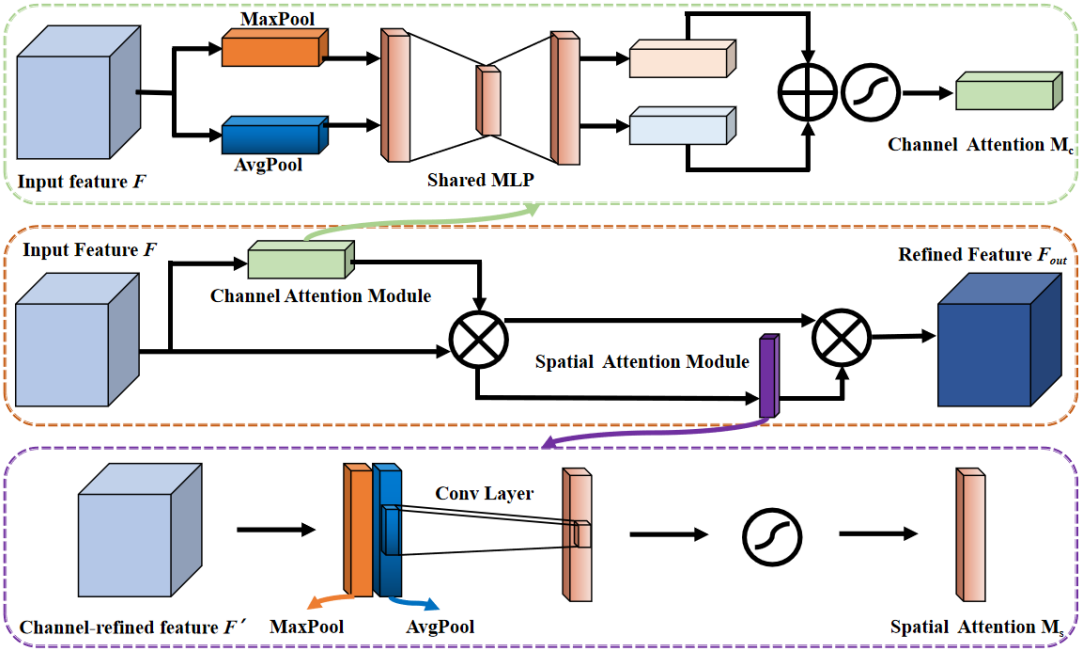

Fig. 2 The architecture of convolutional block attention module.

The discriminator network. A discriminator is like a "judge", which is able to distinguish the realistic clear images from the fake clear images generated by the generator. To let the discriminator be more intelligent, we inherit the wisedom from Isola *et al.* [37]. They propose a PatchGAN discriminator and take the advantage of both the local information and the global information in an image generating sharper images than a standard discriminator. We take a further step to combine their discriminator with our proposed generator together. In this way, it is essential for our proposed network to learn both global information and local information from the training data. As shown in Fig. 1, the input and output of the discriminator are both a three-channel RGB image. The architecture of our discriminator is a 4-kernel-2-stride-2-padding convolutional network, and the channels are set to 3, 64, 128, 256, 512, and 1 in this module, respectively. Together with the generator network, the discriminator network uses the dataset to alternately train to address the min-max problem, which can be expressed as

$$\min_{G}\max_{D} E\left[\log D_r\left(I_c\left(x\right)\right)\right]+E\left[\log\left(1-D_r\left(G_{\theta}^{F_{out}^{i}}\left(I_b\left(x\right)\right)\right)\right)\right] \quad (7)$$

where θ is the learnable parameter in the generator network, E denotes the mean.

3.2 Training Objective

An image translation GAN framework is notoriously hard to train. In previous works [26] and [36], a weighted loss function showed satisfactory performance in training a complex GAN mapping framework. We inherited the priors of the weighted loss function and proposed a novel loss function aiming at improving the quality for the blurry underwater images. It is a three-term loss function, which consists of the content loss L_{con}, the adversarial loss L_{adv}, and the perceptual loss L_{per}. Among them, the content loss L_{con} can yield over smoothened pixel-space outputs [36, 38]. As the underwater scenes are usually dark, and the camera or the object is in a motion condition, the captured underwater images suffer from different degrees of blur. Fine details of the original underwater scenes cannot be reconstructed effectively. In order to reconstruct the blurry areas and the main features, L_{con} is utilized as the first term of our loss function, which is defined as

$$L_{con} = \sum_{x} \left| I_c(x) - G_{\theta}^{F_{out}^i}(x) \right| \quad (8)$$

However, L_{con} alone cannot generate satisfactory resultant images, the resultant images are still blurry and usually lack high frequency details [20]. Hence, relativistic average least squares GAN [36] (RaLSGAN) objective loss (the adversarial loss L_{adv}), is used to further to improve the high frequency details in the images. It has proven in [17] that L_{adv} can allow the network to learn sharper edges and more detailed textures by estimating the probability that the original image is more realistic or not than the blurry image reconstructed by the generator. L_{adv} is expressed as

$$L_{adv} = E[(G_{\theta}^{F_{out}^i}(I_c(x)) - E[G_{\theta}^{F_{out}^i}(I_b(x))] + 1)^2] + E[(G_{\theta}^{F_{out}^i}(I_b(x)) - E[G_{\theta}^{F_{out}^i}(I_c(x))] - 1)^2] \quad (9)$$

Previous works [20], [38], and [9] introduced the perceptual loss L_{per} as a part of their loss function. In terms of L_{per}, it aims to measure the CNN feature space differences between the generated images and the target images, which shows excellent performance in weakening or

eliminating the artifacts. To remove the inevitable artifacts in resultant images, we regard L_{per} as a suitable training objective in our proposed loss function. L_{per} is defined as

$$L_{\text{per}} = \sum_{x} \left| \varphi\left(I_c(x)\right) - \varphi\left(G_{\theta}^{F_{out}^{i}}\left(I_b(x)\right)\right) \right| \quad (10)$$

All the loss functions mentioned above are used as the metrics to compare the reconstructed images and the original ones during the training process. Thus, our loss function can be defined as

$$L = \lambda_c \cdot L_{con} + \lambda_a \cdot L_{adv} + \lambda_p \cdot L_{per} \quad (11)$$

where λ_c, λ_a, and λ_p are, respectively, the weighted parameters of the corresponding loss function.

3.3 Training Datasets

Ground truth clear images cannot be obtained in the underwater scenes. The synthesized underwater image datasets [62, 63] address this issue to some extent, and sufficient training data can be provided for deep learning based CNNs. However, existing underwater datasets [40-42, 62, 63] mainly aim at addressing the issue of object recognition and image enhancement. To our best understanding, available underwater deblurring datasets are rare for training a deep deblurring neural network. To produce sufficient training data, current mainstream works synthesize deblurring images from the clear images captured on land. The synthesis methods can be divided into two categories: 1) convolving clear images with real-world or generated blur kernels [18, 27, 28], and 2) averaging consecutive clear frames from videos captured by a high-speed motion camera [8, 43, 44, 45]. The convolving-based method is a simplified image formation model, and all pixels in the generated image share the same blur kernel trajectory. Thus the synthetic images look different from the real-world motion blur images, which are more similar to those captured with the camera out of focus. To overcome the drawbacks of the convolving-based method, the averaging-based method adopts a multi-frame accumulation

strategy that would be equivalent to collect real-world blurry images containing the camera or the object motion. As both of the out-of-focus blur and motion blur exist in practice, we determine to inherit the above approaches to propose high-quality underwater deblurring datasets for training deep neural networks. The construction of the underwater deblurring datasets are carried out with due consideration to an AUV's operating environment and scenario, thus the parameters of the datasets are configured in conjunction with an AUV's motion characteristics. Based on this, we propose two datasets, namely the underwater convolutional deblurring dataset (UCDD) and the underwater multi-frame averaging deblurring dataset (UMADD). Both datasets contain two image sequences of the same contents: one blurry image sequences with blur by a shakable camera, and another one is the corresponding clear image sequences. Our divers manually used GoPro 8 Hero Black camera to capture 19 videos (120frames per second in the linear mode) at 1920×1080 resolution in Bali. The videos we capture take full account of the content diversity and dynamic motion transformation. Then the clear images are extracted from these videos. Both UCDD and UMADD are generated from these videos, and we describe the production of the datasets in detail below.

UCDD: By considering the cableless underwater robotics with limited energy to collect images in underwater scenes, the camera often cooperates with an auxiliary light source. The acquired images are often blurry and dim, which are like images under the condition of an out-of-focus imaging blur. Inspired by the works in [4] and [6], we propose the model of random trajectories generation to simulate realistic and complex blur kernels that has similar blurring effect when acquiring images with underwater motion platforms. It takes us a further step to generate the blurry and dim data, which is equivalent to the images captured in dark environment under shaking conditions. The blur kernels are stimulated by applying sub-pixel interpolation to

the trajectory vector. For each trajectory vector, it is a complex valued vector corresponding to the discrete positions of an object following 2D random motion in a continuous domain. In this process, Markov process is used to generate the blur trajectory, and the position of the next point of the blur trajectory is randomly generated based on the previous point velocity and position, Gaussian perturbation, impulse perturbation, and deterministic inertial component [9]. To render blurry images in different levels, we extract frames from the videos and set the exposure time as 0.5s, 0.25s, 0.125s, and 0.0625s to generate blurry images. The exposure time setting is appropriate for underwater visualization by an AUV, which often use strobe lights in conjunction with cameras for visual image acquisition. For each exposure time, we generate the same number of blurry images. Examples of clear and blurry image pairs are displayed in Fig. 3. In total, we generate 36, 204 pairs of synthetic blurry images and the corresponding ground truth clear images. The UCDD is publicly available at: https://drive.google.com/file/d/1N-IqijFyiMBAr9henV07a7SccNGgeUt7/view?usp=sharing.

Fig. 3 Examples of clear and blurry image pairs in UCDD. The exposure time for clear images and blurry images is 0.5s and 0.25s in the first row, and 0.125s and 0.0625s in the second row.

UMADD: The pipeline of averaging-based method contains underwater videos collection, frame interpolation, and blur synthesis. The blurry images are generated by accumulating clear images stimulation at every time during the camera exposure [8, 44]. It can be approximately

defined as averaging the pixel values at the same location in high-speed consecutive video frames

$$I_b = \frac{1}{T}\int_{t=0}^{T} F(t)dt \approx \frac{1}{N}\sum_{n=0}^{N-1} F[n] \quad (12)$$

Where T is the exposure time of the camera, $F(t)$ is the light signal at time t, N is the number of frames, $F[n]$ is the light signal of the n-th clear image.

Fig. 4 Examples of clear and blurry image pairs in UMADD.

When recording these video frames, the camera should use a high-frame rate mode to ensure that a large number of video frames are captured in the same exposure time. Meanwhile, special attention should be paid to the quality of each frame since we aim to average these clear frames to generate a blurry one. The GoPro 8 Hero Black camera can be set to a maximum 240 fps when capturing a video, which can satisfy the need of capturing enormous video frames. However, high-frame rate data capture is achieved at the expense of video frame quality for most of high-speed cameras. The consumer-level cameras (including Gopro Hero Black Series) have a limited computational ability in recording all light signals in the cell arrays during the readout time. It is strictly related by the exposure time that leads to a tradeoff between the noise and the blur. Short exposures can reduce the blur at the cost of the increasing noise, while long exposures reduce the noise at the cost of the increasing blur [4]. Thus we inherit the previous wisedom [44]

and set the frame rate as 120 fps with satisfactory compromise to the quality and quantity of the captured video. Then an advanced video interpolation technique is applied to expand the frame rate from 120 to 1920, which aims to make the blur more natural and smooth. When the object or camera move very fast, the averaging operation on the video can produce unnatural result from two adjacent frames [46]. In this situation, the video interpolation technique can help to adjust the frame rate to a high enough level to alleviate or eliminate these unnatural steps. In this paper, an adaptive separable convolution video interpolation [47] is utilized to address the problem unnatural steps and aid nonlinear motion blur generation. Different from the standard optical flow method, the adaptive-separable-convolution-based video interpolation formulates frame interpolation as local separable convolution over input frames using pairs of one dimensional kernels, which can produce more visually pleasing frames [47]. After the video interpolation operation, we average 241 successive clear images to generate one blurry image and define the 121^{st} clear image as the corresponding ground-truth image. For example, the first blurry image is the mean from 1^{st} frame to 241^{st} frame, the second blurry image is the mean from 241^{st} frame to 481^{st} frame. The operation of averaging 241 frames is able to simulate the maximum exposure time of GoPro 8 Hero Black camera since the camera can capture maximum 240 frames in one second. It is consistent with our goal of obtaining as much experimental data as possible. Finally, we generate 2, 842 pairs of blurry and clear images at 1280×720 resolution. Examples of clear and blurry image pairs are displayed in Fig. 4. The UMADD is publicly available at: https://drive.google.com/file/d/1rfY3ha_CJ2YJU6mK9OHizbmS4ZmKmenr/view?usp=sharing.

3.4 Details of Training

We trained the network on datasets UCDD and UMADD, respectively. As mentioned in the previous part, 36, 204 image pairs are generated in UCDD. We split UCDD into 32, 588 image pairs as the training set and 3, 616 image pairs as the validation set. The image pairs in these sets contains the

same proportion of four kinds of exposure time. For UMADD, we augmented the dataset by rotating the original image clockwise by 90°, 180° and 270°. Finally, there are 11, 368 pairs in the dataset, including 10, 231 image pairs in the training set and 1, 137 image pairs in the validation set. To improve the training efficiency, all the data was resized to 512×512 resolution for training and testing. For the training objective, the weights of the content loss L_{con}, the adversarial loss L_{adv}, and the perceptual loss L_{per} are set as 0.5, 0.006, and 0.01, respectively. The training objective is optimized to minimizing the distance between the generated image and the ground truth. We trained the network with both UCDD and UMADD for 100 epoches, the initial learning rate is set to 0.0001, and the batch size is set to 4.

4 Experimental Results and Analysis

Fig. 5 The experimental scene in Jiaozhou Bay, Qingdao. Three GoPro 8 Hero Black cameras are mounted on the head of AUV. Tow areas marked by red rectangles are the head of AUV and the GoPro 8 Hero Black camera we used in the experiments, respectively.

In this section, we evaluate the proposed network with both the validation sets and the sea trial dataset. The sea trial dataset (in total 25 images) are realistic blurry underwater images collected by our AUV, which consists of the underwater natural scenes and sediments. To demonstrate the effectiveness of our proposed network, we compare it with several representative methods proposed in recent years, including the deep learning-based methods and the conventional methods. These selected comparison methods are proposed to address the problem of image deblurring or underwater image restoration in recent years. The methods of Nah *et al.* [8], Tao *et al.* [11], Kupyn *et al.* [26], and Mao *et al.* [61] are specially deep

learning-based image deblurring methods. The method of Krishnan *et al*. [13] is one of the conventional optimization-based deblurring method. The method of Wang *et al*. [60] is a typical representative transformer-based algorithm for image restoration. Additionally, the methods of Peng *et al.*[1] and Fu *et al.*[19] are selected as the comparison methods as they are the latest conventional optimization-based underwater image enhancement methods. The source codes of the selected comparison methods are all provided by the authors on Github [48]. All comparison methods except Nah *et al.* [8] and Tao *et al.* [11], Kupyn *et al*. [26], Wang *et al*. [60], and Mao *et al*. [61] are implemented on MATLAB R2019b [49] framework with a Win 10 platform. The deep learning-based comparison methods are implemented on PyTorch [50] with an Nvidia GTX 1070Ti GPU, Ubuntu platform. In the testing stage, we test all the comparison methods on both the UCDD validation set (including 3, 316 images) and the UMADD validation set (including 1, 137 images). We also conduct experiments in Jiaozhou Bay and collect a sea trial dataset (including 25 images). Our AUV platform is equipped with one GoPro 8 Hero Black camera, which is mounted on the bottom of the AUV. The camera's field of view is in the direction of the seafloor. Fig. 5 shows our AUV platform and the experimental site.

We evaluate our proposed method in qualitative and quantitative ways. The qualitative evaluations mainly depend on the evaluation of image quality by human visual system. As for quantitative evaluations, two full-reference evaluation metrics and several non-reference evaluation metrics are used. The full-reference evaluation metrics are Structural Similarity Image Metric (SSIM) [51] and Peak Signal to Noise Ratio (PSNR) [52]. Several commonly used non-reference image quality evaluation metrics are employed to compare the performance of different methods in this paper. They are non-reference image spatial quality evaluator (BRISQUE) [53], naturalness image quality evaluator (NIQE) [54], patch-based contrast quality index (PCQI) [55], and underwater image quality metric (UIQM) [56], respectively. The score of BRISQUE is

based on a support vector regression (SVR) model trained on an image database that contains images with different distortions (e.g., blurring, artifacts, and noise). It can intuitively represent the perceptual image quality and the blur recovery capability. NIQE is an evaluation metric to judge the natural state of the image globally, which is based on constructing a series of features to measure image quality and fitting these features to a multivariate Gaussian model. These features are extracted from simple and highly regular natural landscapes to measure the differences in the multivariate distribution of an image. For BRISQUE and NIQE, the smaller scores the better image quality. As for PCQI, it provides accurate predictions on the human perception of contrast variations using a metric based on an adaptive representation of local patch structure. In terms of UIQM, it is a specific underwater image quality metric, which is obtained by assigning carefully calculated weights to UICM on color, UISM on sharpness, and UIConM on contrast. The higher scores of PCQI, UIQM, UICM, UISM, and UIConM indicate the image has better quality.

4.1 Ablation Study and Analysis

Fig. 6 shows the qualitative comparison results of the ablation study on the UCDD validation set. To verify the effectiveness of the attention mechanism, we start from the original FPN-based GAN for image deblurring, then we add the attention mechanism into the FPN to form the FPAN. Instead of using 512×512 image pairs, we use 1280×720 image pairs to carry out the ablation study. This is because the performance of the attention module is more visible in high-resolution images.

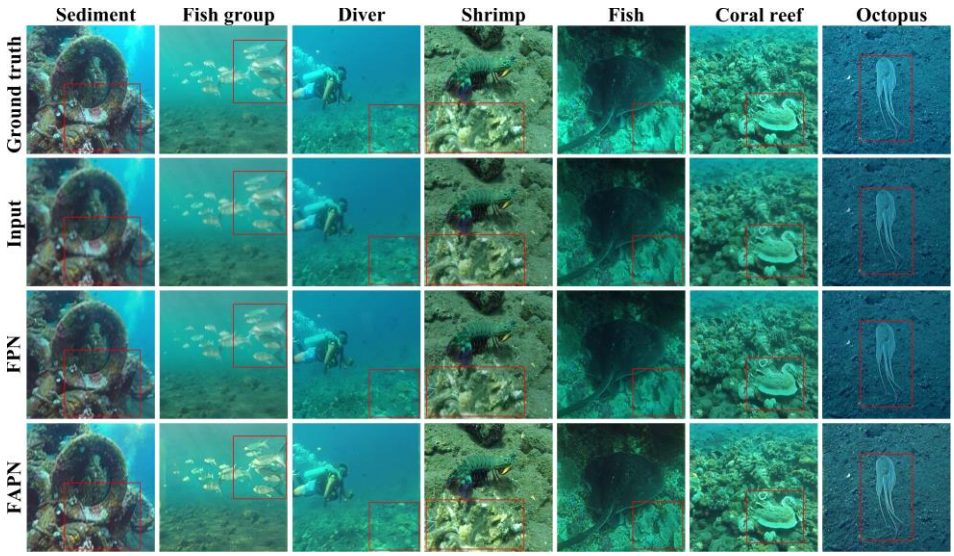

Fig. 6 Qualitative comparison of different network architectures in the ablation study.

Table 1 Quantitative comparison of different network architectures in the ablation study using non-reference metrics. The values indicate the average scores of the images on the UCDD validation set[a, b].

Methods	SSIM	PSNR	BRISQUE*	NIQE*	PCQI	UIConM	UICM	UISM	UIQM
FPN	22.246(2)	**0.611(1)**	28.210(2)	4.077(2)	7682.2(2)	0.662(2)	-82.019(2)	3.511(2)	1.092(2)
FPAN	**24.155(1)**	0.594(2)	**19.071(1)**	**3.308(1)**	**8699.8(1)**	**0.730(1)**	**-74.813(1)**	**3.917(1)**	**1.559(1)**

[a]The values in bold represents the best results.
[b]The number in brackets refers to the ranking 1-2 of a method on the metric.

As shown in Fig. 5, we observe that both FPN and FPAN architectures are able to remove the blur in the images. Our network architectures FPAN outperform FPN in removing the blur and improving the brightness, especially in the images of Fish group, Coral reef, and Octopus in Fig. 6. From the qualitative results, we can confirm that the attention mechanism plays an important role in restoring the object details in the blurry images (see the seafloor and fish in Fish group), and the results generated by FPAN are much closer to the ground truth. We also report the quantitative results in Table 1, the proposed FPAN ranks the first in eight of nine metrics. For PSNR metric, the quantitative results of FPAN and FPN are very close. Based on the ablation study, it can be confirmed that the FPAN is able to learn more information from the training data. Our data are collected at different depths in the underwater environments and the light suffers from different level of attenuation due to the varied depths, which needs a strong learning

network to extract weak textures and features. Thus, FPAN is suitable to address the ill-posed image translation task.

4.2 UCDD validation set

The qualitative comparisons on the UCDD validation set are shown in Fig. 7, from which we can observe that the input images are much more blurry than the ground truth images. Fu's method [19] shows strong ability in compensating for the color, but it generates images with significant fogging mask. Peng's method [1] can improve the image quality by removing color casts although some color and bright differences exist, however, Peng's method [1] shows an unsatisfactory deblurring result even though the image blur is very slight (e.g., the Sediment and the Fish in Fig. 7). Krishnan's method [13] shows limited deblurring performance, it generates slight artifacts in the resultant images. For the deep learning-based methods, the qualitative results of Nah's method [8] and Tao's method [11] are similar in removing the blur. Wang's method [60] can remove the "noise points" for a degraded image that seems to be smooth globally but show limited ability in removing the blur. The results after processing by Kupyn's method [26] are close to those of Tao's method [11]. As for Mao's method, it shows a competitive result in removing the blur on UCDD validation set. Our proposed method generates high quality images with much better visual appearances as shown in Fig. 7. Except the excellent deblurring ability, our proposed method can evenly improve the brightness in the entire image compared with other comparison methods. We notice that some inevitable slight artifacts exit in the results of our proposed method (e.g., the left edge in Fish group processed using our proposed method). It is reasonable since there is a brightness gradient in the image of Fish group, it is dark-bright-dark from the top to the down. The FPN-based framework with the attention mechanism can learn such information and alleviate the artifact problem. The divers made a significant effort to collect a wide range of underwater scenes and animals, it inevitably captured images with large illumination and darkness changes. Nevertheless, the qualitative result of our proposed method is still the best among all the comparison methods.

In addition, we report the quantitative comparison of different methods on the validation set in Table 2 using both full-reference metrics and non-reference metrics. Our proposed method shows superior performance than other methods, it ranks the first in 5 of 9 metrics. For the other four indicators, there is only a small gap between our results and the best results. For the full-reference metrics, it indicates our proposed method shows competitive performance. We get the highest score with UIQM, which is consistent with the qualitative analysis in terms of the contrast, color, and sharpness among the comparison methods. The BRISQUE metric can reflect the ability to restore image distortions and the NIQE metric evaluates the results in terms of the proximity of the restored images to the natural underwater images. Our results outperform other methods in both BRISQUE metric and NIQE metric. As PCQI metric is computed based on an adaptive representation of local patch structure for providing accurate predictions on the human perception of contrast variations [55]. Fu's method ranks the first and its resultant images are more consistent with the human perception, thus the score is reasonable.

Table 2 Quantitative experimental results of different comparison approaches on the UCDD validation set using full-reference metrics and non-reference metrics. The values indicate the average scores of the images[a, b].

Methods	SSIM	PSNR	BRISQUE*	NIQE*	PCQI	UIConM	UICM	UISM	UIQM
Fu *et al.* [19]	0.618 (4)	19.973 (9)	35.733 (8)	4.031 (3)	**9156.7 (1)**	0.466 (8)	−70.915 (2)	6.528 (8)	1.595 (8)
Peng *et al.* [1]	0.534 (8)	21.220 (8)	31.793 (6)	5.135 (7)	8318.6 (3)	0.717 (2)	−70.925 (3)	6.629 (7)	2.519 (2)
Krishnan *et al.* [13]	0.508 (9)	22.079 (6)	33.133 (7)	5.609 (8)	7752.1 (5)	**0.722 (1)**	−81.481 (4)	6.820 (5)	2.297 (3)
Nah *et al.* [8]	0.660 (2)	21.965 (7)	29.847 (4)	4.049 (4)	7618.0 (9)	0.696 (4)	−83.308 (9)	6.843 (3)	2.159 (4)
Tao *et al.* [11]	0.615 (5)	22.306 (5)	30.828 (5)	4.174 (5)	7701.6 (7)	0.678 (6)	−82.112 (8)	6.750 (6)	2.101 (7)
Kupyn *et al.* [26]	0.625 (3)	22.354 (4)	29.315 (3)	4.288 (6)	7695.5 (8)	0.672 (7)	−82.040 (7)	6.892 (2)	2.123 (6)
Wang *et al.* [60]	**0.701 (1)**	**22.771 (1)**	53.831 (9)	5.725 (9)	7743.9 (6)	0.464 (9)	−81.842 (5)	6.345 (9)	1.226 (9)
Mao *et al.* [61]	0.594 (6)	22.515 (3)	28.444 (2)	3.916 (2)	7762.3 (4)	0.681 (5)	−82.001 (6)	6.832 (4)	2.141 (5)
Ours	0.588 (7)	22.599 (2)	**27.174 (1)**	**3.461 (1)**	8462.5 (2)	0.7054 (3)	**−70.280 (1)**	**6.938 (1)**	**2.589 (1)**

[a]The values in bold represents the best results.
[b]The number in brackets refers to the ranking 1-9 of a method on the metric.

4.3 UMADD validation set

In this subsection, we test our proposed method on UMADD validation set, the qualitative results and quantitative results are shown in Fig. 8 and Table 3, respectively. Although the generated blurry images are different from the images in UCDD, the qualitative results are similar to the results in Fig. 7. The

resultant images of our proposed method present superior perceptual quality over that of other methods. The restored images shows good potential in improving the brightness and sharpness. As the quantitative results reported in Table 3, Tao's method [11] and Mao's method [61] outperform other methods in terms of full-reference metrics, since they are designed for removing the motion blur generated in an averaging multi-frame way. However, they show limited ability in non-reference metrics. On the UMADD validation set, our proposed method achieves the first place in terms of BRISQUE, NIQE, UIConM and UIQM, and also ranks top four for PCQI.

Table 3 Quantitative experimental results of different comparison approaches on the UMADD validation set using full-reference metrics and non-reference metrics. The values indicate the average scores of the images[a, b].

Methods	SSIM	PSNR	BRISQUE*	NIQE*	PCQI	UIConM	UICM	UISM	UIQM
Fu *et al.* [19]	0.727 (9)	20.735 (9)	40.240 (7)	4.740 (6)	9585.0 (8)	0.294 (9)	**−77.467 (1)**	4.914 (9)	0.317 (9)
Peng *et al.* [1]	0.784 (6)	24.837 (8)	31.178 (2)	3.984 (4)	9568.6 (9)	0.575 (3)	−87.629 (2)	6.774 (7)	1.586 (2)
Krishnan *et al.* [13]	0.781 (7)	30.372 (5)	30.299 (3)	5.264 (8)	**9828.7 (1)**	0.542 (5)	−100.953 (4)	**7.374 (1)**	1.269 (4)
Nah *et al.* [8]	0.820 (2)	31.248 (4)	41.872 (8)	4.954 (7)	9782.2 (7)	0.501 (7)	−101.402 (8)	6.930 (5)	0.977 (7)
Tao *et al.* [11]	**0.822 (1)**	31.504 (3)	39.470 (6)	3.712 (2)	9825.2 (2)	0.543 (4)	−102.128 (9)	6.836 (6)	1.082 (6)
Kupyn *et al.* [26]	0.818 (3)	31.576 (2)	33.542 (4)	3.832 (3)	9824.8 (3)	0.595 (2)	−101.107 (7)	7.218 (2)	1.408 (3)
Wang *et al.* [60]	0.767 (5)	29.303 (6)	54.767 (9)	5.481 (9)	9801.8 (6)	0.395 (8)	−101.012 (5)	6.313 (8)	0.427 (8)
Mao *et al.* [61]	0.818 (3)	**31.865 (1)**	36.503 (5)	4.000 (5)	9810.6 (5)	0.535 (6)	−101.066 (6)	7.050 (3)	1.143 (5)
Ours	0.751 (8)	26.255 (7)	**23.421 (1)**	**3.326 (1)**	9818.5 (4)	**0.633 (1)**	−88.968 (3)	7.043 (4)	**1.835 (1)**

[a]The values in bold represents the best results.
[b]The number in brackets refers to the ranking 1-9 of a method on the metric.

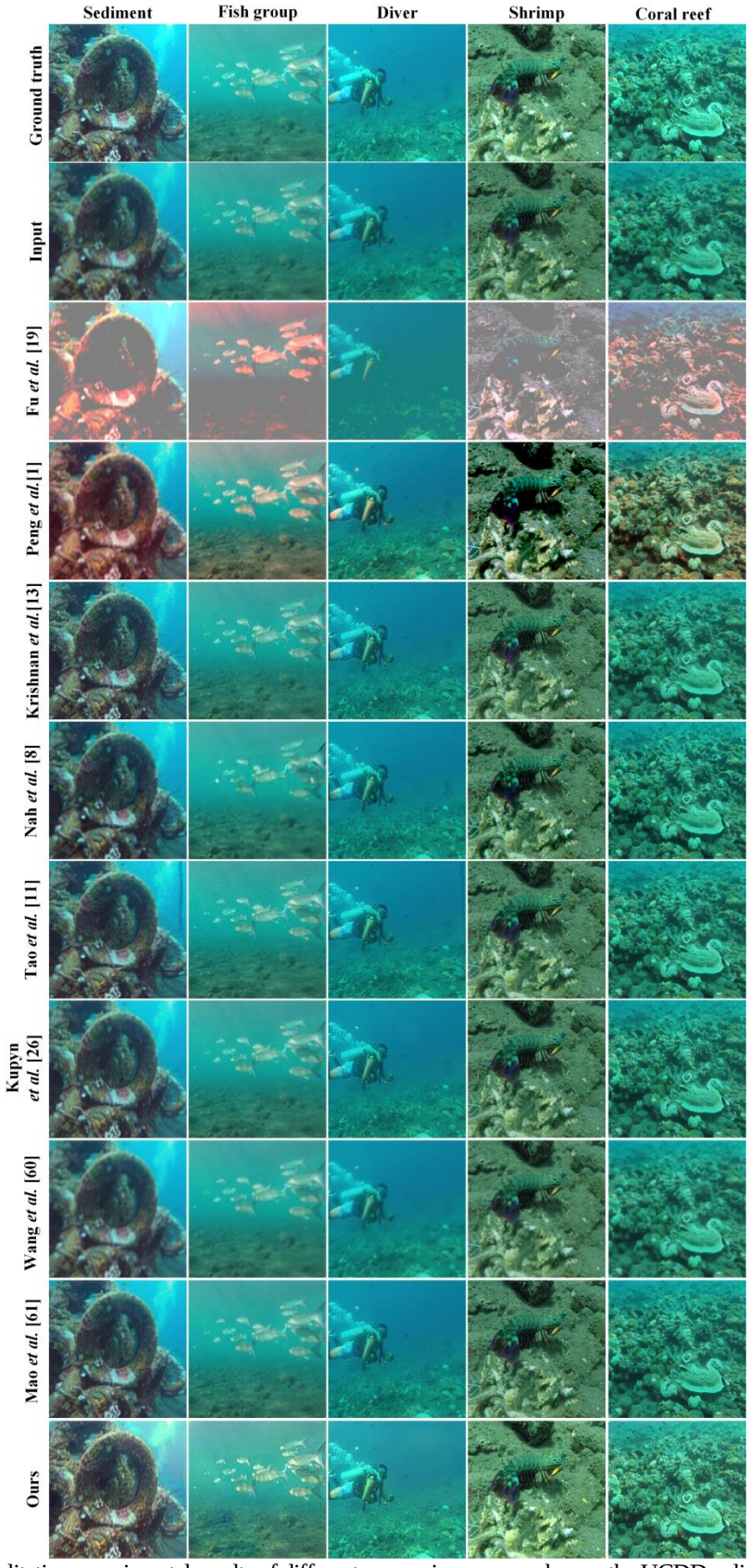

Fig. 7 Qualitative experimental results of different comparison approaches on the UCDD validation set.

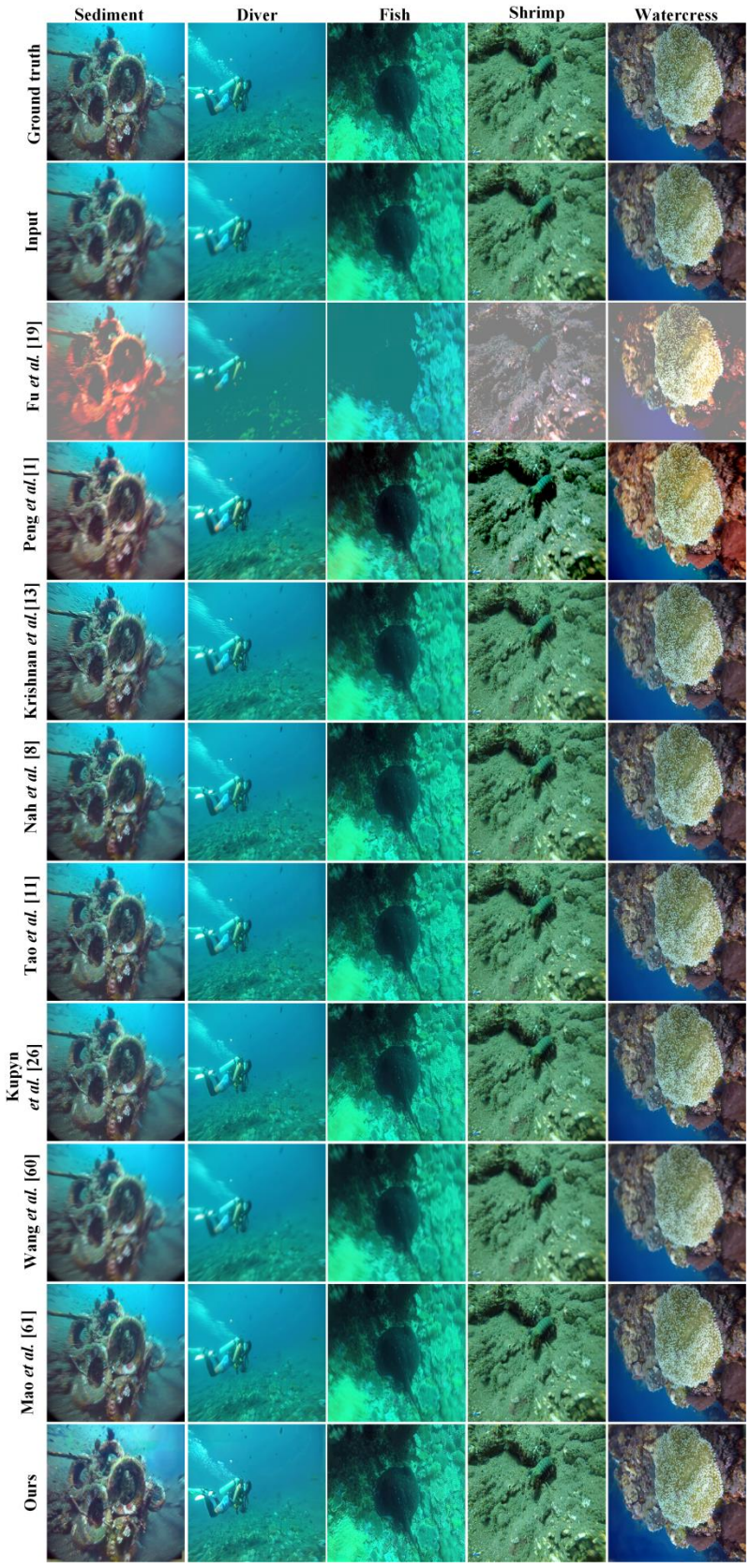

Fig. 8 Qualitative experimental results of different comparison approaches on the UMADD validation set.

4.4 Sea trial dataset

In the sea trial scenario, a Gopro 8 Hero Black was fixed on the bottom of our AUV. The objects in our captured images are sediments, stones, and the marine life such as the starfishes and crabs. The frame rate is 240 fps, and our AUV is powered by the onboard battery and the propellers. The sea trial dataset contains 25 real-world blurry underwater images and the image images are resized to 720×540 resolution. Typical examples from the sea trial dataset and the results of different comparison methods are displayed in Fig. 9. The qualitative results of the comparison methods are consistent with their qualitative results on the UCDD and the UMADD validation sets. In the meanwhile, we have evaluated the performance of different methods using the non-reference metrics on the sea trial dataset, and the average score of each metric is shown in Table 4. Our proposed method still ranks the first in three of six non-reference metrics on the sea trial dataset, this is contributed to its excellent performance in removing blur in the underwater images. We notice that the images of the sea trial dataset suffer from color degrade and image fogging. Peng's method [1] has an effect on these issues, while other methods show a limited ability in solving these issues. Thus, we conduct an image post-processing using an advanced underwater image enhancement method, which is introduced in Sec. 5.

Table 4 Quantitative experimental results of different comparison approaches on the the sea trial dataset using non-reference metrics. The values indicate the average scores of the images[a, b].

Methods	BRISQUE*	NIQE*	PCQI	UIConM	UISM	UIQM
Fu *et al.* [19]	40.371(6)	4.497(8)	**11077(1)**	0.347(9)	2.066(9)	1.925(7)
Peng *et al.* [1]	33.453(5)	**2.898(1)**	8739(9)	0.575(6)	3.077(7)	2.534(5)
Krishnan *et al.* [13]	25.336(2)	3.042(2)	10077(6)	0.700(2)	3.791(4)	2.973(3)
Nah *et al.* [8]	42.383(8)	3.789(6)	9738(8)	0.525(7)	2.395(8)	1.904(8)
Tao *et al.* [11]	41.635(7)	3.609(5)	9910(7)	0.618(4)	3.106(6)	2.478(6)
Kupyn *et al.* [26]	29.036(3)	3.891(7)	10615(2)	0.665(3)	3.652(5)	2.801(4)
Wang *et al.* [60]	54.567(9)	5.828(9)	10181(5)	0.423(8)	6.345(2)	1.226(9)
Mao *et al.* [61]	32.046(4)	3.295(3)	10191(4)	0.615(5)	**6.660(1)**	3.513(2)
Ours	**24.237(1)**	3.347(4)	10411(3)	**0.824(1)**	4.781(3)	**3.701(1)**

[a]The values in bold represents the best results.
[b]The number in brackets refers to the ranking 1-9 of a method on the metric.

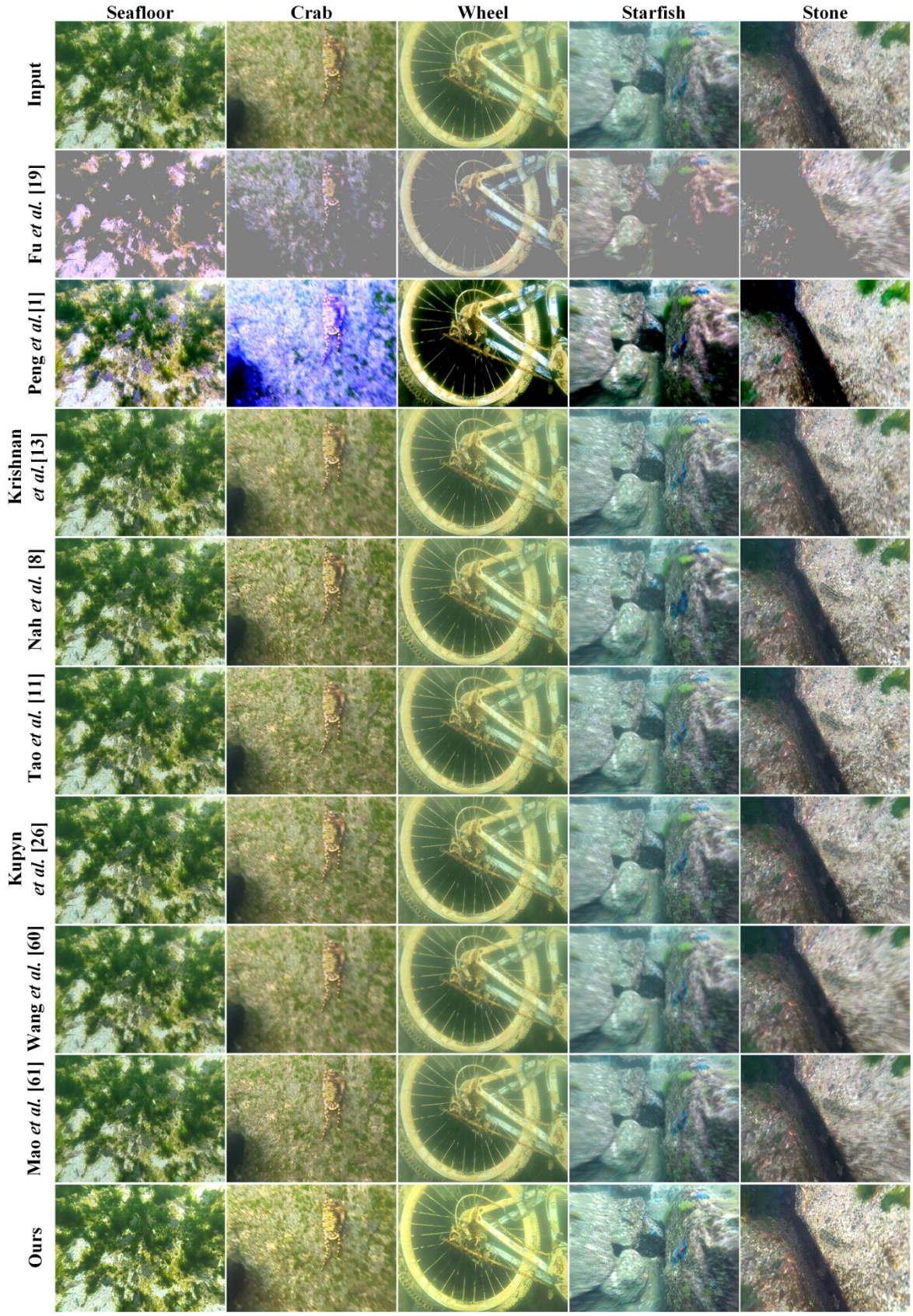

Fig. 9 Qualitative experimental results of different comparison approaches on the the sea trial dataset.

4.5 Efficiency test

We also report the processing time of different methods on the sea trial dataset. All the experiments are conducted using the facility mentioned in Sec. 3. The results of the average testing time for 25 images on the sea trial dataset are as shown in Table 5. Fu's method [19] is the most efficient one in processing a blurry image. Our proposed method ranks the fifth among the nine different methods in restoring an image, and outperforms Peng's method [1], Krishnan's method [13], Tao's method [11], and Kupyn's method [26]. The methods of Wang *et al*. [60], Mao *et al*. [61], and Nah *et al*. [8] process an image in an average time of less than one second. As for the conventional methods, Krishnan's method [13] and Peng's method [1] are very time-consuming. The computational efficiency of deep learning algorithms is generally higher than that of traditional methods according to the above evaluation.

Table 5 The average processing time of different methods for an image in the sea trial dataset[a, b].

Methods	**Fu *et al.* [19]**	**Peng *et al.* [1]**	**Krishnan *et al.* [13]**	**Nah *et al.* [8]**	**Tao *et al.* [11]**	**Kupyn *et al.* [26]**	**Wang *et al.* [60]**	**Mao *et al.* [61]**	**Ours**
Time (s)	**0.70 (1)**	37.76 (8)	57.72 (9)	0.99(3)	5.47(7)	4.304(6)	0.76(2)	0.99(3)	3.93(5)

[a]The values in bold represents the best results.
[b]The number in brackets refers to the ranking 1-9 of a method on the metric.

5 Post-processing

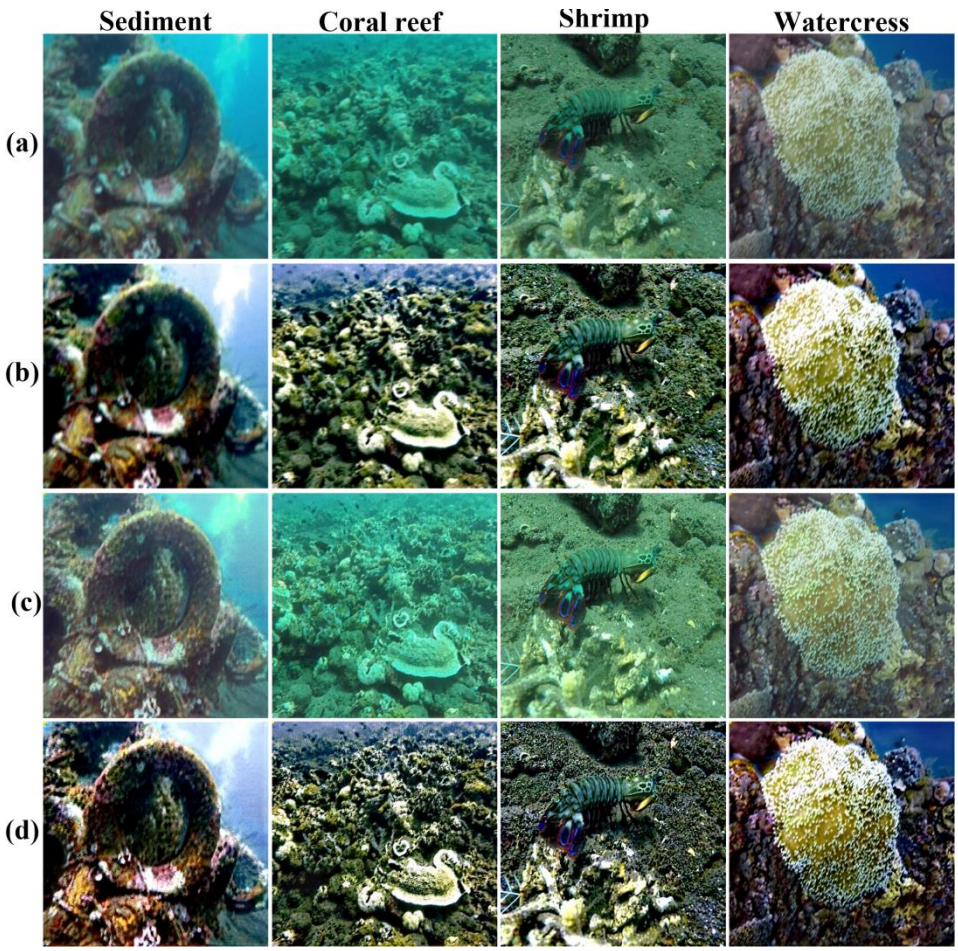

Fig.10 Typical experimental results on UCDD validation set. (a) the input images; (b) the results of processing the input images using the color restoration method; (c) the results of processing the input images using our proposed method; (d) the results of processing the deblurring images.

The proposed underwater image deblurring framework can significantly improve the sharpness of the underwater images. However, the images still suffer from the inherent color distortion. A CNN-based method cannot well-handle the blur effects and color distortion at the same time. Thus, we employ our own color restoration method [57] to address the color distortion issue and generate images with higher quality. As is shown in Figs. 10 and 11, the image quality is greatly improved using our proposed method and the post-processing approach.

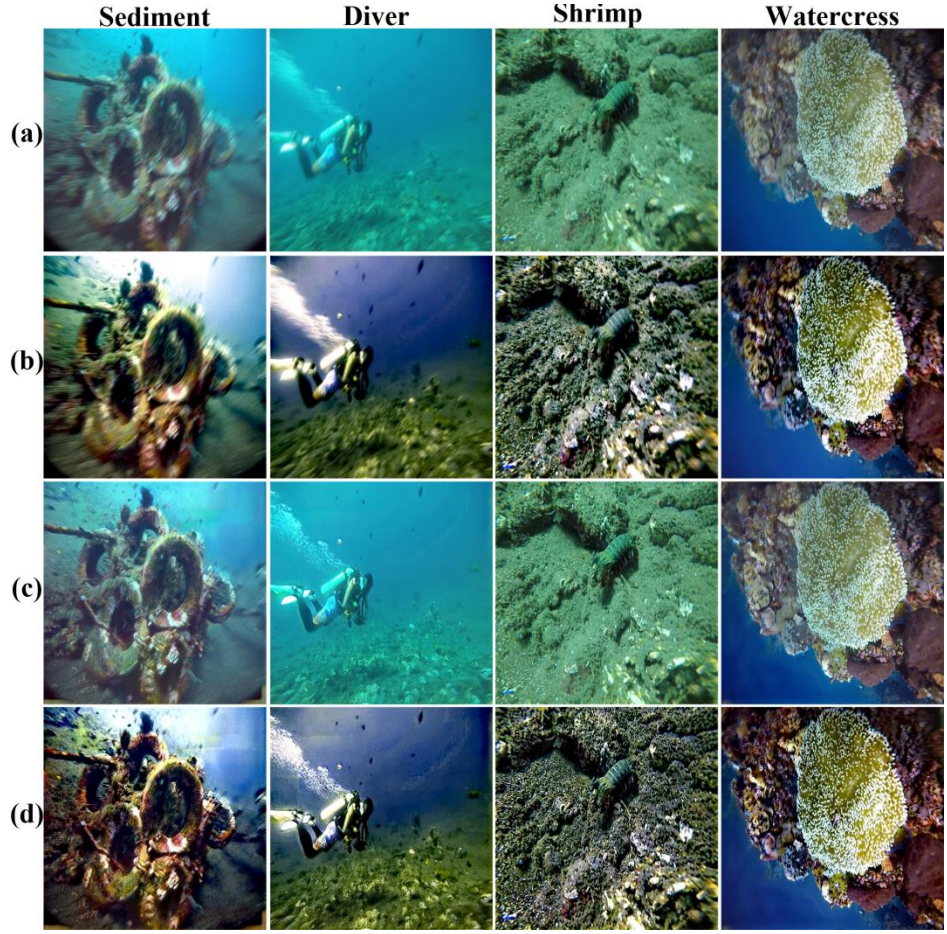

Fig.11 Typical experimental results on UMADD validation set. (a) the input images; (b) the results of processing the input images using the color restoration method; (c) the results of processing the input images using our proposed method; (d) the results of processing the deblurring images.

6 Conclusion and Future Work

In this paper, we proposed an end-to-end deep learning-based approach FPAN to remove the underwater motion blur. By combining the FPN structure with the attention mechanism, FPAN demonstrates clearly superior perceptual quality in removing the blur and restoring the brightness in underwater images. Moreover, due to the lack of publicly available dataset for training the deep deblurring networks, we provide two large-scale underwater deblurring datasets, namely UCDD and UMADD. The proposed method is verified on the validation sets and the sea trial dataset. Qualitative and quantitative experimental results show the effectiveness and robustness

of our proposed method. The proposed method is not only suitable for removing the motion blur, but also has a strong ability to restore the brightness for underwater images.

Our proposed method achieves satisfactory results, however, there are still some limitations. Firstly, our proposed method cannot meet the real-time requirement, hence, it cannot be applied to real-time applications carried out by AUVs. Secondly, unexpected artifacts might appear as mentioned in this paper, this is because the model parameter tuning regarding the water environment requires further optimization. We will make improvements in the future.

Funding

This work was supported by National Natural Science Foundation of China (Grant No. 62001443) and Natural Science Foundation of Shandong province (Grant No. ZR2020QE294).

Acknowledgments

The authors thank the divers for helping collect the valuable underwater images, especially to Yang Yang. They express their gratitude to the team of Underwater Vehicle Laboratory (UVL) for generous help in conducting the sea trial experiments. They also thank Junlin Liu for helping generate the UCDD dataset and Qianqian Yang for discussing the deep network. Additionally, Tengyue Li thanks the China Scholarship Council (CSC) and the Ocean University of China Scholarship for supporting him to conduct his research in the field of image processing at University of Leicester of United Kingdom.

Disclosures

The authors declare no conflicts of interest.

References

1. Y. Peng and P. Cosman, "Underwater image restoration based on image blurriness and light absorption," IEEE Trans. Image Process. 26, 1579-1594 (2017).
2. L. Chen *et al.*,"Perceptual underwater image enhancement with deep learning and physical priors," IEEE Trans. Circ. Syst. Vid. 31, 3078-3092 (2021).
3. J. Wu and X. Di, "Integrating neural networks into the blind deblurring framework to compete with the end-to-end learning-based methods," IEEE Trans. Image Process. 29, 6841-6851 (2020).

4. G. Boracchi and A. Foi, “Modeling the Performance of Image Restoration From Motion Blur,” IEEE Trans. Image Process. 21, 3502-3517 (2012).
5. R. Fergus *et al.*, “Removing camera shake from a single photograph.” ACM Trans. Graphic, 25, 787-794 (2006).
6. M. Tico and M. Vehvilainen, “Estimation of motion blur point spread function from differently exposed image frames”, in Proceedings of the 14th European Signal Processing Conference, pp. 1-4 (2006).
7. Y. Yitzhaky *et al.*, “Direct method for restoration of motion-blurred images,” J. Opt.Soc. Am. A, 15, 1512-1519 (1998).
8. S. Nah *et al.*, “Deep Multi-scale convolutional neural network for dynamic scene deblurring,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 257-265 (2017).
9. O. Kupyn *et al.*, “DeblurGAN: Blind motion deblurring using conditional adversarial networks,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8183-8192 (2018).
10. T. Nimisha *et al.*, “Blur-invariant deep learning for blind-deblurring,” in Proceedings of the IEEE Conference on International Conference on Computer Vision, pp. 4762-4770 (2017).
11. X. Tao *et al.*, “Scale-recurrent network for deep image deblurring,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8174-8182 (2018).
12. Y. Kageyama *et al.*, “ProDebNet: projector deblurring using convolutional neural network,” Opt. Express, 28, 20391-20403 (2020).
13. D. Krishnan *et al.*, “Blind deconvolution using a normalized sparsity measure,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 233-240 (2011).
14. J. Pan *et al.*, “Blind image deblurring using dark channel prior,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1628-1636 (2016).
15. R. Yan and L. Shao, “Blind image blur estimation via deep learning,” IEEE Trans. Image Process. 25, 1910-1921 (2016).
16. D. Ren *et al.*, “Neural blind deconvolution using deep priors,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3341-3350 (2020).
17. J. Pan *et al.*, “Cascaded deep video deblurring using temporal sharpness prior,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3043-3051 (2020).
18. J. Sun *et al.*, “Learning a convolutional neural network for non-uniform motion blur removal,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 769-777 (2015).
19. X. Fu *et al.*, “Two-step approach for single underwater image enhancement,” in Proceedings of IEEE International Symposium on Intelligent Signal Processing and Communication Systems, pp. 789-794 (2017).
20. Z. Li *et al.*, “Learning to see through turbulent water,” in Proceedings of the IEEE Winter Conference on Applications of Computer Vision, pp. 512-520 (2018).
21. P. Agrawal *et al.*, “Learning to see by moving,” in Proceedings of the IEEE Conference on International Conference on Computer Vision, pp. 37-45 (2015).
22. A. Kanazawa *et al.*, “Warpnet: Weakly supervised matching for single-view reconstruction,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3253-3261 (2016).
23. M. Jaderberg *et al.*, “Spatial transformer networks,” in Advances in Neural Inform. Process. System., 28, 2017-2025 (2015).
24. X. Yu *et al.*, “Deep deformation network for object landmark localization,” in Proceedings of the IEEE Conference on European Conference on Computer Vision, pp. 52-70 (2016).
25. M. Arjovsky *et al.*, “Wasserstein gan,” arXiv preprint arXiv:1701.07875, pp. 1-32 (2017).
26. O. Kupyn *et al.*, “DeblurGAN-v2: Deblurring (orders-of-magnitude) faster and better,” in Proceedings of the IEEE Conference on International Conference on Computer Vision, pp. 8878-8887 (2019).

27. C. Schuler *et al.*, "Learning to deblur," IEEE Trans. Pattern Anal. Mach. Intell. 38, 1439-1451 (2016).
28. A. Chakrabarti. "A neural approach to blind motion deblurring," in Proceedings of IEEE Conference on European Conference on Computer Vision, pp. 221-235 (2016).
29. L. Sun *et al.*, "Edge-based blur kernel estimation using patch priors," in Proceedings of IEEE International Conference on Computational Photography, pp. 1-8 (2013).
30. Y. Mei *et al.*, "Pyramid attention networks for image restoration," arXiv:2004.13824, pp. 1-19 (2020).
31. R. Ranjbarzadeh *et al.*, "Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi-modalities brain images," Scientific Reports, 11, 1-17 (2021).
32. Y. Li *et al.*, "Occlusion aware facial expression recognition using cnn with attention mechanism," IEEE Trans. Image Process. 28, 2439-2450 (2018).
33. J. Li *et al.*, "Attention mechanism-based CNN for facial expression recognition," Neurocomputing, 411, 340-350 (2020).
34. T. Lin *et al.*, "Feature pyramid networks for object detection," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 2117-2125 (2017).
35. H. Zhang *et al.*, "Self-attention generative adversarial networks," in Proceedings of International Conference on Machine Learning, pp. 7354-7363 (2019).
36. T. Li *et al.*, "Distorted underwater image reconstruction for an autonomous underwater vehicle based on a self-attention generative adversarial network," Appl. Opt. 59, 10049-10060 (2020).
37. P. Isola *et al.*, "Image-to-image translation with conditional adversarial networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1125-1134 (2017).
38. C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4681-4690 (2017).
39. S. Woo *et al.*, "CBAM: convolutional block attention module," in Proceedings of the IEEE Conference on European Conference on Computer Vision, pp.1-17 (2018).
40. R. Liu *et al.*, "Real-world underwater enhancement: challenges, benchmarks, and solutions under natural light," IEEE Trans. Circ. Syst. Vid. 30, 4861-4875 (2020).
41. C. Li *et al.*, "An underwater image enhancement benchmark dataset and beyond," IEEE Trans. Image Process. 29, 4376-4389 (2020).
42. M. Islam *et al.*, "Fast underwater image enhancement for improved visual perception," IEEE Robot. Autom. Lett. 5, 3227-3234 (2020).
43. S. Su *et al.*, "Deep Video Deblurring for Hand-Held Cameras," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 237-246 (2017).
44. S. Nah *et al.*, "NTIRE 2019 challenge on video deblurring and super-resolution: Dataset and study," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 1-10 (2019).
45. S. Nah *et al.*, "NTIRE 2021 challenge on image deblurring," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 149-165 (2021).
46. P. Wieschollek *et al.*, "Learning blind motion deblurring," in Proceedings of the IEEE Conference on International Conference on Computer Vision, pp. 231-240 (2017).
47. S. Niklaus *et al.*, "Video frame interpolation via adaptive separable convolution," in Proceedings of the IEEE Conference on International Conference on Computer Vision, pp. 261-270 (2017).
48. Github: https://github.com/.
49. MATLAB: https://www.mathworks.com/.
50. PyTorch: https://pytorch.org/.
51. Z. Wang *et al.*, "Image quality assessment: from error visibility to structural similarity," IEEE Trans. Image Process. 13, 600-612 (2004).
52. Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," Electron. Lett., 44, 800-801 (2008).
53. A. Mittal *et al.*, "No-reference image quality assessment in the spatial domain," IEEE Trans. Image Process. 21, 4695-4708 (2012).

54. A. Mittal *et al.*, “Making a “completely blind” image quality analyzer,” IEEE Signal Process. Lett. 20, 209-212 (2012).
55. S. Wang et al., “A patch-structure representation method for quality assessment of contrast changed images,” IEEE Signal Process. Let. 22, 2387-2390 (2015).
56. K. Panetta *et al.*, “Human-visual-system-inspired underwater image quality measures,” IEEE J. Oceanic Eng. 41, 541-551 (2015).
57. T. Li *et al.*, “Underwater image enhancement using adaptive color restoration and dehazing,” Opt. Express. 30, 6216-6235 (2022).
58. J. Xie *et al.*, “A variational framework for underwater image dehazing and deblurring,” IEEE Trans. Circ. Syst. Vid., 1-14 (2021).
59. E. Park *et al.*, “Underwater image restoration using geodesic color distance and complete image formation model,” IEEE Access., 8, 157918-157930 (2020).
60. Z. Wang *et al.*, “Uformer: A general U-shaped transformer for image restoration,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-17 (2022).
61. X. Mao *et al.*, “Deep residual fourier transformation for single image deblurring,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-10 (2021).
62. C. Li *et al.*, “Underwater scene prior inspired deep underwater image and video enhancement,” Pattern Recognit., 98, 1-11 (2020).
63. G. Hou *et al.*, “Benchmarking underwater image enhancement and restoration, and beyond”, IEEE Access., 8, 122078-122091 (2020).
64. X. Li *et al.*, “Enhancing underwater image via adaptive color and contrast enhancement, and denoising”, Eng. Appl. Artif. Intel., 11, 1-14 (2022).
65. C. Li *et al.*, “Underwater image enhancement via medium transmission-guided multi-color space embedding,” IEEE Trans. Image Process., 30, 4985-5000 (2021).
66. G. Hou, *et al.*, “An efficient nonlocal variational method with application to underwater image restoration”, Neurocomputing, 369, 106-121 (2019).

Tengyue Li is a PhD candidate at Ocean University of China (OUC). He received his BS degree in electronic information science and technology from OUC in 2013, and received his MS degree in optical engineering from OUC in 2015. In 2020, he was funded by the China Scholarship Council (CSC) and the Ocean University of China Scholarship to conduct his research in the field of image processing at University of Leicester of UK. His current research interests include underwater image processing, underwater 3-D reconstruction, underwater object recognition, and underwater robotics technology.

Shenghui Rong is a lecturer at School of Information Science and Engineering, Ocean University of China. He received his BS degree in electronic science and technology from Xidian University in 2011 and received his PhD in physical electronics from Xidian University in 2018. In 2016, he was funded by the the China Scholarship Council to conduct his research in the direction of 3-D image processing and recognition at the Griffith University in Australia. His current research interests include optoelectronic countermeasures, computer vision, and pattern recognition.

Long Chen received his BS degree from Northeast Normal University in 2013 and his MS degree in computer architecture at the Vision Lab of Ocean University of China. He is currently pursuing his PhD in the School of Informatics, University of Leicester of UK. His research interests are in the areas of computer vision and machine learning.

Huiyu Zhou is currently a full professor with the School of Informatics, University of Leicester, UK. He received his BS degree from Huazhong University of Science and Technology and MS degree from University of Dundee of UK. He was awarded a PhD degree in Computer Vision from Heriot-Watt University of UK. His research work has been or is being supported by the U.K. EPSRC, MRC, EU, Royal Society, Leverhulme Trust, Puffin Trust, Invest NI, and industry. His current research interests include image analysis and machine learning.

Bo He is a full professor of Ocean University of China (OUC). He received his MS and PhD degrees from Harbin Institute of Technology in 1996 and 1999, respectively. From 2000 to 2003, he worked in Nanyang Technological University, Singapore, as a postdoctoral fellow and his research work focused on the precise navigation, control, and communication for the platform of mobile robots and unmanned vehicles. His current research interests include image analysis, AUV design and applications, and machine learning.