

Collaborative Underwater Image Enhancement and Super-Resolution with Multi-stage Mutual Transmission of Associated Features

Kunqian Li^a, Chunyan Li^a, Zhihao Huang^a, Shaobao Hu^a, Guojia Hou^b, Dalei Song^{a,c,*}

^a*College of Engineering, Ocean University of China, Qingdao 266100, China.*

^b*College of Computer Science and Technology, Qingdao University, Qingdao 266071, China.*

^c*Institute for Advanced Ocean Study, Ocean University of China, Qingdao 266100, China*

Abstract

In underwater scenes, dynamically changing underwater imaging conditions and inevitable underwater imaging degradation severely degrade the quality of underwater images. In addition, due to bandwidth limitation in underwater communication, underwater images often have to be compressed at the cost of reduced resolution. Therefore, robust underwater image enhancement (UIE) and super-resolution (SR) techniques are urgently needed to improve the visual quality and restore the resolution of degraded underwater images. Considering that both UIE and SR are essentially image-to-image tasks oriented to the degraded images, we propose a unified end-to-end deep network, i.e., CoESR-Net, for joint underwater image enhancement and super-resolution. Specifically, to fully exploit the supervision information of scene-associated images, we introduce a multi-stage mutual transmission strategy of associated features for collaborative learning. Moreover, to prepare better and robust feature transformation for the UIE and SR networks, we introduce encoder-decoder structures into the fusion-based enhancement module. Both quantitative/qualitative experiments and ablation study conducted on the widely used datasets, including UFO-120, UIEB, SQUID, Color-Check7 and USR-248, demonstrate the effectiveness of the proposed new framework and key modules.

Keywords: Underwater image enhancement, underwater image super-resolution, deep learning, convolutional neural network, collaborative learning.

1. Introduction

Recently, autonomous underwater vehicles (AUV) and remote operated vehicles (ROV), the main underwater observation platforms, have developed rapidly [38]. To get underwater vision data, most underwater vehicles have been equipped with high-performance cameras [42]. As one of the main sources of underwater information, vision can provide intuitive observations [20] and guidance for underwater operation [44, 5, 48, 35]. Good visual quality and high image resolution are increasingly becoming important requirements for image acquisition modules of underwater vehicle.

However, underwater imaging conditions are dynamically changing and much more complicated than that in the air. Moreover, due to wavelength-dependent absorption, forward scattering, backward scattering, turbulence disturbing, and suspended sediment, the degradation of underwater images is usually inevitable. Therefore, underwater image enhancement (UIE) [32], which is aimed at improving the visual quality of underwater images, has received a lot of attention recently. The

quality of the enhancement even determines the success or failure of subsequent underwater applications.

In addition, although most underwater vehicles are equipped with high-performance cameras, the transmission of captured image data is still a thorny issue. Limited by the high cost of cableless data transmission, transmission of high-resolution images is extremely difficult in long-range underwater missions. Sending low-resolution images, followed by super-resolution (SR) in the decode stage for subsequent missions is an effective compromise. Therefore, both underwater image enhancement and underwater image super-resolution are crucial for underwater observation and operation.

Unfortunately, as both low-level image processing tasks, although underwater image enhancement and single image super-resolution have been widely studied separately [53, 2, 55], only few studies have explored them as a whole end-to-end task [11, 9, 8, 17]. The two tasks are identical in their goal of improving image visual quality, and the enhancement task can help the super-resolution to better capture the degraded information of underwater images.

Moreover, we also noticed that the majority of existing underwater image enhancement algorithms are designed for isolated single images [39] and few researchers have addressed the task of collaborative super-resolution of non-overlapping images from correlated scenes. Even for the few joint enhancement and super-resolution algorithms that are available [11, 9, 8, 17], the algorithm input is also isolated single im-

*Corresponding author

Email addresses: likunqian@ouc.edu.cn (Kunqian Li),
lichunyan6767@stu.ouc.edu.cn (Chunyan Li),
huangzhihao@stu.ouc.edu.cn (Zhihao Huang),
hushaobao@stu.ouc.edu.cn (Shaobao Hu), hgjouc@126.com (Guojia Hou), songdalei@ouc.edu.cn (Dalei Song)

age. Considering that images acquired from correlated scenes contain a large amount of complementary information, a collaborative processing approach is more conducive to learning a unified, effective deep model with high visual unity and generalizability.

In this paper, we propose a collaborative underwater image enhancement and super-resolution method for images from correlated scenes. To give a brief overview, we summarize our contribution as follows:

- 1) We propose a unified end-to-end deep model for both underwater image enhancement and super-resolution, namely CoESR-Net, which can simultaneously and collaboratively process paired images from associated scenarios.
- 2) In the fusion-based enhancement module, we introduce encoder-decoder structures to perform consistent feature transformation and reconstruction for consistent enhancement of underwater images from related scenes.
- 3) We introduce a multi-stage mutual transmission strategy of associated features, which establishes links for feature communication of paired images from related scenes and helps to improve the visual unity of final results and model generalizability.

2. Related Works

2.1. Underwater Image Enhancement with Deep Learning

In recent decades, great developments have been made with the underwater image enhancement problem [53, 2, 64, 61, 52]. However, due to the complicated underwater environment and diverse imaging interference factors, underwater image enhancement in real-world applications is still an unsolved problem. Although deep learning has been highly successful in many vision tasks, it is extremely limited by the scarcity of training images with reference in underwater image enhancement task.

To facilitate the study of underwater image enhancement algorithms based on deep learning framework, researchers synthesized underwater images with random pixel disruption [49], milk-noise simulation [43], generative adversarial network (GAN) [13, 19], underwater imaging models and optical properties of underwater scenes [3]. These training datasets created convenient conditions for deep learning based underwater image enhancement and a group of deep models are proposed, which can ideally be divided into two main categories[2], i.e., CNN-based [49, 3, 43, 24] and GAN-based algorithms [13, 27, 57, 34, 56, 19]. But, there still remains a clear gap between the synthesized images and real-world underwater images. The synthetic samples can hardly describe the complex and diverse degradation in real underwater scenarios.

More recently, Li et al. [25] and Qi et al. [39] have successively established two real-world underwater image datasets with enhancement reference obtained by manual voting. To some extent, they contribute to the image enhancement quality of real underwater scenes. However, such training samples

are extremely costly to obtain and small in quantity, and how to utilize and mine their information more effectively is an important research direction for the future. To fully mine the implied common information of scene associated images, Qi et al. [39] initially explored the joint learning strategy for underwater image enhancement, while the research on collaborative processing mechanism for multi-stage and multi-task scenarios is still a gap.

2.2. Image Super-Resolution with Deep Learning

Single image super-resolution [55], thanks to the rapid development of deep learning, has also made great strides in the past few years. In deep learning-based framework, super-resolution task is abstracted as learning a mapping model from low-resolution images to high-resolution outputs. Depending on the learning strategy, existing SR models can also be broadly classified into two categories, namely CNN-based models and GAN-based models. CNN-based models, such as SRCNN [12], DSRCNN [36], SRRResNet [55] and SRDRM [18], are consisted of a sequence of non-linear filters which are learned from a large number of training images. On the other hand, for GAN-based models, such as SRGAN [22], ESRGAN [47], EDSRGAN [33], a couple of generator and discriminator are iteratively learned by playing a two-player min-max game, so as to generate high-resolution images that look like the real ones. Although GANs usually produce results with high visual quality, they do not help much in recovering ground-truth information.

More recently, many researchers have noticed the importance of underwater image super-resolution in practical underwater exploration [18, 7, 45]. Islam et al. [18] employed deep residual multiplier as basic super-resolution block to from deep underwater SR model, namely SRDRM. Chen et al. [7] introduced a progressive training strategy for their CNN-based SR model with channel-wise attention module. Wang et al. [45] proposed lightweight underwater SR model, whose number of parameters is effectively reduced with their information distillation modules.

In addition to single image super-resolution, multiple-image super-resolution has also received the attention of researchers [51, 21, 16]. However, these studies mainly focused on the super-resolution of overlapped low-resolution images, such as multiple remote sensing observations for the same area [21, 16]. The super-resolution of non-overlapping images from correlated scenes, which can provide collaborative constraints and improve the consistency of enhancements, is seldom studied.

2.3. Collaborative Strategy in Computer Vision

In many typical image-based computer vision tasks, the information that a single image can provide is often quite limited and ambiguous. Then, researchers have begun to explore whether multiple images can be processed simultaneously in a collaborative manner to improve the result for each of them and eliminate errors introduced by such ambiguity.

The earliest explorations of collaborative strategy appeared in image segmentation [41, 63], where the so called image co-segmentation task aims at mining common foreground targets in multiple images, using complementary information to

Table 1: Annotations of the domain mappings in underwater image enhancement and super-resolution task.

Source domain	Target domain	Mapping annotation
low-resolution distorted	high-resolution enhanced	$\mathcal{X} \rightarrow \mathcal{Y}$
low-resolution distorted	low-resolution enhanced	$\mathcal{X} \rightarrow \mathcal{X}_e$
low-resolution enhanced	high-resolution enhanced	$\mathcal{X}_e \rightarrow \mathcal{Y}$
high-resolution distorted	high-resolution enhanced	$\mathcal{Y}_d \rightarrow \mathcal{Y}$

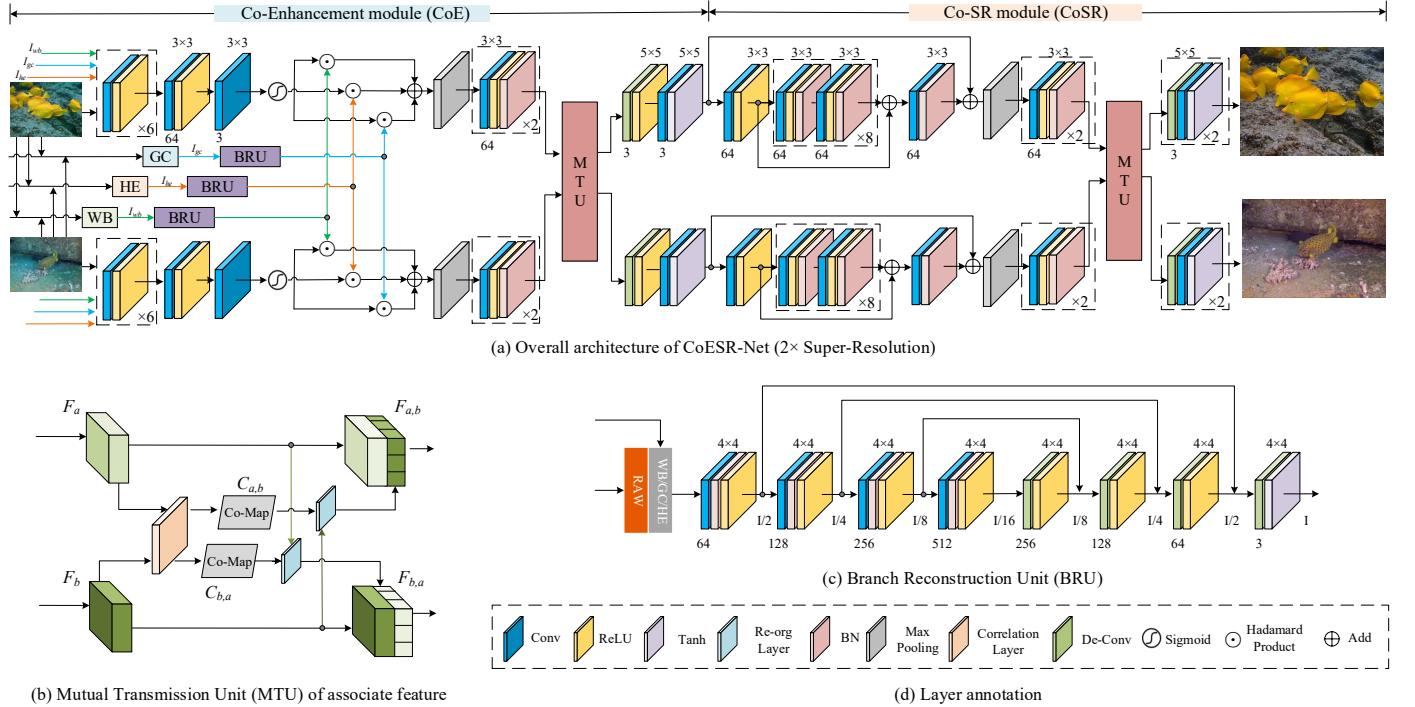


Figure 1: The overall structure diagram of the proposed collaborative underwater image enhancement and super-resolution network (CoESR-Net).

suppress the background and obtain more accurate target segmentation [30, 29]. Thereafter, there have been many studies on collaborative strategy in related fields, such as video co-segmentation [59, 46], co-saliency detection [28, 58] and object co-localization [50]. The above-mentioned collaborative strategy that does not resort to additional supervised information but only exploits the implicit commonality constraints between images is also considered as a weakly supervised or semi-supervised strategy. Beyond that, within the fully supervised learning framework, collaborative strategies have also attracted the attention of researchers, which are often served as feature sharing channels between multiple images to provide additional information [31, 6, 23, 39]. While, in multi-task integrated framework, especially for multi-stage tasks such as high-magnification image super-resolution, collaborative strategy has never been fully explored.

3. Problem Formulation and Proposed Deep Model

3.1. Problem Formulation

Joint underwater image enhancement and super-resolution task aims to improve the visual quality of underwater im-

ages while increasing their resolution, i.e., generating high-resolution enhanced (HRE) underwater images from low-resolution distorted (LRD) underwater images. We formulate this problem as learning an end-to-end mapping G from the LRD domain \mathcal{X} to the HRE domain \mathcal{Y} . For ease of presentation, the global mapping can be further divided into several sub-mappings as Table 1 presented. Moreover, we explore 2 \times , 4 \times and 8 \times super-resolution in this paper.

3.2. CoESR-Net

To solve the formulated problem mentioned above, we propose the CoESR-Net, which is a unified end-to-end deep network that can simultaneously and collaboratively process paired images from association scenarios. The overall structure diagram of the proposed CoESR-Net is given in Figure 1. Unlike traditional enhancement and super-resolution pipelines, which only handle one image at a time in both training and inferring stage, the CoESR-Net is organized as a Siamese structures by introducing multi-stage mutual feature transmission, so as to fully exploit the collaborative information and constraints from scene-associated images. As a generative model, CoESR-Net consists of Co-Enhancement module (CoE) and

Co-Super-Resolution module (CoSR). From the perspective of domain transformation, the proposed CoESR-Net sequentially performs transformation $X \rightarrow X_e$ with the CoE module and transformation $X_e \rightarrow Y$ with the CoSR module as Figure 1 (a) presented. Such multi-task learning and end-to-end trainable network helps break down the barriers between the two tasks, which manifests itself as reducing the incompatibility of feature expression in the task transition. Section 4 will give more analysis and comprehensive demonstration on this issue. The following subsections of this part will give more details about each module of CoESR-Net.

3.2.1. Mutual Transmission Unit

In a two-branch collaborative framework, an effective information transmission path between the two branches helps gather shared high-level associated features and provide additional constraints for the learning of recovering mappings. In CoESR-Net, we follow [39] to introduce a correlation layer-based mutual transmission unit (MTU) for the mining and reorganization of cross-branch associated features. As presented in Figure 1(b), MTU takes the features of two different but scenario-associated images, i.e., F_a and F_b , as the two-branch inputs, which are independently generated from previous convolutional feature transforming branch. Then, the correlation layer is applied to these two deep feature maps to alternately generate two mutual correlation maps (Co-Map), i.e. $C_{a,b}$ and $C_{b,a}$, which show how similar the feature on each location is to each of the features from another branch. With the Co-Maps, for the feature maps on each branch, for example, the F_a with $h \times w \times d$ -dimension elements, we will split F_b and reorganize it as a $h \times w \times d$ collaborative feature map to catenate F_a , while ensuring that F_a catenate the most associated d -dimension feature from F_b on each position. Thus, after the mutual transmission of MTU, the size of output feature map from each branch becomes $h \times w \times 2d$, which effectively complements information from synergistic branch for the subsequent collaborative enhancement and super-resolution.

3.2.2. Co-enhancement Module

In this paper, we propose a co-enhancement module integrating a two-stage feature transform and fusion mechanism, where the first-stage aims to perform feature transformation and fusion between different reconstructed feature components of the same underwater image, and the second-stage is an MTU-based feature transform and fusion of the correlated features between different images from related scenarios.

As presented in Figure 1(a), the sub-branch of the first-stage feature transform and fusion mechanism mainly follows the backbone of Water-Net [25], which is a simple but effective fusion-based generative structure for clear underwater image feature recovery. Its main-branch consists of a set of convolutional and activation layers. It is used to transform the combined raw image I_{raw} and rough refinements I_{gc} , I_{he} and I_{wb} , achieved by gamma correction (GC), histogram equalization (HE) and white balance (WB), respectively, to confidence map C_{gc} , C_{he} and C_{wb} for weighted sub-branch fusion. While, the original structure of Water-Net, whose three sub-branches, i.e.,

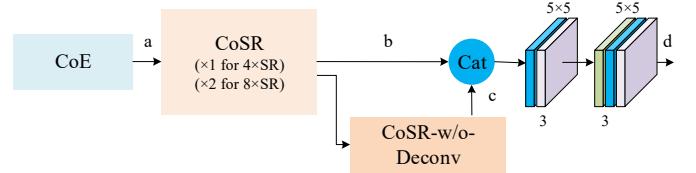


Figure 2: The structure diagram of CoESR-Net with 4 \times or 8 \times super-resolution.

Table 2: The feature size on different nodes of the diagram in Figure 2.

Magnification	a	b	c	d
4 \times	$w \times h$	$2w \times 2h$	$2w \times 2h$	$4w \times 4h$
8 \times	$w \times h$	$4w \times 4h$	$4w \times 4h$	$8w \times 8h$

the image feature transformer unit (FTU) only consists of three conv+ReLU blocks, usually have trouble in generating visually consistent transform for images from similar scenarios. We improve it by designing a robust image feature transformer, which can be also viewed as an image reconstruction unit for preliminary enhancement.

As presented in Figure 1(c), similar to the generator of UGAN [13], we introduce a “U-Net”-like Branch Reconstruction Unit (BRU) to serve as the robust feature transformer. It is built upon an encoder-decoder structure. The BRU consists of four pairs of conv-deconv layers. By setting the convolution stride as 2, the stacked convolutional layers can achieve larger receptive fields and perceive more broad and high-level features. And the deconvolution layers, which are combined with skip connections, can further integrate information from different levels. In this way, the rough refinements for the images from similar scenarios can receive more consistent transformation and thus provide more visually consistent branch information for fusion. Finally, the first-stage fused deep feature can be achieved by adding the three sub-branches with the predicted fusion confidence.

The second stage of feature transform and fusion is established upon the mutual transmission unit. A max pooling and two stacked Conv-ReLU-BN blocks are arranged to prepare the inputs for MTU. While for the outputs of MTU, we set a Deconvolution-ReLU block and a Convolution-Tanh block to recover the size and further fuse the features from two branches.

3.2.3. Co-Super-Resolution Module

As Figure 1 presented, the CoSR module, which is consists of two Siamese branches, directly takes the outputs of CoE module as the inputs for super-resolution. By following [18], each branch of CoSR is organized with 8 short and 1 long residual learning blocks. After that, the features from two branches will be first downsized with a 2 \times 2 max pooling layer independently and then collaboratively receive correlation feature transmission with the MTU. Finally, the two stacked deconvolutional and activation layer blocks will give the feature maps a 4 \times size amplification to realize a 2 \times super-resolution.

As to 4 \times and 8 \times super-resolution tasks, multiple CoSR modules will be stacked as Figure 2 presented. To cover information

from various levels for better deep feature presentation, we keep a short-cut for multi-level feature catenation. In the stacked MTUs, associated features will receive multi-stage comprehensive mutual transmission, which provides conditions for using shared information to achieve detail complementation in co-super-resolution tasks.

3.3. Network Training Loss

Given a collaborative underwater image enhancement and super-resolution network defined as in Figure 1(a), we intend to train a mapping model G , from the LRD domain X to the HRE domain Y , with paired real-world underwater images. As an end-to-end multi-task model, we use a total of three kinds of loss terms upon the outputs of CoE module and CoSR module to provide comprehensive learning constraints, i.e., pixel error loss, perceptual loss and SSIM loss.

3.3.1. Pixel Error Loss

Given model G , raw underwater image X , and its reference Y , the adopted pixel-error loss measures the pixel-wise ℓ_1 distance between the ground-truth reference Y and the model or intermediate outputs of $G(X)$ as follows:

$$L_{\ell_1} = \frac{1}{NM} \sum_{i,j} |Y_{i,j} - G(X)_{i,j}|, \quad (1)$$

where i and j are the pixel coordinates in the images, N and M are the width and length of the input underwater images, respectively. As shown in Figure 3, L_{ℓ_1} is applied to both the outputs of CoE module and CoSR module. When we use it to measure the loss of the outputs of CoE module, the ground-truth reference Y is refer to low-resolution enhanced image. While for the loss of the outputs of CoSR module, the ground-truth reference Y is refer to high-resolution enhanced image accordingly.

3.3.2. Perceptual Loss

To reduce the visual content difference between the output image and the reference, the perceptual loss is defined as

$$L_{Perceptual} = \frac{1}{N^2 M^2} \sum_{i,j} (\Phi_{VGG}(Y)_{i,j} - \Phi_{VGG}(G(X))_{i,j})^2, \quad (2)$$

where $\Phi_{VGG}(Y)$ and $\Phi_{VGG}(G(X))$ mean extracting the deep features from the specified layer (conv4_2) of the VGG19 network with the reference Y and the mapping results $G(X)$ as the inputs, respectively. N and M are the width and length of those feature maps, i and j are the coordinates for the locations in the such feature maps. $L_{Perceptual}$ is only applied to the outputs of CoSR module.

3.3.3. SSIM Loss

To impose extra constraints on the texture and structure similarity between the reference and the final super-resolution output, we also add SSIM loss [60] to the objective function as Figure 3 presented. We define the SSIM loss as follows,

$$L_{SSIM} = 1 - \frac{1}{P} \sum_p \text{SSIM}(p), \quad (3)$$

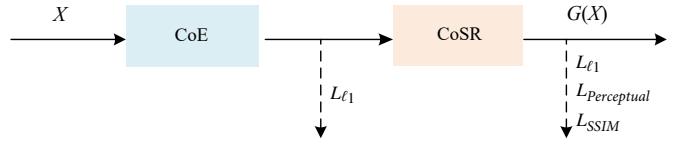


Figure 3: The applied loss terms for the training of CoESR-Net.

$$\text{SSIM}(p) = \frac{2\mu_{G(x)}\mu_y + C_1}{\mu_{G(x)}^2 + \mu_y^2 + C_1} \cdot \frac{2\sigma_{G(x)}\sigma_y + C_2}{\sigma_{G(x)}^2 + \sigma_y^2 + C_2}, \quad (4)$$

where p is defined as a center pixel of an 11×11 image patch and P is the total number of pixels in the image, $\text{SSIM}(p)$ means the SSIM value between patch $p_{G(x)}$ from output image $G(x)$ and patch p_Y from ground-truth image Y , whose center are both on p , $\mu_{G(x)}$ (μ_Y) and $\sigma_{G(x)}$ (σ_Y) are the mean and standard deviation of $p_{G(x)}$ (p_Y), respectively. The constants C_1 and C_2 are set to 0.01 and 0.03 as the default in SSIM definition [62].

Finally, the loss function for CoESR-Net can be formulated as follows:

$$\mathcal{L} = \mathcal{L}_{CoE} + \mathcal{L}_{CoSR}, \quad (5)$$

$$\mathcal{L}_{CoE} = L_{\ell_1}, \quad (6)$$

$$\mathcal{L}_{CoSR} = L_{\ell_1} + \alpha L_{Perceptual} + \beta L_{SSIM}, \quad (7)$$

where the weights α and β are empirically set to 0.05 and 0.75 to balance the scales of different terms and fixed during the training stage on different datasets.

4. Experiments

4.1. Implementation Details and Experimental Setup

We implemented the proposed network using the TensorFlow platform and conducted experiments on an NVIDIA RTX 2080Ti GPU. We take Adam as the optimizer and set the learning rate to 1e-4. Taking both efficiency and memory space into consideration, the training batch size is set to 2. Given the image size of desired outputs as 640×480 , the size of input images are set to 320×240 , 160×120 and 80×60 when we perform $2\times$, $4\times$ and $8\times$ super-resolution, respectively.

We conducted experiments on 5 publicly available datasets, i.e., UFO-120 [17], UIEB [25] USR-248 [18], SQUID [4] and Color-Check7 [1]. UFO-120 dataset is designed for simultaneous enhancement and super-resolution of underwater image, which contains 1500 LRD-HRE paired samples for training and 120 paired samples for test. UIEB dataset including 950 real-world underwater images is originally designed for underwater image enhancement task, 890 of which have the manually selected reference images from multiple enhancement candidates. While, by changing the resolution of the input image, it can also be used for super-resolution task. In our experiments, by randomly selecting, the training, validation and test sets of UIEB contain 690, 100 and 100 images, respectively. USR-248 dataset is only used for the training of underwater single image super-resolution models, which contains 1060 paired samples for training/validation and 248 paired samples for test.

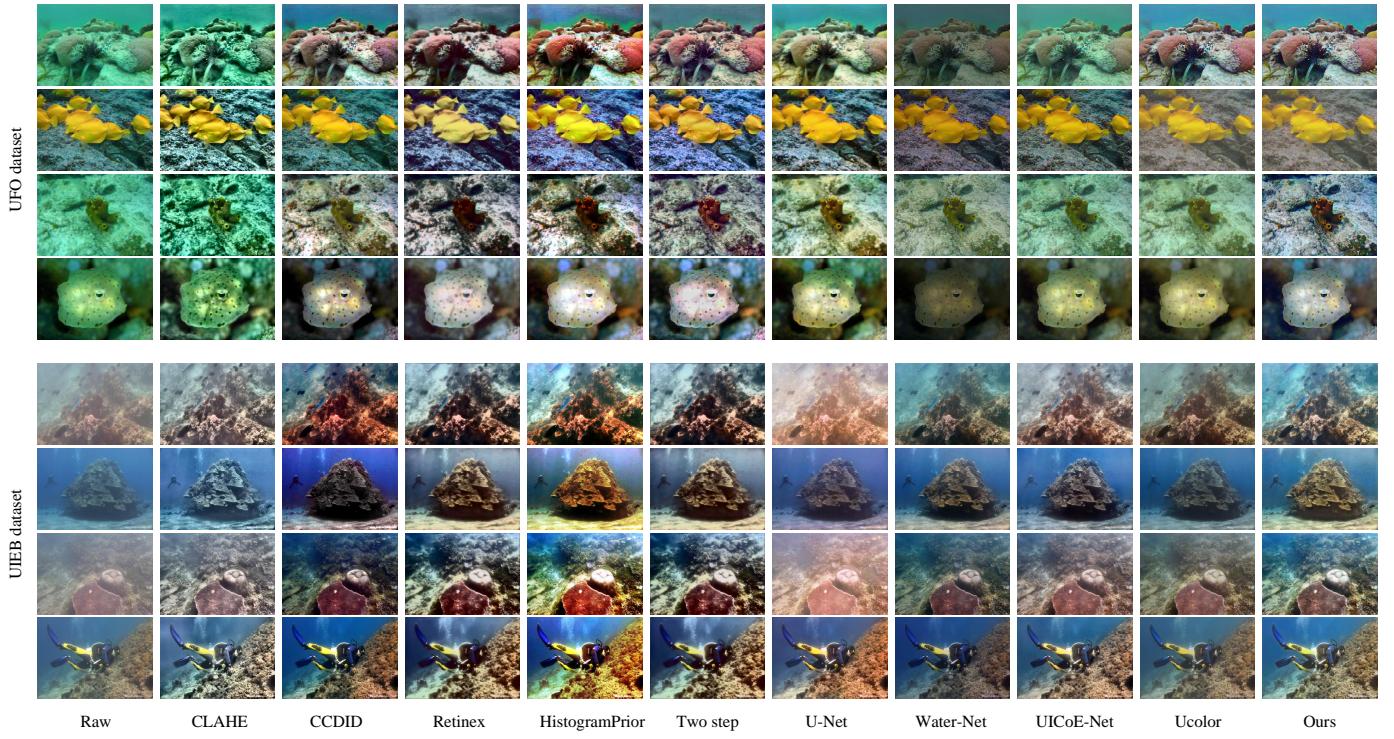


Figure 4: The enhancement comparison on the UFO-120 and UIEB test sets. Traditional methods including CLAHE [65], CCDID [10], Retinex [15], HistogramPrior [26], TwoStep [14] and latest deep enhancement models including U-Net [40], Water-Net [25], UICoE-Net [39], Ucolor [24] are used as the competitors.

Table 3: Evaluation of the enhancement performance on UFO-120 and UIEB datasets. For clarity, the traditional and deep learning-based approaches in the table are separated into two parts with a line.

Datasets	UFO-120		UIEB	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑
CLAHE [65]	15.25	0.5965	17.08	0.8053
CCDID [10]	18.65	0.7486	16.50	0.7966
Retinex [15]	15.39	0.6306	18.10	0.7843
HistogramPrior [26]	14.71	0.6202	20.61	0.8163
Two-step [14]	16.51	0.6801	18.09	0.7813
U-Net [40]	19.68	0.7771	18.81	0.8337
Water-Net [25]	18.99	0.7744	19.93	0.8523
UICoE-Net [39]	25.51	0.8523	20.35	0.8669
Ucolor [24]	23.68	0.7998	20.16	0.8379
Ours	24.10	0.8197	20.70	0.8735

SQUID and Color-Check7 are another two more challenging underwater image enhancement benchmarks. SQUID is widely used to evaluate the performance on relatively deep water images. Color-Check7 contains images photographed with different cameras to test the robustness and accuracy of color correction.

To give comprehensive experiments and insight analysis, we first evaluate the performance of the CoE module and CoSR

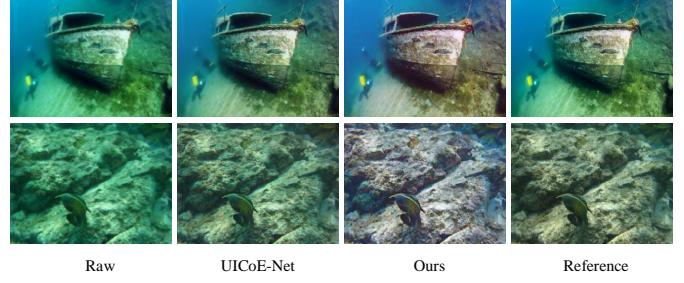


Figure 5: Enhancement examples of the proposed enhancement model which have much better visual quality. From left to right, the raw images, enhancements of the UICoE-Net and our CoE module, their corresponding references given by UFO-120 are presented, respectively.

module by comparing them with a number of underwater image enhancement and super-resolution approaches in Section 4.2 and Section 4.3, respectively. Then, for the complete CoESR-Net, we conducted comprehensive ablation experiments to verify the effectiveness of each module and the results are reported in Section 4.4. Besides, we perform a comparison of the two training strategies for the enhancement module and super-resolution module, i.e. separated training and end-to-end training, which demonstrates the effectiveness and necessity of the unified collaborative framework for both underwater image enhancement and super-resolution. Finally, run-time and efficiency analysis are given in Section 4.5.

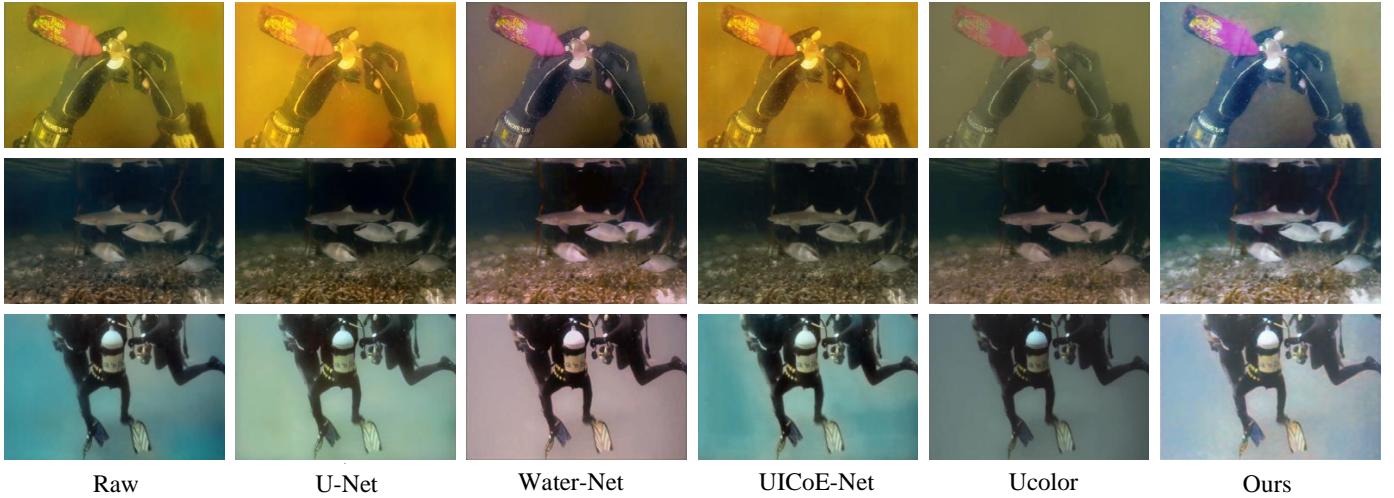


Figure 6: The enhancement comparison on UIEB challenging set. From left to right, enhancements of U-Net [40], Water-Net [25], UICoE-Net [39], Ucolor [24] and our model are presented. The deep models are trained on the UIEB training set.

4.2. Experiments on Enhancement Module

4.2.1. Compared Methods

To evaluate the performance of the proposed collaborative enhancement module, we compare it with 9 typical enhancement methods, including 5 traditional approaches, i.e., CLAHE [65], CCDID [10], Retinex [15], HistogramPrior [26], Two-step [14], and 4 deep UIE models, i.e., U-Net [40], Water-Net [25], UICoE-Net [39], Ucolor [24].

CLAHE is an improved histogram equalization method, which has been widely used to improve the unbalanced color distribution in underwater images. CCDID is a dark channel prior-based dehaze method, which also takes color cast into consideration. Considering that the degradation of underwater images likewise show obvious hazy appearance and color distortion, CCDID can be used as a UIE approach as well. Retinex method proposes a variational retinex model to compute the reflectance and illumination of the image. Then, two histogram-based methods are used to enhance them separately. The HistogramPrior method consists of an underwater image dehazing algorithm and a contrast enhancement algorithm. As to the Two-step method, Fu et al. proposed a piece-wise linear transformation to correct the distorted color appearance. Then, they further proposed an optimal contrast strategy for hazy detail improvement.

U-Net is a classical network which has been widely used in image-to-image tasks, such as image enhancement, image segmentation, image style transfer, etc. As we introduced in Section 3.2.2, Water-Net is fusion strategy-based deep enhancement model, which tries to fuse the results from different enhancements. UICoE-Net and Ucolor are two latest underwater image enhancement networks. UICoE-Net introduces multi-branch correlation for feature complement and joint learning. Different from traditional deep networks, which mainly enhance the underwater image in an independent color space, Ucolor introduces multi-color space encoder with channel attention for better color correction.

Table 4: The performance comparison on UIEB challenging set of deep enhancement models, which are trained on UIEB training set. Non-reference metrics UIQM, UCIQE and perceptual score are reported.

Methods/Metrics	UIQM↑	UCIQE↑	PS↑
U-Net [40]	0.3033	0.5044	1.54
Water-Net [25]	0.3594	0.5497	1.75
UICoE-Net [39]	0.2699	0.5089	1.59
Ucolor [24]	0.2589	0.4990	1.43
Ours	0.6235	0.5315	2.50

To comprehensively compare the performance of the proposed enhancement module and the above-mentioned competitors, we perform both qualitative and quantitative comparison on 4 underwater image enhancement benchmarks, including UFO-120, UIEB, SQUID and Color-Check7. Specifically, we only trained those deep models on UFO-120 and UIEB datasets which have paired reference. On SQUID and Color-Check7 datasets, we only perform the test.

4.2.2. Comparison on UFO-120 and UIEB Datasets

The quantitative evaluation on UFO-120 and UIEB datasets are reported in Table 3. For more intuitive visual comparison, we present enhancement examples of different approaches on Figure 4. According to the two commonly used reference-based quality metrics, i.e., PSNR and SSIM, our Co-enhancement module achieves the best performance on UIEB dataset, and the second best performance on UFO-120 dataset, which is only behind the latest UICoE-Net. Part of the reason is that unlike the proposed method which only performs feature collaboration in one scale, UICoE-Net established a multi-scale feature correlation mechanism. It helps to capture the complementary information from multiple scales for image detail recovery, but also inevitably introduces tremendous extra computation. For

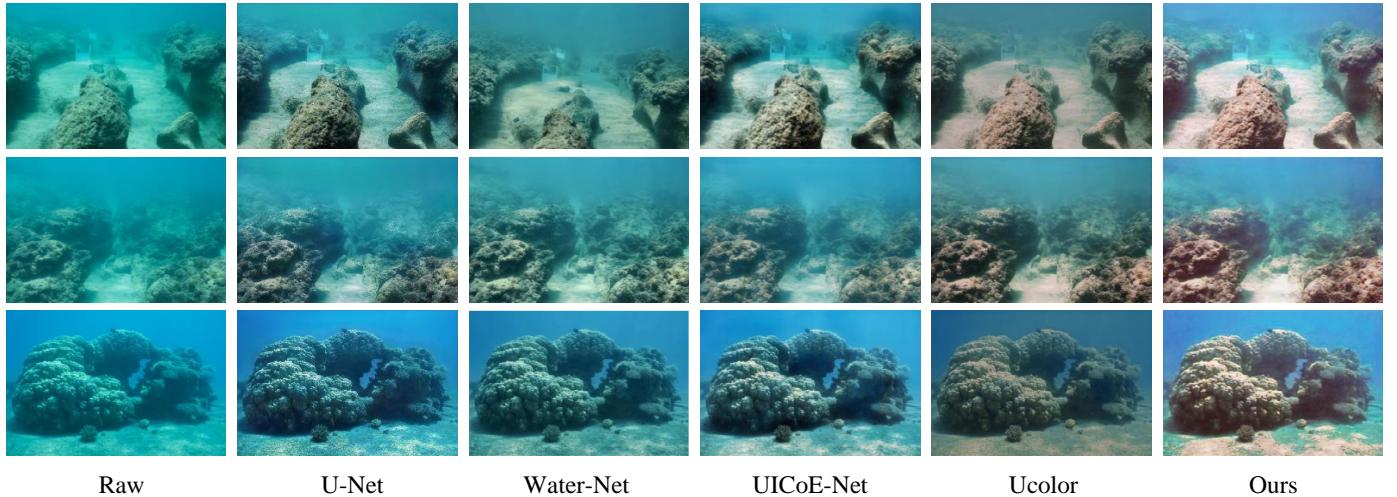


Figure 7: The enhancement comparison on SQUID dataset. From left to right, enhancements of U-Net [40], Water-Net [25], UICoE-Net [39], Ucolor [24] and our model are presented. The deep models are trained on the UIEB training set.

Table 5: The performance comparison on SQUID dataset of deep enhancement models. Deep models trained on UIEB and UFO-120 datasets are both presented. Non-reference metrics UIQM, UCIQE and perceptual score are reported.

Training sets	UFO-120			UIEB		
	Methods/Metrics	UIQM↑	UCIQE↑	PS↑	UIQM↑	UCIQE↑
U-Net [40]	0.0947	0.4257	1.55	0.1232	0.4264	1.37
Water-Net [25]	0.0085	0.4115	1.25	0.0552	0.3851	1.46
UICoE-Net [39]	0.0168	0.3880	1.45	0.1492	0.4453	1.55
Ucolor [24]	0.0192	0.3624	1.24	0.1204	0.4894	1.76
Ours	0.7256	0.5023	2.02	0.6519	0.5375	2.12

the collaborative UIE and SR task we concern, we proposed to perform mutual transmission of associated features on different stages of the UIE module and those cascaded CoSR blocks by balancing efficiency against performance. As presented in Figure 5, another reason why the scores of the proposed method are lower than those of UICoE-Net is that the quality of some enhancement references provided by UFO-120 is not very satisfactory. The references presented in Figure 5 still keep obvious greenish distortion. While, the results of our enhancement module deliver much more impressive visual perception than the given reference. However, for these cases, UICoE-Net will receive higher scores according to the reference-based metrics since its results are more close to the references, though our results are much better. These cases also show the phenomenon that directly learning the color correction mappings from degraded domain to the clear domain may not be suitable for those images suffering severe color cast. Take the results in Figure 4 for example, both U-Net and UICoE-Net left greenish cast on UFO-120 cases. On the one hand, this is due to the poor quality of the training reference and small number of such severely degraded samples, and on the other hand, because it is very difficult to directly learn such complex mappings.

Thanks to the delicate strategies designed for contrast en-

hancement and color correction, traditional methods are generally able to deal with the color cast issues more easily. While, for those underwater images with only slight color cast, such as the UIEB cases presented in Figure 4, traditional method usually produce severe over-correction, such as showing obvious red distortion and over-saturation. Therefore, in terms of both metrics and visual quality, the traditional methods do not perform well on the above datasets.

Water-Net and our proposed CoE module both take white balance, histogram equalization and Gamma correction as the pre-processing for network input initialization, which can effectively deal with the issues of severe color cast. Moreover, since our proposed BRUs provide the network with more robust feature transformer for images from related scenes or with similar degradation and MTU shares their complementary information, our CoE module generate much more vivid enhancement than that of Water-Net. On both UFO-120 and UIEB datasets, Ucolor achieves quite good color correction performance since it introduces multi-color space encoder. While, detail structure recovery is not fully explored in Ucolor framework, which leads to obvious hazy appearance and relatively low SSIM scores.

Since the volunteers can not reach consensus on their optimal enhancement candidates, 60 underwater images are viewed

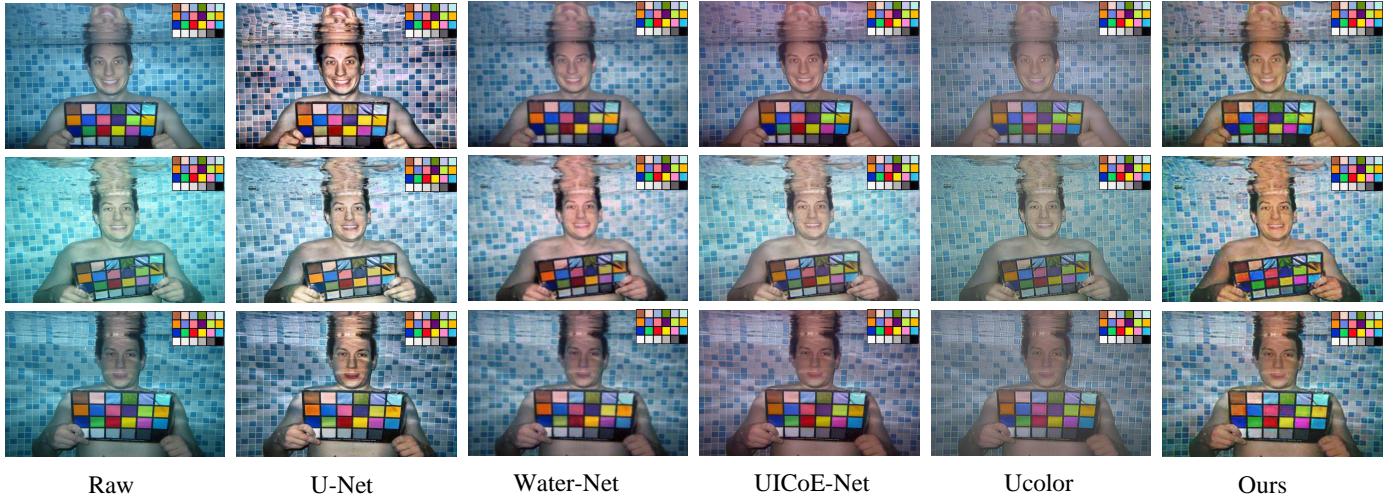


Figure 8: The enhancement comparison on Color-Check7. From left to right, enhancements of U-Net [40], Water-Net [25], UICoE-Net [39], Ucolor [24] and our model are presented. The deep models are trained on the UFO-120 training set. Ground truth colors are presented on the top-right corner of each image.

Table 6: Comparison of color correction performance on Color-Check7. Deep models trained on UFO-120 Training are evaluated. The color dissimilarity between corrected color and ground truth color in terms of the CIEDE2000 are reported.

Methods	Can_D10	Fu_Z33	Oly_T60	Oly_T80	Pan_TS1	Pen_W60	Pen_W80	Average
Raw	10.33	16.64	14.98	19.30	16.15	11.96	14.12	14.78
U-Net [40]	11.20	11.48	11.03	14.09	17.02	9.95	11.87	12.38
Water-Net [25]	12.03	12.67	10.71	15.60	14.49	11.08	14.93	13.07
UICoE-Net [39]	12.29	10.31	10.26	19.69	22.97	10.53	17.35	14.77
Ucolor [24]	10.75	13.20	12.33	18.80	18.05	10.34	13.90	13.91
Ours	10.32	11.86	7.58	10.68	7.25	10.95	10.96	9.95

as challenging cases and form the UIEB challenging set without corresponding reference. We test all the deep enhancement models on this dataset and the evaluation is reported in Table 4. We adopted the commonly used non-reference metrics for quantitative evaluation, i.e., UIQM [37] and UCIQE [54]. Our Co-enhancement module achieves the best UIQM score and second best UCIQE score. While, as presented in Figure 6, by visually comparing the enhanced images, our results are more natural than that of Water-Net. Due to the lack of robust and consistent feature transformation provided by the proposed BRU and MTU, the enhancements of Water-Net will be more easily dominated by those fragile enhancements of pre-processing.

In addition to the above metrics, we further conducted user study to give subjective evaluation according to the visual perceptual quality. 8 volunteers with basic image processing knowledge are invited to give a score to each enhanced image. The perceptual scores (PS) range from 1 to 5, which present the worst to the best. For each image, the order of enhancements of different approaches are randomly shuffled to ensure fairness. Impressively, our Co-enhancement module achieves the highest average perceptual score. This result indicates that the enhancement results obtained by our method are more consistent

Table 7: The performance comparison on Color-Check7 dataset of deep enhancement models, which are trained on UFO-120 training set. Non-reference metrics UIQM, UCIQE and perceptual score are reported.

Methods/Metrics	UIQM↑	UCIQE↑	PS↑
U-Net [40]	0.9480	0.4579	2.73
Water-Net [25]	0.8479	0.4501	2.19
UICoE-Net [39]	0.7667	0.4307	2.03
Ucolor [24]	0.6951	0.3681	1.35
Ours	0.9956	0.4591	3.51

with human visual perception on this challenging dataset.

4.2.3. Enhancement and Generalization Comparison on SQUID and Color-Check7 Datasets

SQUID dataset contains 57 stereo pairs from relative deep water areas in Israel (from 3-6 meters to 20-30 meters), which therefore have severe degradations, such as diverse heavy color distortion, low contrast and hazy appearance. Color-Check7 dataset includes 7 images photographed with different cameras in a swimming pool. It is commonly used for robust test and color correction accuracy evaluation since color checker is pho-

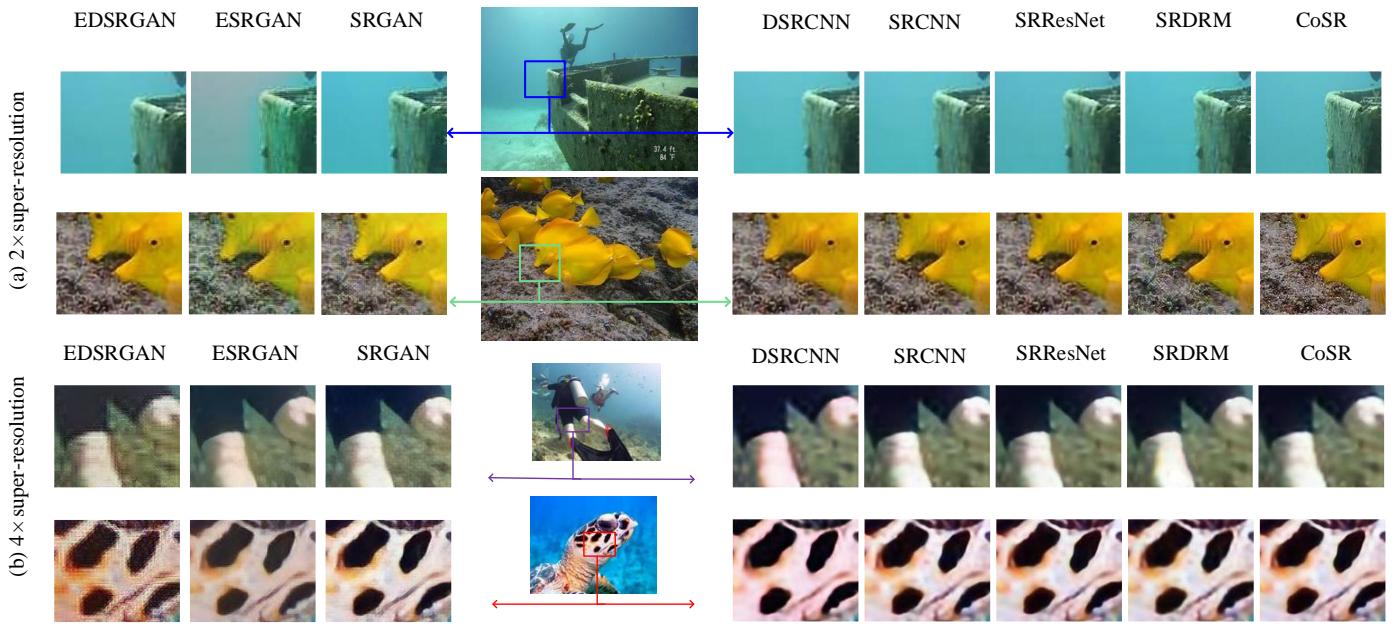


Figure 9: The super-resolution performance comparison on images from USR-248 dataset. From left to right, the magnified areas present the results of EDSRGAN [33], ESRGAN [47], SRGAN [22], DSRCNN [36], SRCNN [12], SRResNet [55], SRDRM [18] and the proposed CoSR.

Table 8: Evaluation of super-resolution performance. The results on 2x, 4x and 8x super-resolution are presented in a form as 2x score / 4x score / 8x score.

Datasets	UFO-120			USR-248	
	Methods/Metrics	PSNR↑	SSIM↑	PSNR↑	SSIM↑
SRGAN [22]		27.62 / 24.63 / 22.11	0.80 / 0.72 / 0.61	28.05 / 24.76 / 20.14	0.78 / 0.69 / 0.60
ESRGAN [47]		26.47 / 23.09 / 21.44	0.78 / 0.65 / 0.59	26.66 / 23.79 / 19.75	0.75 / 0.66 / 0.58
EDSRGAN [33]		27.31 / 23.28 / 21.38	0.87 / 0.74 / 0.57	27.12 / 21.65 / 19.87	0.77 / 0.65 / 0.58
SRResNet [55]		28.43 / 24.54 / 22.01	0.87 / 0.75 / 0.66	25.98 / 24.15 / 19.26	0.72 / 0.66 / 0.55
SRCNN [12]		28.17 / 24.16 / 22.12	0.86 / 0.75 / 0.64	26.81 / 23.38 / 19.97	0.76 / 0.67 / 0.57
DSRCNN [36]		28.07 / 24.31 / 22.06	0.84 / 0.73 / 0.65	27.14 / 23.61 / 20.14	0.77 / 0.67 / 0.56
SRDRM [18]		28.46 / 24.44 / 22.16	0.88 / 0.76 / 0.66	28.36 / 24.64 / 21.20	0.80 / 0.68 / 0.60
Ours		28.51 / 24.68 / 22.18	0.89 / 0.77 / 0.68	28.59 / 25.89 / 21.79	0.82 / 0.71 / 0.61

tographed into each image. With the deep enhancement models trained on UFO-120 and UIEB dataset, the testing experiment on SQUID and Color-Check7 datasets can effectively check their enhancement performance and generalization.

The quantitative comparison on SQUID and Color-Check7 datasets are presented in Tables 5 and 7, respectively. On both SQUID and Color-Check7 datasets, our CoE module achieves the best scores on UIQM and UCIQE metrics. Take the examples presented in Figure 7 for example, since the images taken from the deep water areas are severely degraded to dark greenish tone, traditional deep enhancement models can hardly correct this color distortion effectively and completely. While, thanks to the proposed collaborative strategy for joint learning, complementary supervision from the collaborative images facilitates the learning of color correction mapping. In addition to the evaluation of non-reference metrics, the perceptual

scores given by the volunteers reflect that our enhancement also have better visual quality. Notably, our method provides more consistent color correction on images captured with different cameras, which demonstrate the robustness of our enhancement model.

To further evaluate the color correction accuracy of different methods, we measure the color dissimilarity between corrected colors and the ground truth colors in the color checker according to the CIEDE2000. As the results shown in Table 6, the corrected colors of our model are much more close to the given ground truth color. This is consistent with the visual effect presented in Figure 8.

4.3. Experiments on Super-Resolution Module

To evaluate the performance of the proposed CoSR module, we compared it with 7 deep learning-based super-resolution methods, including 3 GAN-based approaches, i.e., SRGAN

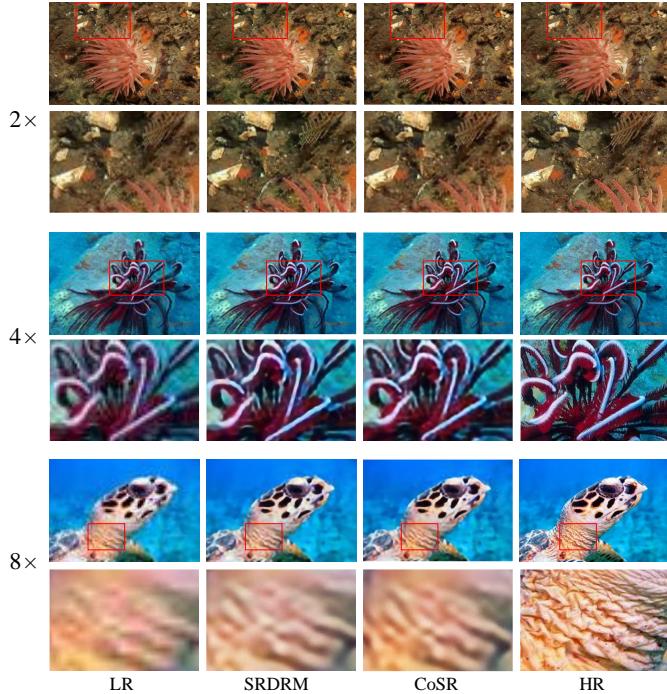


Figure 10: The super-resolution performance comparison on images from USR-248 dataset between the baseline SRDRM [18] and our CoSR. From left to right, the low resolution (LR) input image, outputs of SRDRM, outputs of our CoSR and the high resolution (HR) ground truth are presented, respectively.

[22], ESRGAN [47], EDSRGAN [33] and 4 CNN-based approaches, i.e., SRResNet [55], SRCNN [12], DSRCNN [36], SRDRM [18]. On Table 8, we present the quantitative evaluation results of these SR approaches on UFO-120 and USR-248 datasets. We tested different super-resolution magnifications, i.e., 2 \times , 4 \times and 8 \times , to test the robustness. On both PSNR and SSIM metrics, our CoSR module achieves the best scores on all the super-resolution magnifications. Examples of super-resolution on images from USR-248 dataset are presented in Figure 9. It can be noticed that, the super-resolution results of the proposed CoSR module are usually recovered more local details, which shows as more sharpen edges (such as the edges of the ship’s side on the 1st row), finer textures (such as the sand on the 2nd row) and less missing details (such as shadow on the leg on the 3rd row). While, the performance of GAN-based approaches is no stable and robust enough.

Notably, the competitor SRDRM, which has no multi-stage mutual transmission units for feature complement, can be viewed the base model of the proposed CoSR module. By comparing CoSR with SRDRM, the quantitative scores of CoSR also show obvious advantages, which demonstrates the effectiveness of mutual transmission units for the underwater image super-resolution task. In Figure 10, we present several 2 \times , 4 \times and 8 \times examples of SRDRM and our CoSR module for visually detail comparison. It can be noticed that, for diverse underwater scenes with quite limited training samples, the super-resolution of SRDRM may suffer obvious texture defects, such as the 2 \times result in Figure 10. While, thanks to the mutual feature

transmission strategy, the correlated features bring additional constraint for better and robust texture recovery. From the 4 \times results in Figure 10, it shows that, in addition to producing better recovery of local details, our method can also better preserve color appearance. It is a good property since the color degradation can usually arises during the compression and non-robust super-resolution. As to the severe detail compression shown in the 8 \times super-resolution case, the recovery of textures, edges and even color using CoSR also shows obvious advantages.

4.4. Ablation Study and Experiments on End-to-End Strategy

To explore the effectiveness and contribution of different key modules of the proposed CoESR-Net, we conducted ablation study on both UFO-120 and UIEB datasets. The qualitative and quantitative comparison are presented in Figure 11 and Table 9, respectively. Specifically, the ablation setting include

- * CoESR-w/o-BRU: The CoESR-Net without the branch reconstruction units for robust and consistent feature transformation;
- * CoESR-w/o-CoSR: The CoESR-Net without mutual transmission unit for collaborative super-resolution;
- * CoESR-w/o-CoE: The CoESR-Net without mutual transmission unit for collaborative enhancement;
- * CoESR-w/o-UIENet: The CoESR-Net without underwater image enhancement module, which only uses the super-resolution module to learn the whole enhancement and super-resolution mapping task.

Accordingly, we could draw the following conclusions from the quantitative and qualitative comparison:

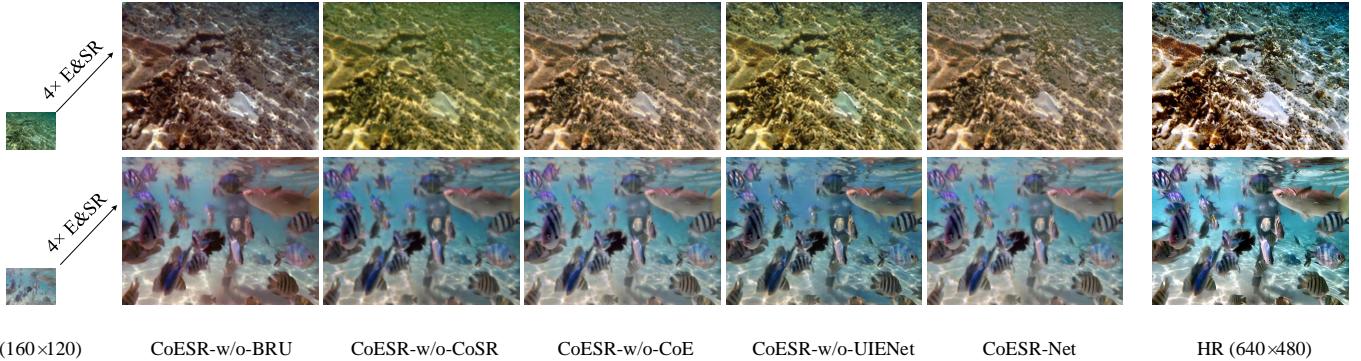
(a) When compared with all the ablated models mentioned above, our full CoESR-Net achieves relative better quantitative on both UFO-120 and UIEB datasets, which demonstrate the effectiveness of BRU, CoSR and CoE.

(b) In the fusion-based enhancement module, the proposed BRU effectively avoids over-correction (such as the red tone of the results of CoESR-w/o-BRU) due to the inappropriate pre-enhancement by performing robust feature reconstruction.

(c) The contribution of CoSR mainly lies in generating more accurate texture and detail structures, which leads to the big gap of SSIM scores between the CoESR-w/o-CoSR and the full model. Besides, by compare the results of CoESR-w/o-CoSR and CoESR-Net in Figure 11, we can notice that CoSR also provides partial enhancements, such as color correction.

(d) The quantitative improvement of CoE module is relatively less than that of CoSR, to some extent this is because that the following cascaded MTUs of CoSR can also provides rich associated features to compensate the loss of CoE. While, the contribution of CoE module is mainly for the robust recovery with the compensation supervision.

(e) Removing the UIE module and only keeping the original SR model to handle both enhancement and super-resolution tasks will also lead to the significant performance lose. As shown in Figure 11, the performance degradation is mainly



LR (160×120) CoESR-w/o-BRU CoESR-w/o-CoSR CoESR-w/o-CoE CoESR-w/o-UIENet CoESR-Net HR (640×480)

Figure 11: Examples of the ablation study. From left to right, the low resolution distorted inputs, results of CoESR-w/o-BRU, CoESR-w/o-CoSR, CoESR-w/o-CoE, CoESR-w/o-UIENet, CoESR-Net and the high resolution enhanced reference are presented, respectively.

Table 9: Evaluation results of ablation experiments. The scores report the performance of the abated models on end-to-end enhancement and super-resolution task. The results on the tasks with 2 \times , 4 \times and 8 \times super-resolution are presented in a form as 2 \times score / 4 \times score / 8 \times score.

Datasets	UFO-120			UIEB	
	Methods/Metrics	PSNR \uparrow	SSIM \uparrow	PSNR \uparrow	SSIM \uparrow
CoESR-w/o-BRU	28.43 / 25.42 / 21.98	0.87 / 0.80 / 0.73	21.67 / 20.96 / 20.24	0.81 / 0.78 / 0.69	
CoESR-w/o-CoSR	28.54 / 25.22 / 22.12	0.88 / 0.76 / 0.72	21.94 / 21.48 / 20.35	0.83 / 0.80 / 0.72	
CoESR-w/o-CoE	28.76 / 25.61 / 22.23	0.89 / 0.80 / 0.71	22.32 / 22.27 / 21.12	0.88 / 0.84 / 0.74	
CoESR-w/o-UIENet	28.41 / 24.78 / 22.22	0.89 / 0.75 / 0.73	21.92 / 21.39 / 20.64	0.88 / 0.80 / 0.73	
CoESR	28.78 / 25.43 / 22.37	0.90 / 0.83 / 0.75	22.65 / 22.42 / 21.75	0.89 / 0.84 / 0.72	

manifested in color correction. Thanks to the preserved CoSR module, the recovery of local details, structures and contrast is satisfactory. This indicates that retaining independent enhancement module is necessary for underwater image-to-image task.

In this paper, we proposed an collaborative network for both underwater image enhancement and super-resolution, which is organized and trained in an end-to-end way. An intuitive concern is that whether end-to-end training is more suitable for this joint task than separated training. Therefore, on both UFO-120 and UIEB datasets, we trained the CoE module and CoSR module separately. On the testing stage, we feed the CoSR module with the results of CoE module to generate the enhancements with large resolution. The quantitative and qualitative comparison between these two training strategies are presented in Table 10 and Figure 12, respectively.

According to the quantitative results, we can find that the end-to-end training strategy shows impressive advantages. It demonstrates that taking the challenging underwater image enhancement and super-resolution as collaborative tasks within a unified framework is necessary and promising. As shown in Figure 12(b) and Figure 12(c), for the separated training strategy, the quality of final high resolution results could be greatly limited by the performance of separated enhancement module. While, the result shown in Figure 12(d) demonstrates the potential of an end-to-end learning framework for both UIE and SR. This is partly due to the fact that underwater image enhancement and super-resolution are both image-to-image tasks, where feature information could be shared. And on the other

hand, the unified network provides a deeper structure for learning complex image-to-image mappings, which could be quite hard for separated networks, especially the enhancement network.

4.5. Run-time Analysis

In this section, we measure the run-time costs of the proposed CoESR-Net and its ablated versions to give a comprehensive time-cost analysis about the model and key modules. For underwater images with resolutions of 80 × 60, and 160 × 120 and 320 × 240, we perform 2 \times , 4 \times and 8 \times super-resolution and enhancement to generate clear underwater images with resolution of 640×480, respectively. In addition to the run-time of the full CoESR-Net, we also test the run-time costs of CoESR-w/o-CoE, CoESR-w/o-CoSR and CoESR-w/o-UIENet. In Table 11, both run-time (ms) and FPS are reported. For the joint task of 8 \times super-resolution and enhancement, which is the most computationally intensive one in our test, it costs 215 ms per image, i.e., 4.65 FPS. Removing the MTUs in CoE or CoSR modules can improve the operation efficiency to some extent. However, the run-time reduction is quite limited compared to the total time cost (from 215.00 to 199.17 and 203.33, respectively). This shows that the introduced collaborative strategy does not significantly increase the computational cost. The CoESR-Net-w/o-UIENet model, which only keeps CoSR module for both UIE and SR tasks, achieves 0.66 FPS boost compared with the full model. But, as reported in Table 9, the speed boost does

Table 10: The comparison of the two training strategies for the enhancement module and super-resolution module, i.e. separated training and end-to-end training. The results on the tasks with 2x, 4x and 8x super-resolution are presented in a form as 2x score / 4x score / 8x score.

Datasets	Training Strategy	PSNR↑	SSIM↑
UFO-120	Separated training	23.86 / 22.41 / 20.66	0.76 / 0.73 / 0.71
	End-to-end training	28.78 / 25.43 / 22.37	0.90 / 0.83 / 0.75
UIEB	Separated training	21.68 / 21.51 / 19.93	0.84 / 0.78 / 0.71
	End-to-end training	22.65 / 22.42 / 21.75	0.89 / 0.84 / 0.72

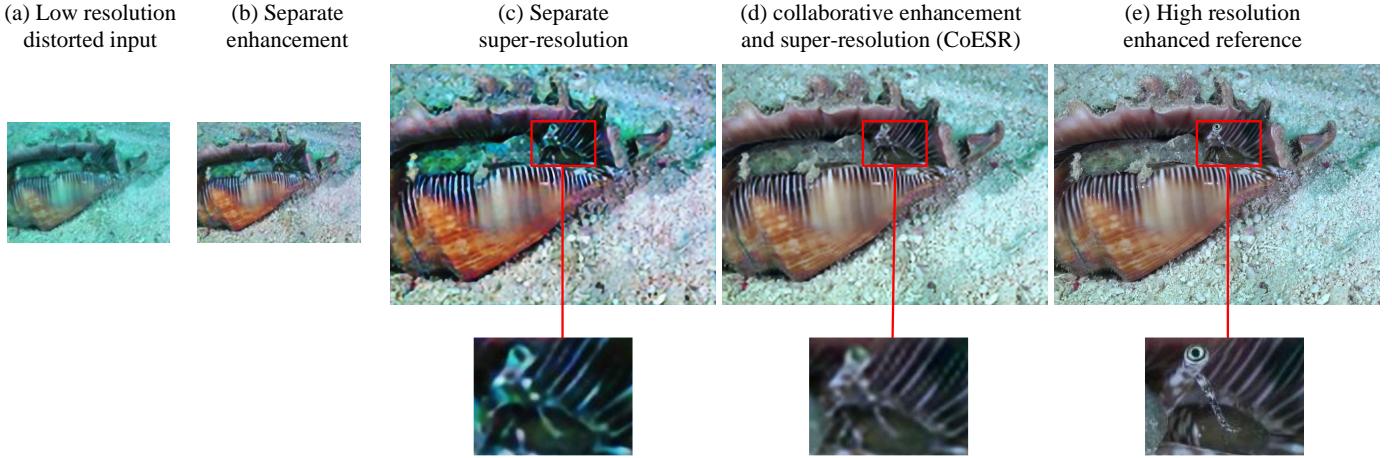


Figure 12: An example of the comparison between the separated enhancement/super-resolution strategy and the collaborative enhancement and super-resolution strategy with end-to-end training. (a) shows the low resolution distorted images for the inputs, (b) and (c) are the outputs of separate enhancement module and separate super-resolution model, respectively. (d) shows the outputs of the proposed collaborative enhancement and super-resolution model. (e) is the reference of high resolution enhancement images.

Table 11: Execution efficiency of the proposed model and the ablated models. Test on tensorflow platform with a NVIDIA GeForce RTX 2080Ti GPU. Both run-time (ms) and FPS are reported in the form of run-time/FPS.

Models/Tasks	2x E&SR	4x E&SR	8x E&SR
CoESR-Net	132.50 / 7.54	144.16 / 6.93	215.00 / 4.65
-w/o-CoE	121.67 / 8.21	130.83 / 7.64	199.17 / 5.02
-w/o-CoSR	124.17 / 8.05	135.04 / 7.40	203.33 / 4.92
-w/o-UIENet	114.17 / 8.77	123.33 / 8.10	188.33 / 5.31

not make up for the performance losses. Therefore, it is necessary to introduce independent enhancement modules in the joint UIE and SR task. For 2x and 4x E&SR tasks, the full models achieve 7.54 FPS and 6.93 FPS, respectively. For two joint image-to-image tasks, the efficiency of the proposed model is promising.

5. Conclusion

In this paper, we propose a unified solution for joint underwater image enhancement and super-resolution. To fully mine associated supervision from other scene-related images, we further introduce a multi-stage mutual feature transmission strategy to establish a collaborative learning framework. Moreover,

we introduce branch reconstruction unit to perform robust feature transformation for fusion. Extensive experiments on diverse benchmarks for underwater image enhancement, super-resolution and the joint task demonstrated the effectiveness of the collaborative strategy and the end-to-end joint learning framework. Ablation study and the run-time analysis on the ablated models give an insight into the effectiveness and efficiency of the key components.

Acknowledgment

The research has been supported by the National Natural Science Foundation of China under Grant 61906177, in part by the Natural Science Foundation of Shandong Province under Grant ZR2019BF034, in part by Significant Applied Technology Innovation Projects for Agriculture of Shandong Province under Grants SD2019NJ020, in part by Fundamental Research Funds for the Central Universities under Grants 201964013.

References

- [1] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, P. Bekaert, Color balance and fusion for underwater image enhancement, *IEEE Transactions on Image Processing* 27 (1) (2018) 379–393.
- [2] S. Anwar, C. Li, Diving deeper into underwater image enhancement: A survey, *Signal Processing: Image Communication* 89 (2020) 115978.

- [3] S. Anwar, C. Li, F. M. Porikli, Deep underwater image enhancement, ArXiv abs/1807.03528.
- [4] D. Berman, D. Levy, S. Avidan, T. Treibitz, Underwater single image color restoration using haze-lines and a new quantitative dataset, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43 (8) (2021) 2822–2837.
- [5] M. Cai, Y. Wang, S. Wang, R. Wang, Y. Ren, M. Tan, Grasping marine products with hybrid-driven underwater vehicle-manipulator system, *IEEE Transactions on Automation Science and Engineering* 17 (3) (2020) 1443–1454.
- [6] H. Chen, Y. Huang, H. Nakayama, Semantic aware attention based deep object co-segmentation, in: *Asian Conference on Computer Vision*, 2018, pp. 435–450.
- [7] X. Chen, S. Wei, C. Yi, L. Quan, C. Lu, Progressive attentional learning for underwater image super-resolution, in: *International Conference on Intelligent Robotics and Applications*, 2020, pp. 233–243.
- [8] Y. Chen, J. Sun, W. Jiao, G. Zhong, Recovering super-resolution generative adversarial network for underwater images, in: *International Conference on Neural Information Processing*, 2019, pp. 75–83.
- [9] N. Cheng, T. Zhao, Z. Chen, X. Fu, Enhancement of underwater images by super-resolution generative adversarial networks, in: *International Conference on Internet Multimedia Computing and Service*, 2018.
- [10] S. K. Dhara, M. Roy, D. Sen, 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* 31 (5) (2020) 2076–2081.
- [11] X. Ding, Y. Wang, Z. Liang, J. Zhang, X. Fu, Towards underwater image enhancement using super-resolution convolutional neural networks, in: *International Conference on Internet Multimedia Computing and Service*, 2018, pp. 479–486.
- [12] C. Dong, C. C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38 (2) (2016) 295–307.
- [13] C. Fabbri, M. J. Islam, J. Sattar, Enhancing underwater imagery using generative adversarial networks, in: *IEEE International Conference on Robotics and Automation*, 2018, pp. 7159–7165.
- [14] X. Fu, Z. Fan, M. Ling, Y. Huang, X. Ding, Two-step approach for single underwater image enhancement, in: *International Symposium on Intelligent Signal Processing and Communication Systems*, 2017, pp. 789–794.
- [15] X. Fu, P. Zhuang, Y. Huang, Y. Liao, X. Zhang, X. Ding, A retinex-based enhancing approach for single underwater image, in: *IEEE International Conference on Image Processing*, 2014, pp. 4572–4576.
- [16] Z. He, J. Li, L. Liu, D. He, M. Xiao, Multiframe video satellite image super-resolution via attention-based residual learning, *IEEE Transactions on Geoscience and Remote Sensing* (2021) 1–15.
- [17] M. J. Islam, P. Luo, J. Sattar, Simultaneous Enhancement and Super-Resolution of Underwater Imagery for Improved Visual Perception, in: *Robotics: Science and Systems*, 2020.
- [18] M. J. Islam, S. Sakib Enan, P. Luo, J. Sattar, Underwater image super-resolution using deep residual multipliers, in: *IEEE International Conference on Robotics and Automation*, 2020, pp. 900–906.
- [19] M. J. Islam, Y. Xia, J. Sattar, Fast underwater image enhancement for improved visual perception, *IEEE Robotics and Automation Letters* 5 (2) (2020) 3227–3234.
- [20] K. Katija, P. L. D. Roberts, J. Daniels, A. Lapides, K. Barnard, M. Risi, B. Y. Ranaan, B. G. Woodward, J. Takahashi, Visual tracking of deep-water animals using machine learning-controlled robotic underwater vehicles, in: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 860–869.
- [21] M. Kawulok, P. Benecki, S. Piechaczek, K. Hrynczenko, D. Kostrzewska, J. Nalepa, Deep learning for multiple-image super-resolution, *IEEE Geoscience and Remote Sensing Letters* 17 (6) (2020) 1062–1066.
- [22] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, W. Shi, Photo-realistic single image super-resolution using a generative adversarial network, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 105–114.
- [23] B. Li, Z. Sun, Q. Li, Y. Wu, H. Anqi, Group-wise deep object co-segmentation with co-attention recurrent neural network, in: *IEEE/CVF International Conference on Computer Vision*, 2019, pp. 8518–8527.
- [24] C. Li, S. Anwar, J. Hou, R. Cong, C. Guo, W. Ren, Underwater image enhancement via medium transmission-guided multi-color space embedding, *IEEE Transactions on Image Processing* 30 (2021) 4985–5000.
- [25] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, D. Tao, An underwater image enhancement benchmark dataset and beyond, *IEEE Transactions on Image Processing* 29 (2020) 4376–4389.
- [26] C. Li, J. Guo, R. Cong, Y. Pang, B. Wang, Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior, *IEEE Transactions on Image Processing* 25 (12) (2016) 5664–5677.
- [27] C. Li, J. Guo, C. Guo, Emerging from water: Underwater image color correction based on weakly supervised color transfer, *IEEE Signal Processing Letters* 25 (3) (2018) 323–327.
- [28] H. Li, F. Meng, K. N. Ngan, Co-salient object detection from multiple images, *IEEE Transactions on Multimedia* 15 (8) (2013) 1896–1909.
- [29] K. Li, S. Qi, H. Yang, L. Zhou, D. Song, Extensible image object co-segmentation with sparse cooperative relations, *Information Sciences* 521 (2020) 422–434.
- [30] K. Li, J. Zhang, W. Tao, Unsupervised co-segmentation for indefinite number of common foreground objects, *IEEE Transactions on Image Processing* 25 (4) (2016) 1898–1909.
- [31] W. Li, O. Hosseini-Jafari, C. Rother, Deep object co-segmentation, in: *Asian Conference on Computer Vision*, 2018, pp. 638–653.
- [32] X. Li, G. Hou, K. Li, Z. Pan, Enhancing underwater image via adaptive color and contrast enhancement, and denoising, *Engineering Applications of Artificial Intelligence* 111 (2022) 104759.
- [33] B. Lim, S. Son, H. Kim, S. Nah, K. M. Lee, Enhanced deep residual networks for single image super-resolution, in: *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017, pp. 1132–1140.
- [34] X. Liu, Z. Gao, B. M. Chen, Mifcgan: Multilevel feature fusion-based conditional gan for underwater image color correction, *IEEE Geoscience and Remote Sensing Letters* (2019) 1–5.
- [35] A. Manzanilla, S. Reyes, M. Garcia, D. Mercado, R. Lozano, Autonomous navigation for unmanned underwater vehicles: Real-time experiments using computer vision, *IEEE Robotics and Automation Letters* 4 (2) (2019) 1351–1356.
- [36] X.-J. Mao, C. Shen, Y.-B. Yang, Image restoration using convolutional auto-encoders with symmetric skip connections, arXiv:1606.08921.
- [37] K. Panetta, C. Gao, S. Agaian, Human-visual-system-inspired underwater image quality measures, *IEEE Journal of Oceanic Engineering* 41 (3) (2016) 541–551.
- [38] Y. R. Petillot, G. Antonelli, G. Casalino, F. Ferreira, Underwater robots: From remotely operated vehicles to intervention-autonomous underwater vehicles, *IEEE Robotics Automation Magazine* 26 (2) (2019) 94–101.
- [39] Q. Qi, Y. Zhang, F. Tian, Q. M. Jonathan Wu, K. Li, X. Luan, D. Song, Underwater image co-enhancement with correlation feature matching and joint learning, *IEEE Transactions on Circuits and Systems for Video Technology* 32 (3) (2022) 1133–1147.
- [40] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: *Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234–241.
- [41] C. Rother, T. Minka, A. Blake, V. Kolmogorov, Cosegmentation of image pairs by histogram matching-incorporating a global constraint into mrfs, in: *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, 2006, pp. 993–1000.
- [42] P. J. B. Sánchez, M. Papaelias, F. P. G. Márquez, Autonomous underwater vehicles: Instrumentation and measurements, *IEEE Instrumentation Measurement Magazine* 23 (2) (2020) 105–114.
- [43] X. Sun, L. Liu, Q. Li, J. Dong, E. Lima, R. Yin, Deep pixel-to-pixel network for underwater image enhancement and restoration, *IET Image Processing* 13 (3) (2019) 469–474.
- [44] P. Trslík, M. Rossi, L. Robinson, C. W. O'Donnell, A. Weir, J. Coleman, J. Riordan, E. Omerdic, G. Dooley, D. Toal, Vision based autonomous docking for work class rovs, *Ocean Engineering* 196 (2020) 106840.
- [45] H. Wang, H. Wu, Q. Hu, J. Chi, X. Yu, C. Wu, Underwater image super-resolution using multi-stage information distillation networks, *Journal of Visual Communication and Image Representation* 77 (2021) 103136.
- [46] L. Wang, G. Hua, R. Sukthankar, J. Xue, Z. Niu, N. Zheng, Video object discovery and co-segmentation with extremely weak supervision, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (10) (2017) 2074–2088.

- [47] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, C. C. Loy, Ergan: Enhanced super-resolution generative adversarial networks, in: European Conference on Computer Vision Workshops, 2019, pp. 63–79.
- [48] Y. Wang, C. Tang, M. Cai, J. Yin, S. Wang, L. Cheng, R. Wang, M. Tan, Real-time underwater onboard vision sensing system for robotic gripping, *IEEE Transactions on Instrumentation and Measurement* 70 (2021) 1–11.
- [49] Y. Wang, J. Zhang, Y. Cao, Z. Wang, A deep cnn method for underwater image enhancement, in: *IEEE International Conference on Image Processing*, 2017, pp. 1382–1386.
- [50] X.-S. Wei, C.-L. Zhang, J. Wu, C. Shen, Z.-H. Zhou, Unsupervised object discovery and co-localization by deep descriptor transformation, *Pattern Recognition* 88 (2019) 113–126.
- [51] J. Wu, T. Yue, Q. Shen, X. Cao, Z. Ma, Multiple-image super resolution using both reconstruction optimization and deep neural network, in: *IEEE Global Conference on Signal and Information Processing*, 2017, pp. 1175–1179.
- [52] H. Yang, F. Tian, Q. Qi, Q. J. Wu, K. Li, Underwater image enhancement with latent consistency learning-based color transfer, *IET Image Processing* 16 (6) (2022) 1594–1612.
- [53] M. Yang, J. Hu, C. Li, G. Rohde, Y. Du, K. Hu, An in-depth survey of underwater image enhancement and restoration, *IEEE Access* 7 (2019) 123638–123657.
- [54] M. Yang, A. Sowmya, An underwater color image quality evaluation metric, *IEEE Transactions on Image Processing* 24 (12) (2015) 6062–6071.
- [55] W. Yang, X. Zhang, Y. Tian, W. Wang, J.-H. Xue, Q. Liao, Deep learning for single image super-resolution: A brief review, *IEEE Transactions on Multimedia* 21 (12) (2019) 3106–3121.
- [56] X. Ye, Z. Li, B. Sun, Z. Wang, R. Xu, H. Li, X. Fan, Deep joint depth estimation and color correction from monocular underwater images based on unsupervised adaptation networks, *IEEE Transactions on Circuits and Systems for Video Technology* 30 (11) (2020) 3995–4008.
- [57] X. Ye, H. Xu, X. Ji, R. Xu, Underwater image enhancement using stacked generative adversarial networks, in: *Advances in Multimedia Information Processing*, 2018, pp. 514–524.
- [58] D. Zhang, D. Meng, J. Han, Co-saliency detection via a self-paced multiple-instance learning framework, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (5) (2017) 865–878.
- [59] J. Zhang, K. Li, W. Tao, Multivideo object cosegmentation for irrelevant frames involved videos, *IEEE Signal Processing Letters* 23 (6) (2016) 785–789.
- [60] H. Zhao, O. Gallo, I. Frosio, J. Kautz, Loss functions for image restoration with neural networks, *IEEE Transactions on Computational Imaging* 3 (1) (2017) 47–57.
- [61] J. Zhou, T. Yang, W. Chu, W. Zhang, Underwater image restoration via backscatter pixel prior and color compensation, *Engineering Applications of Artificial Intelligence* 111 (2022) 104785.
- [62] Zhou Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity, *IEEE Transactions on Image Processing* 13 (4) (2004) 600–612.
- [63] H. Zhu, F. Meng, J. Cai, S. Lu, Beyond pixels: A comprehensive survey from bottom-up to semantic image segmentation and cosegmentation, *Journal of Visual Communication and Image Representation* 34 (2016) 12 – 27.
- [64] P. Zhuang, C. Li, J. Wu, Bayesian retinex underwater image enhancement, *Engineering Applications of Artificial Intelligence* 101 (2021) 104171.
- [65] K. Zuiderveld, Contrast limited adaptive histogram equalization, *Graphics Gems* (1994) 474–485.