Neurocomputing 596 (2024) 127909

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Breaking the water dilemma: Transmission-guided bilevel adaptive learning for underwater imagery

Sihan Xie [a], Peiming Li [a], Jiaxin Gao [a], Ziyu Yue [b], Xin Fan [a,*], Risheng Liu [a,c]

[a] School of Software Technology, Dalian University of Technology, Dalian 116024, China
[b] School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China
[c] The Peng Cheng Laboratory, Shenzhen, 518052, China

ARTICLE INFO

Communicated by L. Celona

MSC:
00-01
99-00

Keywords:
Underwater image enhancement
Image super-resolution
Bilevel optimization
Loss search
Deep learning

ABSTRACT

Simultaneously enhancing the visual effects and resolution of underwater images poses a challenging task as it involves two types of image enhancement tasks, underwater image enhancement and image super-resolution. In spite of the emergence of various deep learning models, almost all existing methods are tailored to one specific enhancement task, rendering them unsuitable for super-resolution in underwater scenes, which consequently resulting in color distortion, unpleasant artifacts and missing high-frequency details. To address this challenge, we propose a multi-level degradation removal enhancer that utilizes underwater transmission prior to improve the quality of underwater images, dubbed as *SimUESR*. Specifically, the proposed *SimUESR* is designed to be guided by multiple sets of transmission-inspired guidance, which are cascaded with multi-stage degradation removal modules via a feature modulation operation. Through this, the underwater prior is used as modulation information to modulate contrast and color deviation, gradually embedded through the transmission-guided modules at the feature level. Then the enhanced features are incorporated into a multi-level degradation removal module to generate lossless image content. To release the burden of manually designing loss, we introduce a novel bilevel adaptive learning strategy that combines finite-difference approximation to automatically search for the desired loss, effectively improving visual perception performance. The experimental results demonstrate the remarkable superiority of the proposed method for underwater enhancement and super-resolution tasks, achieving improvements of 0.57 dB and 2.87 dB in PSNR on the UFO-120 and EUVP datasets, respectively. The code is available at https://github.com/lpm1001/SimUESR.

1. Introduction

Improving the perceptual quality of underwater imagery is a pragmatic and yet challenging task, with a wide range of promising applications in underwater robotics [3], underwater object detection [4], underwater joint luminance-chrominance learning [5], etc. However, diverse water types exhibit various degrees of color distortion, impurities and contrast reduction phenomena in the captured underwater images in the field of underwater vision [6–9], making it particularly challenging to restore the perceptual quality of images (e.g., visibility and high-frequency detail recovery). In addition, Optical images are prone to suffering signal attenuation and resolution loss during the acquisition, compression, and transmission processes, resulting in various degradation issues such as low resolution and image blurring, which hinder the implementation of high-level downstream tasks. Hence, it is natural to consider joint resolution of multiple degradation tasks [5,10–13]. However, the existing super-resolution methods [11–13] are not yet applicable to complex underwater scenes. According to this, this paper focuses on utilizing underwater physical priors to address the joint task of enhancing and resolution improvement of underwater images.

Existing algorithms in the field of underwater image enhancement often exhibit unstable performance when processing different underwater datasets due to different water depth, light conditions and water types [14]. Traditional methods can be divided in the light of whether they are based on physical models. Non-physical algorithms focus on adjusting the pixel points of the image, while physical algorithms account for the degraded image formation process based on underwater imaging principles. Among physical algorithms, the underwater dark channel algorithm is a prominent technique [15,16]. By calculating the scattering and absorption effects in the light propagation path, the color and brightness information of the original image can be recovered. However, adjusting parameters for different environmental conditions limits the applicability of the algorithm. Given the challenges in underwater imaging task and the limitations of existing methods, we propose

* Corresponding author.
E-mail address: xin.fan@dlut.edu.cn (X. Fan).

https://doi.org/10.1016/j.neucom.2024.127909
Received 7 November 2023; Received in revised form 31 March 2024; Accepted 19 May 2024
Available online 23 May 2024

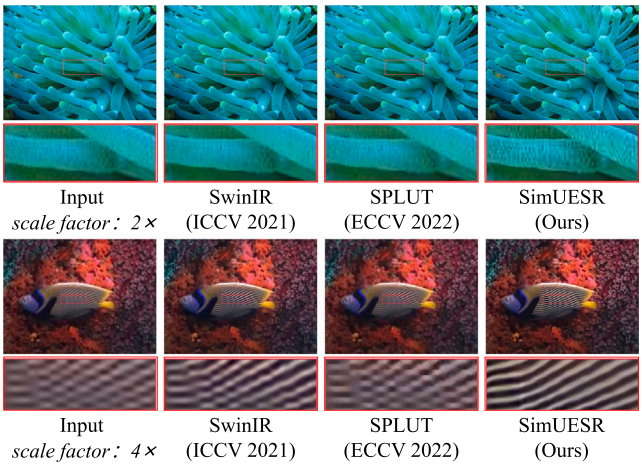


Fig. 1. Visual results of state-of-the-art methods (*i.e.,* SwinIR [1] and SPLUT [2]) and our *SimUESR* on the *USR-248* dataset. For the sake of aesthetics, the size of the input image has been enlarged to the output size.

the introduction of an underwater physical prior to guide the training of deep learning models.

In recent years, various super-resolution reconstruction techniques have been used to restore richer high-frequency details from low resolution images to improve image quality. They have been widely used in remote monitoring, satellite images, and medical imaging fields. However, enhancing the resolution and visibility of underwater images in a single model has not been deeply explored as a practical task. Generally, previous methods for super-resolution were designed for terrestrial images without considering the physical laws of underwater scenes, which directly migrate to underwater scenes with poor recovery results. For instance, some recent efficient super-resolution methods [1,13,17] have introduced advanced network modules (e.g., residual structures, attention mechanism, transformer, etc.) for performance improvement, but have failed to achieve the expected requirements for recovery in underwater super-resolution. For example, SR-VAE [13] employs the VAE network to tackle the task of image super-resolution and proposes a conditional sampling mechanism, but it still fails to deal with the blurring effect in VAE-based sample generation for underwater images. Despite the recent emergence of methods [18] conducting underwater image enhancement and super-resolution, however, they mainly rely on designing the network empirically, ignoring the inherent physical properties and prior design of the underwater scene, making it difficult to achieve the desired enhancement performance. As shown in Fig. 1, directly applying existing super-resolution methods to underwater scenes often leads to color distortion, blurry, and fails to recover the visibility of degraded images. Specifically, SwinIR [1] and SPLUT [17], as advanced methods for image enhancement and super-resolution in regular ground scenes, tend to suffer from luminance deficits, introduce unpleasant artifacts, and fail to generate clear high-frequency texture details. (zoom in to view). In contrast, by introducing underwater physical priors and multi-scale feature enhancement networks, our approach achieves the best visual quality and overcomes the blurring effect, especially in terms of detail recovery, denoising and smoothing. In this regard, we infer that the root cause of these shortcomings can be attributed to the fact that the pre-defined probabilistic distribution assumptions underlying existing advanced super-resolution methods are tailored specifically for single, terrestrial degradation factors, without universality to complex underwater environments.

The previous introduction enables us to leverage the advantages of physics-based models in conjunction with deep learning-based networks. Inspired by Ucolor [19], we use the transmission guidance map as an underwater prior that implies depth and contrast transformation. The pixel values of the transmission guidance map represent the ratio of scene brightness not captured by the camera after being reflected by the medium, indicating that areas with more severe degradation in underwater images will receive larger weight values [19]. It is mapped to a higher dimensional feature dimension through a prior feature extraction module composed of multiple convolutional and activation layer units, serving as auxiliary information with attention function and discrimination mechanism, concatenated with input information. In the Transmission Guidance Module, the input information focuses on areas with severe spatial degradation. Afterwards, the depth and contrast information contained in the prior can be adaptively embedded into multi-level degradation removal modules, extracting fine-grained features from three branches and capturing contextual information, avoiding background interference. SimUESR consists of four groups of transmission guidance modules and multi-level degradation removal modules. The reasons of using the module groups are two-fold: (1) Transmission guidance module can adaptively learn image-specific contrast and color information, which can extract discriminative features and restore perceived image quality. (2) Multi-level degradation removal module transforms features extracted from former modules to ensure the sufficient utilization and correlation between features from multiple scales and enriches the diversity of features. Furthermore, due to the difficulty in acquiring underwater reference images, both supervised and unsupervised loss functions may be required when training the model. The selection of loss functions and hyper-parameter settings not only require professional knowledge in the relevant field, but also consume time and resources. Based on this, we also design a bilevel adaptive learning strategy to fully automate the selection of loss functions and hyper-parameter learning. A finite-difference gradient training strategy is introduced for further performance improvement while ensuring the stability of the training. In summary, our contributions are as follows:

- We innovatively develop an enhancer for simultaneously improving resolution and enhancement for underwater imagery within a singular model, guided by incorporating underwater physical priors (i.e., transmission mapping) that can deliver color-compensated information for high-frequency detail recovery.
- By bringing in feature modulation operations as an intermediate bridge, depth information is progressively embedded as compensation in a multi-scale residual-based feature stream for modulating contrast and color deviation on a pixel-by-pixel basis.
- Instead of constructing a conventional single-level optimization model utilizing a naive alternating strategy, we propose a novel bilevel adaptive learning strategy that constructs efficient solution algorithms (i.e., finite-difference approximation) to automatically search for the desired loss, avoiding the tedium of empirical loss-dependent selection.
- Extensive and sufficient experiments have fully validated the effectiveness of our methodology, with significant advantages for both underwater enhancement and underwater super-resolution dataset benchmarking. The devised novel learning strategy is also effective in terms of visual perception performance and training automation.

The remainder of this paper is organized as follows. Section 2 is dedicated to exploring the related work of underwater image enhancement and super-resolution. The proposed method and learning strategy are outlined in Section 3, which includes the transmission guidance module, multi-level degradation removal module and bilevel adaptive learning. Section 4 entails a comprehensive demonstration of the experiments and ablation studies conducted in support of the proposed method. Finally, we conclude the paper in Section 5.

2. Related work

Over the past decade, the field of underwater image enhancement and super-resolution has seen substantial progress and technological breakthroughs, largely attributed to the significant innovation in deep learning techniques. In this segment, we present a comprehensive review of the recent progress achieved in these two types of image processing tasks.

2.1. Image super-resolution

The Super-Resolution (SR) task focuses on the reconstruction of high-resolution imagery from corresponding low-resolution inputs. In recent times, numerous methods have surfaced in the domain of SR, utilizing both CNN-based and GAN-based models. Utilizing the generative adversarial network in training SRResNet [20] has resulted in more natural and visually appealing images. In order to expand the model size to improve the result quality, methods, such as ESRGAN [21], remove the batch normalization layer in SRResNet for the sake of stacking more layers to extract more features. Nevertheless, the employment of GAN-based training methods leads to a trade-off between image vision and the accuracy of the Peak Signal-to-Noise Ratio (PSNR) metric, potentially resulting in image distortion. To overcome this issue, a lightweight convolutional neural network, known as PAN [17], has been developed, which constructs an attention map on the 3D dimension utilizing pixel attention. Compared to channel and spatial attention, pixel attention requires fewer additional conditions while still achieving superior super-resolution effects. As mentioning the attention mechanism, inspired by Swin transformer [22], Liang et al. [1] introduce SwinIR as a foundational model for tackling a range of image restoration tasks including SR. The shift window mechanism employed within the Residual Swin Transformer Block (RSTB) enables effective modeling of long-range dependencies except for time-consuming and memory consuming aspects. A recent work, SPLUT [2] proposes cascaded lookup tables (LUTs) to expand the field of perception in the field of SR and proposes a new parallel network in an effort to compensate for the loss of accuracy due to discretization. In addition, it is worth mentioning that Islam et al. [23] train SRDRM and SRDRM-GAN on USR-248, an underwater super-resolution dataset, which is the first model specially used for underwater super-resolution.

2.2. Underwater image enhancement

In recent years, there has been a notable increase in the development of techniques aimed at enhancing the quality of underwater imagery. These methods can be broadly classified into three categories: physical methods, non-physical methods and deep learning methods.

Physical Method and Non-physical Method: Physical methods aim to estimate the parameters of the model established for the underwater imaging process, represented by the underwater dark channel prior (UDCP) [15] and the generalized dark channel prior (GDCP) [16] modified from the dark channel prior (DCP) [24] suitable for underwater scenes. These methods assume the existence of pixels in local areas of the image where at least one of the channels has a very low luminance value. In addition, a histogram distribution prior-based contrast enhancement algorithm is proposed by Li et al. [25] that can effectively improve contrast and brightness. However, physical methods are often sensitive to changes in the scene due to the limitations of the parameter assumptions. On the other hand, non-physical methods modify pixel-level values to restore image quality and do not take the imaging process into account. To improve the visibility of distant targets, two versions of the raw underwater images are defined by Ancuti et al. [26] for color correction and contrast enhancement by applying fusion principle. Fu et al. [27] decompose the underwater enhancement problem into two subproblems: contrast enhancement and color correction. However, non-physical methods are prone to losing details and introducing artifacts.

Deep Learning Method: In 2017, Perez et al. [28] introduce deep learning to underwater image enhancement tasks. Thereafter, WaterNet [29] demonstrates a strong generalization capability, while it fails to address the issue of backscattering. UWCNN [30] jointly optimizes the MSE and SSIM loss, preserving image structure and texture besides enhancing clarity. Meanwhile, Ucolor [19] proposes multi-color space coding guided by medium transmission rate. By combining the characteristics of different color spaces [19,31,32], the diversity of characteristics is enhanced. And the attention mechanism is used to adaptively integrate and focus on the most characteristic features extracted from different color spaces. However, the accuracy of these models is limited by the lack of underwater datasets, and this issue has been partially addressed by generative adversarial network. Recently, GAN-based network models such as WaterGAN [33] have emerged as promising solutions.

2.3. Joint tasks

In recent years, it has become common to solve multiple degradation tasks jointly in the interest of solving more problems with less consumption [34,35]. The field of image restoration is witnessing rapid development in algorithms that combine multiple subtasks to form more complex image restoration processes [36–38]. With the increasing demand for image clarity, integrating various image enhancement tasks with super-resolution [39,40] has become a meaningful endeavor. Islam et al. [18] introduce simultaneous enhancement and super-resolution (SESR) problem for the first time and propose Deep SESR model to address it. The USR-248 [23] and UFO-120 [18] datasets are underwater datasets acquired by sensors for underwater super-resolution task. With these datasets, we seek to fill the gap between underwater image enhancement and super-resolution.

3. The proposed method

In this section, we provide an overview of the entire pipeline (Section 3.1), followed by the introduction of two proposed modules, namely Transmission Guidance Modules (TGM) (Section 3.2) and Multi-level Degradation Removal Modules (MDRM) (Section 3.3). We subsequently present the candidate loss search space and bilevel adaptive learning strategy in Section 3.4 and Section 3.5, respectively.

3.1. Overview

We describe the *SimUESR* framework in this section. The overall framework of *SimUESR* is depicted in Fig. 2, which consists of multiple groups of TGM and MDRM for recovering underwater images at the feature level, and the upscale module controls the magnification of high-resolution images. In this regard, the features of the TGM branch are capable of providing abundant contrast and color information in underwater scenes, which are then incorporated into multi-scale feature fusion refinement block to maintain consistency in contrast and color between the generated images and ground-truth images. The implementation of this result depends on the feature modulation operation as an intermediate bridge to cascade with the multi-scale degradation removal module. The MDRM, on the other hand, focuses on modifying the features that are extracted from previous modules to ensure that they are optimally correlated and utilized across various scales. It achieves this by interacting with the spatial features and eliminating complex degradation. Besides, in terms of learning strategy, we avoid the trial-and-error manual selection of loss and empirical design of learning strategy, but introduce a new bilevel adaptive learning strategy. Based on a hierarchical bilevel optimization framework, we introduce a finite-difference approximation algorithm combined with an early stopping mechanism to automatically search for the desired loss. In the following, we provide a detailed description of the proposed key network components and learning strategy.

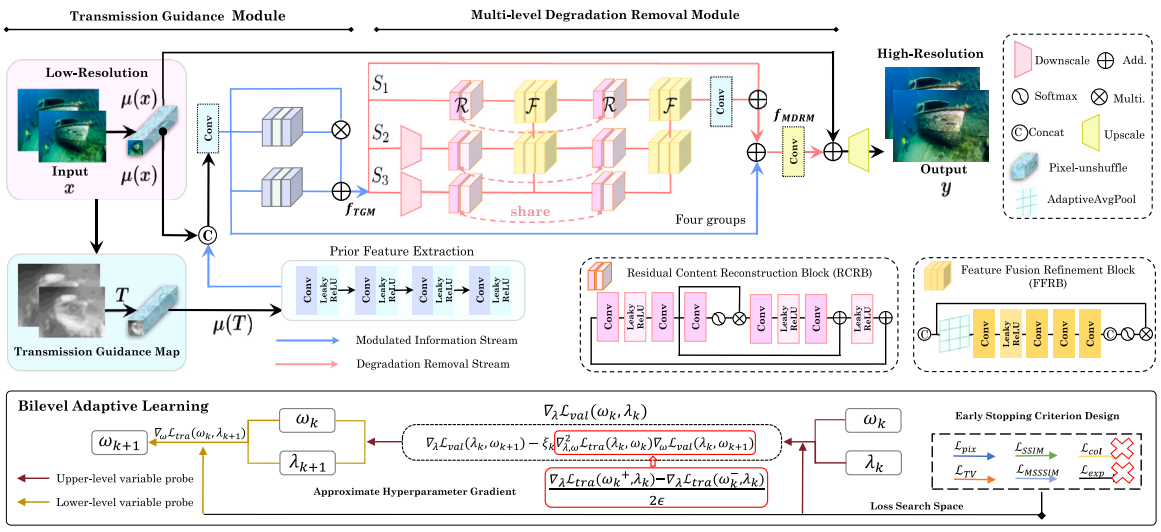


Fig. 2. Network architecture (*Above*) and learning strategy (*Below*) of the proposed *SimUESR* framework. (*Above*) It mainly consists of multiple groups of Transmission Guidance Modules (TGM) and Multi-level Degradation Removal Modules (MDRM), serving to provide rich features for enhancing high-frequency details at the feature level. (*Below*) A hierarchical bilevel adaptive learning strategy is constructed to automate the process of searching for expected losses from the candidate search space, and learning parameters (including network parameters ω and weight hyper-parameters λ).

3.2. Transmission guidance module

The classic underwater image formation model indicates that the light collected by imaging equipment is composed of reflected light from the target object and the remaining atmospheric light after absorption and scattering. However, due to the complex underwater environment, after being absorbed by water, the light scattered by suspended and particulate matter can be lost, resulting in blurred images and reduced contrast. Learned from several classical underwater image restoration algorithms [16,19,41–44], the degraded underwater image is formed as follows, that is, the principle of underwater imaging:

$$I^c(x) = G^c(x) \otimes T(x) \oplus H^c(x) \otimes (1 - T(x)), \tag{1.1}$$

where I is the degraded image, while G refers to the corresponding clear ground truth. And x indicates each individual pixel, T is the transmission rate which represents the proportion of radiant light from the scene that reaches the camera. Furthermore, c denotes one of the RGB channels, while H represents the value of a homogeneous background pixel.

To guide the model in adjusting color deviation and enhancing low contrast, we obtain a transmission guidance map from underwater images, which includes scene brightness and depth information as prior information. Specifically, we obtain the transmission guidance map from the transmission rate, which is observed by $1 - T$ ($1 \in H \times W$ is a full 1 matrix). It can represent the proportion of radiation not captured by the camera, indirectly reflecting the degree of image damage. This means that this prior information is added to subsequent restoration modules to guide the algorithm to pay more attention to areas with severe damage. Although the corresponding ground truth T required for estimating T is unavailable. Inspired by the physical underwater image restoration algorithm [42] and deep learning method Ucolor [19], the transmission rate can be acquired by:

$$\tilde{T}(x) = max_{c,y \in \Phi(x)}\left(\frac{H^c - I^c(y)}{max(H^c, 1 - H^c)}\right), \tag{1.2}$$

where $\tilde{T}$ denotes the estimation of transmission rate, I denotes the degraded image, H is the value of homogeneous background pixel, c is one of RGB channels and $\Phi(x)$ represents a x-centered local window. In [19], the value of homogeneous background pixel is estimated from the color variation caused by the variation in depth in the underwater scene. These depth information contained in the transmission guidance map can correct the blurring and contrast reduction caused by scattering effects, and estimate the propagation path and color change degree of light underwater to correct color deviation. Utilizing the transmission guidance map as a prior effectively utilizes rich underwater information.

We construct an adaptive transmission guidance module that transmits both prior and input information to subsequent degradation removal modules. We firstly employ the pixel-unshuffle operation μ to reduce the spatial pixel size and expand the channel size before feeding inputs x into feature fusion nested block. Thereafter, we append a prior feature extraction block $\mathcal{P}(\cdot)$ to get discriminative features $f_{\texttt{latent}}$ from transmission guidance map, which is transmitted to concat with the unshuffled features $f_{\texttt{unshuffle}}$ of input to perform further feature aggregation, followed by:

$$f_{\texttt{unshuffle}} = \mu(x), \quad f_{\texttt{latent}} = \mathcal{P}(\mu(\tilde{T})), \quad f_{\texttt{concat}} = \mathcal{C}(f_{\texttt{unshuffle}}, f_{\texttt{latent}}),$$

Subsequently, the features refined by the TGM are represented as

$$f_{\texttt{conv}} = \texttt{Conv}(\texttt{LReLu}(\texttt{Conv}(f_{\texttt{concat}}))), \quad f_{\texttt{TGM}} = f_{\texttt{conv}} \times f_{\texttt{concat}} \oplus f_{\texttt{conv}},$$

where $\mathcal{C}$ denotes the concatenation.

3.3. Multi-level degradation removal module

In the realm of underwater imaging, complex scenes present a significant challenge to the successful execution of super-resolution tasks. This is due to the distinct physical laws at play, which give rise to selective attenuation of light of varying wavelengths, leading to color deviations, blurriness, and limited visibility. The multi-level degradation removal module aims to capture contextual information in spatial dimensions and generate fine-grained details for high-resolution output images while simultaneously removing complex degradation. To achieve this, the module leverages color and contrast information from low-resolution representations, thus enhancing the quality of underwater super-resolution output. Specifically, we adopt the multi-scale residual content reconstruction block as the degradation removal module to extract multi-scale features in three parallel branches. In this manner, the multi-scale residual content reconstruction block consists of a series of convolutional operations and non-linear activation function Leaky-ReLU with skip connections, which facilitates an improved hierarchical feature learning. Then feature maps from different branches are passed to the feature fusion refinement block to perform

feature aggregation. Eventually, the convolutional streams from each branch are added at the pixel level and aggregated into f_{MDRM}. The whole process is as follows:

$$S_4 = \mathcal{F}(\mathcal{R}(S_2), \mathcal{R}(S_3)),$$
$$S_5 = \mathcal{F}(\mathcal{R}(S_4), \mathcal{R}(\mathcal{R}(S_3))),$$
$$S_{out} = \mathcal{F}(\mathcal{R}(\mathcal{F}(\mathcal{R}(S_1), S_4)), S_5),$$
$$f_{\mathrm{MDRM}} = Conv_{3\times3}(S_{out}) \oplus f_{\mathrm{TGM}},$$

where S_1, S_2, S_3 are extracted from different scales by applying downscale to f_{TGM}, $\mathcal{F}(\cdot)$ denotes the feature fusion refinement block and $\mathcal{R}(\cdot)$ denotes the residual content reconstruction block. Our motivation for designing this module is to maintain a global attention while having the capacity to learn local information characteristics to ensure better performance and preserve the efficiency.

3.4. Automatic searching objectives

In the pursuit of developing an efficient loss function, significant computational resources are typically required, which can be a costly trade-off for human efforts. To overcome this limitation, a process of hyper-parameter optimization has been proposed to automate the optimization of hyper-parameters and eliminate the need for manual intervention during the training process. In this study, we propose a novel approach that involves designing a training loss search space comprising image quality-oriented losses. This search space enables the exploration and identification of optimal combination of the losses to replace hand-crafted losses. By leveraging this approach, we can automate the search for an effective loss function without compromising computational efficiency, thus enhancing the overall training process. The preliminary candidate losses include L1, color [45], SSIM [46], exposure [47], MSSSIM [46] and TV [45] loss. The end-to-end training of *SimUESR* is supervised by six candidate loss functions mentioned to learn the function G: $X \rightarrow \hat{Y}$. $\hat{Y}$ is the generated output and Y refers to the target. By leveraging this multi-loss supervision strategy, we can effectively optimize the model's performance and improve the quality of the output. Specifically, the use of multiple loss functions provides the model with diverse information during the training process, enabling it to learn and adapt to different types of input data, leading to enhanced generalization and robustness of the model. For each loss, we define the weighting coefficient as the hyper-parameter to be optimized. The overall loss function can be expressed as:

$$\mathcal{L}_{total} = \sum_{i=1}^{6} \lambda_i * \mathcal{L}_i,$$

where $\mathcal{L}_i$ is $\mathcal{L}_{pix}$, $\mathcal{L}_{TV}$, $\mathcal{L}_{SSIM}$, $\mathcal{L}_{MSSSIM}$, $\mathcal{L}_{col}$, $\mathcal{L}_{exp}$, respectively, and λ_i is the coefficient corresponding to $\mathcal{L}_i$.

(1) $\mathcal{L}_{pix}$: Firstly, we choose the widely-used L_1 loss as our pixel loss $\mathcal{L}_{pix}$, defined as follows:

$$\mathcal{L}_{pix} = \|\hat{Y} - Y\|_1, \tag{2.1}$$

(2) $\mathcal{L}_{col}$: In order to correct the color deviation, we introduce color constancy loss as the supervision. The primary objective of this loss is to rectify the color deviation of the output image while establishing meaningful connections between the RGB channels through appropriate adjustments, which can be defined as [45]:

$$\mathcal{L}_{col} = \sum_{\forall(m,n)\in\varepsilon} (\tilde{Y}^m - \tilde{Y}^n)^2, \varepsilon = \{(R,B),(G,B),(R,G)\}, \tag{2.2}$$

where $\tilde{Y}^m$ is the average pixel value of m channel in the generated image, (m,n) is a paired channel selection scheme.

(3) $\mathcal{L}_{\mathcal{TV}}$: In the restoration process, noise may be introduced, and some regular items need to be added to maintain the smoothness of the image [45]. We add the most widely used smoothness loss to reduce noise in each enhanced image $\hat{Y}$:

$$\mathcal{L}_{\mathcal{TV}} = \sum_{c\in\Pi} (|\nabla_x \hat{Y}^c| + |\nabla_y \hat{Y}^c|), \Pi = \{R,G,B\}, \tag{2.3}$$

where ∇_x and ∇_y represent gradient operation in the horizontal direction and gradient operation in the vertical direction, respectively. It can remove salt and pepper noise without affecting high-frequency components.

(4) $\mathcal{L}_{exp}$: Exposure control loss suppresses the low exposure or high exposure area, which is the distance between the good exposure E and the average pixel value of the local area [47].

$$\mathcal{L}_{exp} = \frac{1}{M}\sum_{k=1}^{M} |\tilde{Y}_k - E|, \tag{2.4}$$

where $\tilde{Y}$ is the average pixel value of each window; E is the corresponding grayscale in RGB space; M denotes the number of non-overlapping local windows.

(5) $\mathcal{L}_{SSIM}$: To counteract the common L1 distance's inability to measure the structural similarity of images, we adopt SSIM loss. SSIM focuses on the luminance, contrast and structure of the image, which can be defined as follows [46]:

$$\mathcal{L}_{SSIM}(\hat{y}, y) = \frac{(2\mu_{\hat{y}}\mu_y + c_1)(2\sigma_{\hat{y}y} + c_2)}{(\mu_{\hat{y}}^2 + \mu_y^2 + c_1)(\sigma_{\hat{y}}^2 + \sigma_y^2 + c_2)}, \tag{2.5}$$

in which $\hat{y}$ denotes the generated image and y denotes the ground truth. μ, σ^2 denote the average and variance of the image, and $\sigma_{\hat{y}y}$ is the covariance of $\hat{y}$ and y. Constants c_1 and c_2 are set to $(k_1 L)^2$ and $(k_2 L)^2$, which are used to maintain stability. L is the dynamic range of pixels. In the experiments, we set k_1 and k_2 to 0.01 and 0.03.

(6) $\mathcal{L}_{MSSSIM}$: Multi-Scale Structural Similarity is a multi-scale SSIM index, which can be defined as follows [46]:

$$\mathcal{L}_{MSSSIM}(\hat{y}, y) = 1 - \prod_{m=1}^{M} \left(\frac{2\mu_{\hat{y}}\mu_y + c_1}{\mu_{\hat{y}}^2 + \mu_y^2 + c_1}\right)^{\beta_m} \left(\frac{2\sigma_{\hat{y}y} + c_2}{\sigma_{\hat{y}}^2 + \sigma_y^2 + c_2}\right)^{\gamma_m}, \tag{2.6}$$

in which M refers to different scales, width and height are reduced by factor 2^{M-1}, and β_m, γ_m represent the relative importance of two items.

3.5. Bilevel adaptive learning

Automated machine learning techniques are increasingly emerging, aiming to automate the process of learning hyperparameters and network parameters [48,49]. Inspired by this, we construct the hierarchical bilevel optimization framework to accomplish the process of automating the search for expected losses and parameter learning. Drawing on the ideas of meta-learning and hyper-parameter optimization [34,50], we firstly introduce the constrained coupling relationship of two variables based on the bilevel optimization formulation [51,52], as follows:

$$\min_{\lambda\in\Lambda} F(\lambda, \omega),\ s.t.\ \omega \in S(\lambda),$$
$$\mathrm{where}\ S(\lambda) = \arg\min_{\omega\in\Omega} f(\lambda, \omega),\ (\text{parameterized by } \omega), \tag{3.1}$$

where F and f are the optimization objective functions of the outer variable λ and inner variable ω, respectively. $S(\lambda)$ denotes the optimal solution set of inner problem parameterized by ω. Ω and Λ are feasible domains of variables ω and λ, respectively. During the training period, we denote ω and λ as network weights and auxiliary coefficients under the bilevel optimization framework.

(1) Hierarchical Bilevel Reformulation: The training and the validation loss for training set $\mathcal{D}_{tra}$ and validation set $\mathcal{D}_{val}$ are denoted by $\mathcal{L}_{tra}$ and $\mathcal{L}_{val}$, respectively [53]. In this case, the lower level and upper level learning objective functions are

$$f(\lambda, \omega) = \mathcal{L}_{tra}(\lambda, \omega) = \mathcal{L}_{total}(\lambda, \omega; \mathcal{D}_{tra}),$$
$$F(\lambda, \omega) = \mathcal{L}_{val}(\lambda, \omega) = \mathcal{L}_{total}(\lambda, \omega; \mathcal{D}_{val}). \tag{3.2}$$

Both losses are determined not only by the loss function coefficients λ, but also the weights ω in the network. We minimize the validation loss $F(\omega^*, \lambda^*)$ to find loss function coefficients λ^*, and we minimize the

training loss $\omega^* = argmin_\omega\ f(\lambda, \omega)$ with fixed λ to get the weights ω^* associated with the loss function coefficients.

Based on the idea of explicit gradient approximation [54,55], the core learning objective of λ can be further optimized by:

$$\nabla_\lambda \Gamma(\lambda) = \overbrace{\nabla_\lambda \mathcal{L}_{val}(\lambda, \omega^*(\lambda))}^{\text{Direct Gradient}} + \overbrace{\nabla_\omega \mathcal{L}_{val}(\lambda, \omega^*(\lambda)) \nabla_\lambda \omega^*(\lambda)}^{\text{Coupled Gradient: } G_C}. \tag{3.3}$$

It can be noted that the direct gradient term $\nabla_\lambda \mathcal{L}_{val}(\lambda, \omega^*(\lambda))$ is implemented in most of the gradient based methods [54,55], but it only takes into account the direct dependence on the auxiliary parameters λ. One of the most straightforward means of implementing direct gradients is the commonly used alternating iteration strategy, i.e., by fixing one variable to update another, which usually causes several issues such as low performance, training oscillations, and instability. Motivated by the above, we introduce the coupled gradient term G_C which accurately computes the rate of change of $\omega^*(\lambda)$ with respect to λ, i.e., the underlying coupling between the two variable parameters is linked.

(2) Finite Difference Approximation: Since inner-level optimization involving multiple iterations can lead to significant memory consumption, we introduce simplified and elegant gradient approximation techniques to circumvent this problem. Specifically, for a fixed initialization ω_0, the loop structure of the gradient computation is performed, i.e., $\omega_{k+1} = \omega_k - \eta_k \nabla_\omega \mathcal{L}_{tra}(\lambda_k, \omega_k),\ k = 1, \ldots, K$, where η_k denotes the corresponding step size and K denotes the overall lower-level iterations number. Then we have $\nabla_\lambda \Gamma(\lambda_k) = \nabla_\lambda \mathcal{L}_{val}(\lambda_k, \omega_k(\lambda_k)) + \nabla_\omega \mathcal{L}_{val}(\lambda_k, \omega_k(\lambda_k)) \nabla_\lambda \omega_k(\lambda_k)$. After designing the further training loss, we obtain a stable solution strategy for further improving the performance of our network. To solve the coupled gradient term G_C efficiently, following the learning manner of differentiable architecture search [53], we conduct the simple approximation scheme as follows:

$$\begin{aligned} \nabla_\lambda \Gamma(\lambda_k) &= \nabla_\lambda \mathcal{L}_{val}(\lambda_k, \omega_{k+1}) \\ &\approx \nabla_\lambda \mathcal{L}_{val}(\lambda_k, \omega_{k+1}) - \xi_k \nabla^2_{\lambda,\omega} \mathcal{L}_{tra}(\lambda_k, \omega_k) \nabla_\omega \mathcal{L}_{val}(\lambda_k, \omega_{k+1}), \end{aligned} \tag{3.4}$$

where $\omega_{k+1} = \omega_k - \eta_k \nabla_\omega \mathcal{L}_{tra}(\lambda_k, \omega_k)$ denotes the weights for one-step forward optimizer, and η_k is the learning rate for k_{th} step of inner optimization. Ultimately, we substantially reduce the complexity of Eq. (3.4) by applying the finite difference approximation:

$$\begin{aligned} &\nabla^2_{\lambda,\omega} \mathcal{L}_{tra}(\lambda_k, \omega_k) \nabla_\omega \mathcal{L}_{val}(\lambda_k, \omega_{k+1}) \\ &\approx \frac{\nabla_\lambda \mathcal{L}_{tra}(\omega_k^+, \lambda_k) - \nabla_\lambda \mathcal{L}_{tra}(\omega_k^-, \lambda_k)}{2\epsilon}, \end{aligned} \tag{3.5}$$

where ϵ is a small scalar depending on learning rate and $\omega_k^\pm = \omega_k \pm \epsilon \nabla_\omega \mathcal{L}_{val}(\omega_{k+1}, \lambda_k)$.

(3) Early Stopping Criterion Design: At early beginning, we adopt the traditional alternating iteration mechanism to generally search loss in training loss search space. Empirically, the update stalls if some coefficient reaches a negative number. Therefore we take the policy to stop iterative procedure within an epoch and remove those losses from the training loss space [56]. When the early stopping criterion is met, it indicates that the parameter search for λ has reached a sufficient point. At this stage, further exploration of the parameter space is deemed unnecessary, as the criterion indicates satisfactory progress in optimizing the loss coefficients. By stopping the search at this point, we can ensure that the optimization process has adequately addressed the desired objectives and achieved a reasonable balance between the loss functions, leading to improved model performance. The iterative procedure is outlined in Alg. 1.

4. Experiment

In this section, we first commence the experimental details. Following that, we conduct a series of comparisons between the proposed and other state-of-the-art methods utilized in tasks pertaining to image enhancement and super-resolution. Finally, we scrutinize the efficacy of our proposed method through comprehensive ablation studies.

Algorithm 1 Bilevel Adaptive Learning

Require: Auxiliary coefficients λ and network weights ω, step size ξ and η, small scalar ϵ, dataset $\mathcal{D}$

```
#Select loss from candidate search space.
Set L_total ← Σ_{i=1}^{6} λ_i * L_i.
#Divide D into D_val and D_tra.
repeat
  while not converged and λ_k > 0 do
    #Upper − level variable probe.
    Update loss function coefficient λ_k by descending ∇_λ L_val(ω_k, λ_k).
    #Apply finite difference approximation.
    ∇_λ Γ(λ_k) ← ∇_λ L_val(λ_k, ω_k) − η_k (∇_λ L_tra(ω_k^+, λ_k) − ∇_λ L_tra(ω_k^−, λ_k)) / 2ε.
    λ_{k+1} ← λ_k − ξ_{k_1} ∇_λ Γ(λ_k)
    #Lower − level variable probe.
    Update network weights ω_k by descending ∇_ω L_tra(ω_k, λ_{k+1}).
    ω_{k+1} ← ω_k − ξ_{k_2} ∇_ω L_tra(ω_k, λ_{k+1})
    k ← k + 1
  end while
  #Stop iterative procedure.
  Remove those losses whose coefficient reaches a negative number, update L_total.
until training convergence.
Derive the final loss function coefficient based on the learned λ and the final weights ω.
```

Table 1
Details of the benchmark datasets.

Dataset	Training	Testing	Paired/Unpaired
USR-248 [23]	1060	248	Paired
UFO-120 [18]	1500	120	Paired
UIEB [57]	712	178	Paired
EUVP [58]	7200	4284	Paired

4.1. Implementation details

Training settings: The training process is conducted on a single NVIDIA RTX A6000 GPU within the PyTorch framework. During training, the ground truth is cropped to sizes of $192 \times 192, 160 \times 160, 128 \times 128$ and the batch size is gradually decreased to 8, 5, 4, 2, 1 for progressive training. Additionally, the data is randomly flipped along each axis with a probability of 50% to augment the training set. We set the initial learning rate of 2×10^{-4} and betas of 0.9 and 0.999 for Adam optimizer [59]. The network converges within 1.5×10^5 iterations, which are split into three cycles. The first cycle comprises 46k iterations with a fixed learning rate of 2×10^{-4}, followed by 58k iterations with a learning rate of 3×10^{-4} for the second cycle. Finally, the learning rate is adjusted to a cosine annealing schedule, ranging from 3×10^{-4} to 1×10^{-6} for the last 46k iterations. A bilevel adaptive learning strategy, which utilizes an optimization with an early-stopping criterion design and finite difference approximation, is employed to search the loss function coefficient. Overall, the proposed training process is characterized by careful selection of hyperparameters and adaptive optimization techniques, which contribute to its efficacy and robustness.

Datasets: Initially, we evaluate the effectiveness of our proposed method on the super-resolution task through the usage of two benchmark underwater datasets, namely USR-248 [23] and UFO-120 [18]. Additionally, we evaluate the robustness of the proposed method for underwater enhancement task by incorporating EUVP [58] and UIEB [57] underwater datasets. The dataset specifications for the aforementioned benchmark underwater image datasets are presented in Table 1.

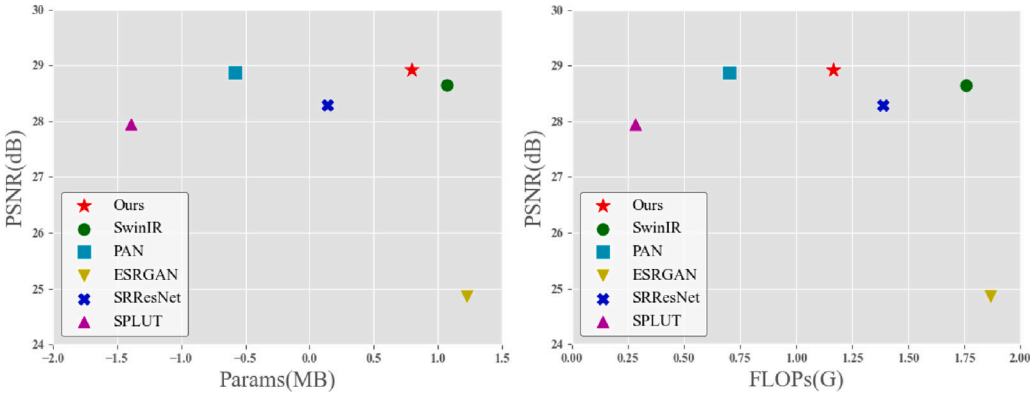


Fig. 3. PSNR (dB) vs. Params (MB) or PSNR (dB) vs. FLOPs (G) among state-of-the-art methods and our *SimUESR*. For aesthetic purposes, we use a log scale on coordinate axis. More details are provided in the Table 2.

Table 2
Model size and FLOPs of state-of-the-art methods and ours on scale 2× and 4× task.

Methods	Scale 2×		Scale 4×	
	Params(M)	FLOPs(G)	Params(M)	FLOPs(G)
SPLUT [2]	0.0403	1.9076	0.0419	1.9819
PAN [17]	0.2614	5.0388	0.2724	8.0477
SRResNet [20]	1.3698	24.3269	1.5175	41.6074
SwinIR [1]	11.7525	57.1544	11.9002	67.1578
ESRGAN [21]	16.7032	73.4952	16.6979	293.8959
Ours	6.2743	14.5779	6.2657	58.1647

Table 3
Quantitative comparison results (PSNR, SSIM, UIQM) of the state-of-the-art methods and ours for 2× and 4× tasks on the *USR-248* dataset. The best result is highlighted in red whereas the second best one is highlighted in blue.

Dataset	USR-248					
Scale	2×			4×		
Metrics	PSNR↑	SSIM↑	UIQM↑	PSNR↑	SSIM↑	UIQM↑
SRDRM [23]	28.0293	0.8017	2.7479	24.6149	0.6364	2.6475
SRGAN [20]	25.0028	0.7682	2.8632	22.3603	0.5569	2.8758
SRResNet [20]	28.2936	0.8169	2.7559	24.7084	0.6463	2.6712
ESRGAN [21]	24.8740	0.7325	3.4736	22.6671	0.5821	3.1754
PAN [17]	28.8887	0.8354	3.0736	25.0874	0.6708	2.5015
SwinIR [1]	28.6547	0.8300	2.6332	25.0146	0.6582	2.2302
SPLUT [2]	27.9581	0.8037	3.2880	24.6361	0.6360	2.8483
Ours	28.9259	0.8437	3.2079	25.2042	0.6850	2.7621

Evaluation Metrics: To provide a comprehensive and fair assessment of existing advanced methods and the proposed method, we employ both reference and non-reference metrics. For assessing the performance of super-resolution, we rely on widely-used metrics including: Peak Signal-to-Noise Ratio (PSNR) [60], Structural Similarity Index Measure (SSIM) [61], and Underwater Image Quality Measure (UIQM) [62] to measure the color balance, sharpness and contrast of images. Additionally, for evaluating the efficacy of the proposed method for underwater image enhancement, we adopt all the metrics (PSNR, SSIM, UIQM and NIQE), among which the role of Natural Image Quality Evaluator (NIQE) is to assess the visual perception effect. In terms of quantifying model parameters and complexity, our enhancement network demonstrates efficiency in both memory usage and computational requirements when compared to recently proposed advanced methods such as SwinIR and ESRGAN (see Fig. 3). The specific data is presented in Table 2.

4.2. Evaluation: Super-resolution

In this section, a comparative analysis of the 2× and 4× SR performance is conducted on two distinct underwater super-resolution datasets, namely USR-248 [23] and UFO-120 [18]. The USR-248 dataset comprises of 1060 pairs of HR-LR images, utilized for training underwater SR models, and a test set consisting of 248 paired images, which is used to evaluate benchmark performance. The benchmark evaluation covers a range of scales, including 2×, 3×, 4×, and 8×. In order to validate the robustness of the proposed method in the underwater environment, the study further employs another underwater dataset, UFO-120. The dataset comprises of a training set with 1500 samples for SESR training and a testing set consisting of 120 samples. The dataset provides 2× images, and we get 4× images by downsampling.

Table 4
Quantitative comparison results (PSNR, SSIM, UIQM) of the state-of-the-art methods and ours for 2× and 4× tasks on the *UFO-120* dataset. The best result is highlighted in red whereas the second best one is highlighted in blue.

Dataset	UFO-120					
Scale	2×			4×		
Metrics	PSNR↑	SSIM↑	UIQM↑	PSNR↑	SSIM↑	UIQM↑
SRDRM [23]	23.8007	0.7048	2.7393	21.5652	0.5622	2.4479
SRGAN [20]	22.1147	0.6549	2.9614	20.8834	0.5355	2.9122
SRResNet [20]	24.1248	0.7092	2.7625	22.8430	0.6063	2.5754
ESRGAN [21]	20.1561	0.6228	2.9665	19.7027	0.5304	2.8332
PAN [17]	24.5707	0.7339	2.7456	23.3952	0.6308	2.2283
SwinIR [1]	24.9232	0.7423	2.6332	23.4949	0.6362	2.2302
SPLUT [2]	20.1057	0.6514	2.0972	19.5453	0.5632	1.6461
Ours	25.4903	0.7680	2.8685	23.6357	0.6600	2.5048

For the evaluation and performance comparison, we take the existing underwater SR model SRDRM [23] into account. We also consider the terrestrial SR models named SRGAN [20], ESRGAN [21], SRResNet [20], PAN [17], SwinIR [1] and SPLUT [2] in the evaluation as benchmarks, among which SPLUT does not provide 2× code, so we get 2× results by changing the downsampling part of the output layer. Based on Tables 3 and 4, it is evident that *SimUESR* exhibits a significant performance advantage over the other models when evaluated quantitatively using PSNR and SSIM metrics, which are widely utilized for assessing the reconstruction quality and structural similarity of enhanced images. Despite the slight disadvantage observed in terms of UIQM, the images generated using our proposed method exhibit exceptional visual appeal.

Conventional super-resolution methods developed for terrestrial images have proven effective in achieving low magnification super-resolution, but often fail to incorporate high-frequency details necessary for attaining high resolution. In contrast, the proposed method yields substantially sharper and higher quality images that closely resemble ground-truth patterns. Qualitative comparisons presented in Figs. 4 and 5 reveal inadequacies in color restoration exhibited by SRDRM, SRGAN, ESRGAN and SRResNet. The resolution of GT is

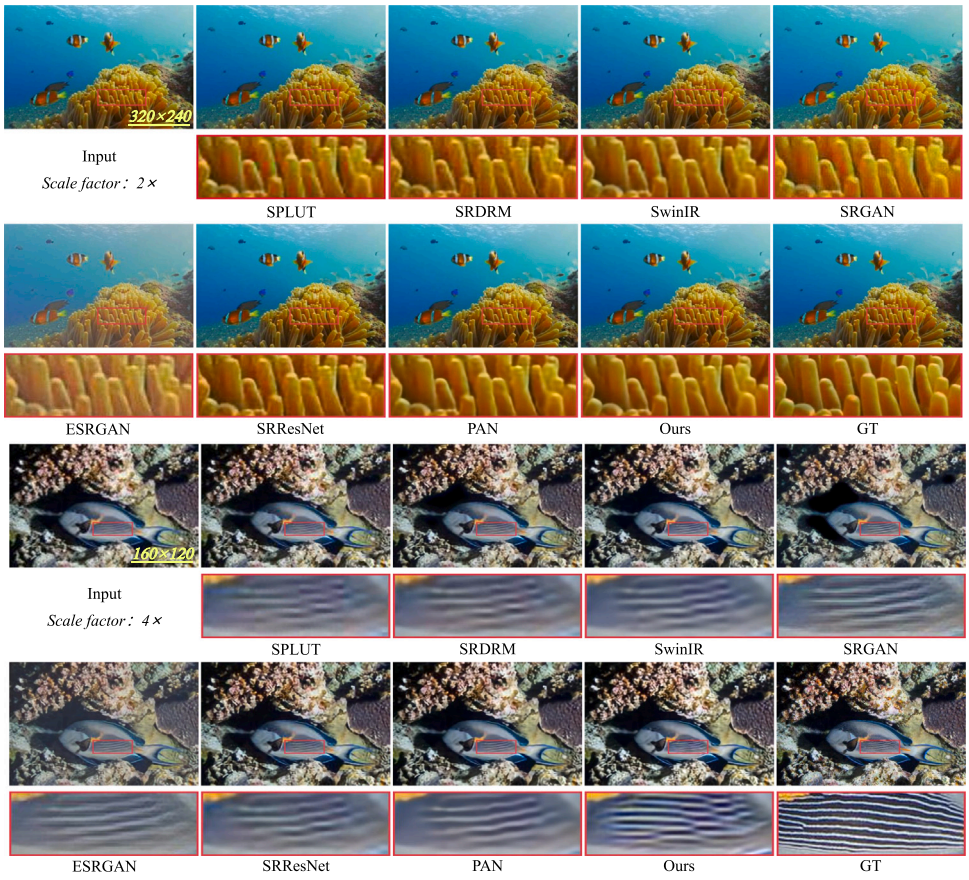


Fig. 4. Visual results of state-of-the-art methods and our *SimUESR* on the *USR-248* dataset. Our method addresses color deviation and blurring issues. For aesthetic purposes, we set input and GT to equal size.

640 × 480. Specifically, in Fig. 4, SPLUT and SRDRM produce color bias due to over-enhancement in both red and green channels, while ESRGAN misses high frequency details. In Fig. 5, SPLUT results in significant color deviation, and SRDRM, SRGAN, and ESRGAN models yield many artifacts. Conversely, the proposed method generates images with superior color consistency and enhanced levels of detail. In terms of both quantitative and qualitative assessments, the proposed method exhibits superiority.

4.3. Evaluation: Enhancement

To assess the robustness of our method in underwater enhancement task, we conduct a comprehensive evaluation of the proposed underwater enhancement method by qualitatively and quantitatively analyzing the color, contrast and sharpness of generated images on two publicly available datasets, namely EUVP [58] and UIEB [57]. We compare our method against several advanced methods including conventional methods EUIVF [26], OCM [25], TSA [27], physical-based methods AIO [63], UDCP [24] and deep learning-based methods UWCNN [30], WaterNet [29], FGAN [64], UGAN [65].

We first show the qualitative comparisons on EUVP and UIEB benchmark in Fig. 6. Our observations indicate that conventional methods such as EUIVF [26], OCM [25], TSA [27] and AIO [63] struggle to effectively restore visibility in underwater images and often produce blurring and unpleasant artifacts. In addition, almost all comparison methods result in a significant decrease in contrast and varying degrees of color deviation, especially on the EUVP dataset and bright color spots appearing on the UIEB dataset. Our proposed method generates underwater images characterized by softer colors, sharper subjects, and more natural textures, while maintaining a balanced contrast. Additionally, the color histogram map reveals that our method achieves the most similar pattern to the ground truth, particularly in the red channel, thus demonstrating the superiority of our method in underwater enhancement scenarios.

Table 5
Quantitative comparison results (NIQE, UIQM, PSNR, SSIM) of state-of-the-art methods and ours on the *EUVP* [58] and *UIEB* [57] datasets in the experiments of single underwater enhancement. The best result is highlighted in red whereas the second best one is highlighted in blue.

Methods	NIQE↓		UIQM↑		PSNR↑		SSIM↑	
	EUVP	UIEB	EUVP	UIEB	EUVP	UIEB	EUVP	UIEB
OCM [25]	4.628	3.877	2.776	2.545	15.62	16.19	0.722	0.759
UDCP [24]	4.398	4.303	2.079	1.772	14.53	11.73	0.716	0.509
EUIVF [26]	4.358	4.059	2.763	2.679	17.06	21.93	0.723	0.823
TSA [27]	5.623	4.165	2.869	1.996	13.21	14.32	0.560	0.641
UGAN [65]	6.467	7.057	3.325	2.528	19.31	17.73	0.795	0.765
WaterNet [29]	4.375	4.484	3.065	2.857	18.68	19.65	0.847	0.824
AIO [63]	4.892	3.994	3.346	3.078	13.85	12.69	0.507	0.466
FGAN [64]	5.175	6.364	3.222	2.512	19.49	18.16	0.848	0.597
UWCNN [30]	4.251	4.441	2.231	3.078	18.37	13.35	0.821	0.665
Ours	4.194	3.117	4.895	3.039	22.36	20.10	0.903	0.637

Table 5 present the quantitative assessment of the proposed method for underwater image enhancement, as measured by four metrics: PSNR, SSIM, NIQE, and UIQM. Our method demonstrates superior performance compared to the state-of-the-art methods on the EUVP dataset, achieving a significant improvement of 2.87 dB in PSNR, 0.055 in SSIM, 0.057 in NIQE, and 1.549 in UIQM.

4.4. Ablation study for training strategy

In order to investigate the impact of the proposed training strategy, we perform an ablation study to assess the contributions of two key

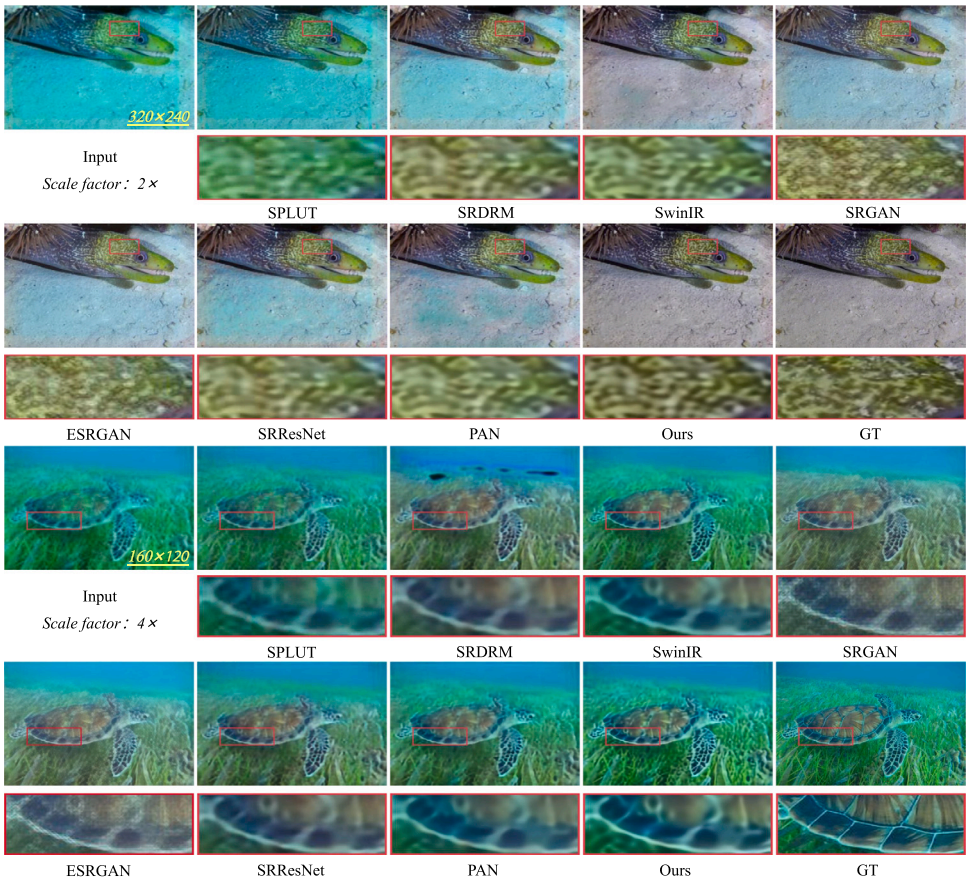


Fig. 5. Visual results of state-of-the-art methods and our *SimUESR* on the *UFO-120* dataset. Our method does not add artifacts. For aesthetic purposes, we set input and GT to equal size.

Table 6
Ablation study with whether to use HBR and FDA on *USR-248* dataset (**2×**).

Setting	Bilevel adaptive learning		PSNR↑	SSIM↑
	HBR	FDA		
S_1	✘	✘	26.33	0.836
S_2	✔	✘	$28.36_{\uparrow 2.03}$	$0.833_{\downarrow 0.003}$
Ours	✔	✔	$\mathbf{28.92}_{\uparrow 2.59}$	$\mathbf{0.0.844}_{\uparrow 0.008}$

elements: Hierarchical Bilevel Reformulation (HBR) and Finite Difference Approximation (FDA). The results in Table 6 demonstrate that the model trained with the bilevel adaptive learning strategy achieves superior quantitative performance compared to the ablated models, indicating the effectiveness of both HBR and FDA either alone or in combination.

(1) Effectiveness of Hierarchical Bilevel Reformulation

To investigate the effect of bilevel adaptive learning strategy, we train our *SimUESR* on USR-248 dataset with the strategy. The experimental results, presented in Table 6 and Fig. 7, indicate that the models trained with the bilevel adaptive learning strategy outperform the native model in terms of PSNR, although slightly less well in terms of SSIM. The outcomes demonstrate that the proposed strategy, which utilizes a loss function with automatically optimized weighting coefficient, is more effective in achieving superior quantitative results than hand-crafted losses.

(2) Effectiveness of Finite Difference Approximation

We explore to reveal how FDA helps. As depicted in Table 6, the results indicate that the PSNR values of our method significantly outperform the naive method. Furthermore, FDA addresses the limitations of SSIM reduction caused by HBR. As illustrated in Fig. 7, the utilization of FDA enhances the training stability and improves the visual perceptual quality scores. These findings suggest that FDA plays a crucial role in optimizing the performance of our proposed method.

4.5. Ablation study for modules

To elucidate the significant impact of each module and assess the functionality of each branch, we perform experiments to validate the role of various modules, including Prior Feature Extraction Block (PFEB), and different numbers of module groups of transmission guidance modules (TGM) and multi-level degradation removal modules (MDRM).

(1) Effectiveness of the Transmission Guidance Module

To assess the contribution of the Transmission Guidance Module (TGM), we train two versions of our *SimUESR* (with and without TGM) on USR-248 dataset. In addition, we present visualization results in Fig. 8 to provide a qualitative comparison of the enhanced images produced by each version of the network, where (a) presents the ground truth, (b) is the enhancement results without TGM, (c) is the result of full model, (d) and (e) denote the error map of (b) and (c), respectively. The error map, which is computed through the divergence between the output and GT, *i.e.,* $\|\hat{Y} - Y\|_1$, further supports the notion that the full model produces images that are more consistent with ground truth. We can see that TGM plays a critical role in regulating contrast and color deviation, and that images enhanced using the full model achieve greater similarity to the ground-truth in both low and high frequency regions. Furthermore, Table 7 demonstrates that TGM improves the performance of our network, as evidenced by an increase of 2.97 dB in PSNR and 0.016 in SSIM.

(2) Effectiveness of Prior Feature Extraction Block

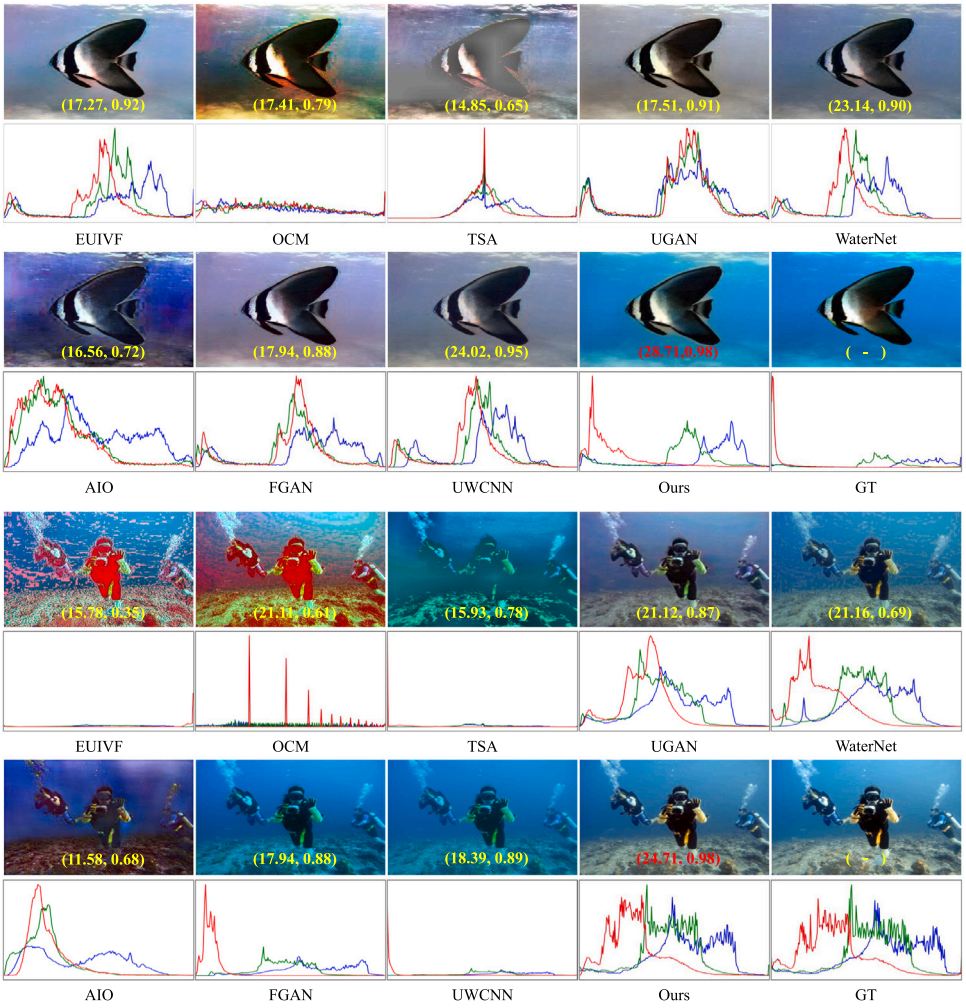


Fig. 6. Visual results of state-of-the-art methods and our *SimUESR* on the *EUVP* dataset (*Above*) and *UIEB* dataset (*Below*). The reference-aware assessments (i.e., PSNR and SSIM) are listed at the bottom. The pixel intensity distribution reflects the superiority of the proposed method in a statistical sense.

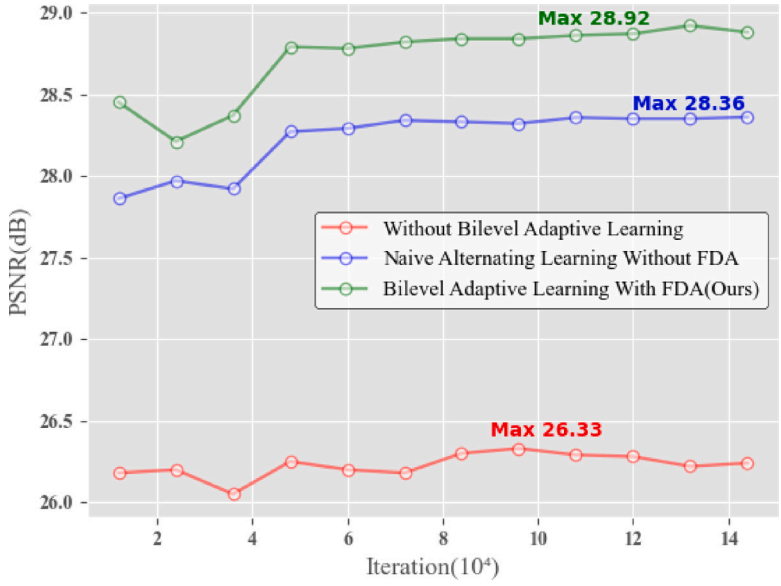


Fig. 7. Performance score accompanying the training process under three learning strategy. Our bilevel adaptive learning strategy is quick and stable to reach peak performance, otherwise the training oscillations converge slowly.

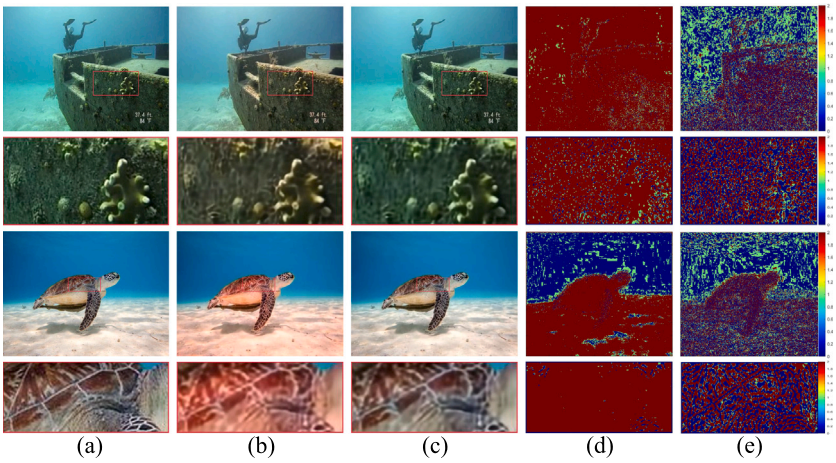


Fig. 8. Ablation study of TGM. (a) GT. (b) Result of w/o TGM. (c) Our final result. (d) Error map of w/o TGM. (e) Error map of final result. The error map is computed through the divergence between the output and GT.

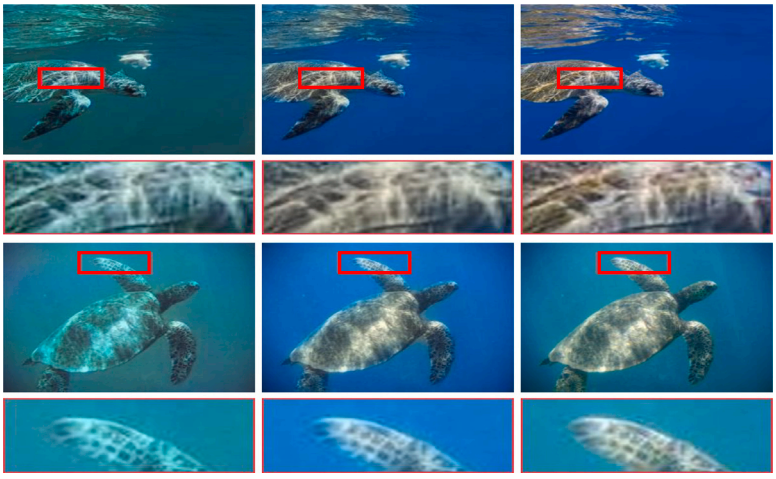


Fig. 9. Limitations. In extreme color distortion, the proposed method reduces local color richness and leads to a bias towards blue.

Table 7
Ablation study for modules on *USR-248* dataset.

Setting	TGM	PFEB	Groups	Training(h)↓	PSNR↑	SSIM↑
S_1	✔	✔	2	6.2	28.81	0.842
S_2	✔	✔	6	15.8	28.80	0.843
S_3	✔	✔	8	19.5	28.70	0.841
S_4	✘	✔	4	7.3	25.95	0.828
S_5	✔	✘	4	7.5	28.83	0.844
Ours	✔	✔	4	8.3	**28.92**	**0.844**

To more explicitly demonstrate the effect of Prior Feature Extraction Block (PFEB), we conduct training on a modified version of the *SimUESR* model which excluded the PFEB component, and instead utilize a basic convolutional layer to embed the transmission map into TGM. Table 7 demonstrates the noteworthy impact of the conditional prior feature extraction block on quantitation, as evidenced by the substantial improvement in both PSNR and SSIM. These results suggest that the extracted priors contain more comprehensive and detailed information.

(3) Effectiveness of different number of TGM and MDRM

We conduct a series of comparative experiments involving various numbers of module groups (*i.e.,* 2, 4, 6, 8) to assess their performance. Our experimental results suggest that our choice is superior to the alternatives. It is worth noting that a trade-off exists between accuracy and speed when selecting different group numbers. As indicated in Table 7, models with a larger number of groups require more computing resources but may not necessarily result in improved performance, while too few modules can have a negative impact on performance despite being more computationally efficient.

4.6. Limitation

Despite its robustness in restoring underwater images for joint tasks, the *SimUESR* has several limitations. Firstly, as illustrated in Fig. 9, when the underwater images are in severe color distortion, the greenish color cast restored by *SimUESR* will cause a bias to blue. Secondly, while the proposed method exhibits exceptional performance in enhancing underwater images, it may sometimes result in a reduction of the local color richness of the generated images. Finally, Our method does not score well for enhancement task on SSIM in some datasets, such as UIEB.

5. Conclusion and futurework

This paper dedicates a novel enhancer for jointly solving the challenging underwater image enhancement and super-resolution tasks within a singular model, dubbed *SimUESR*, Specifically, a transmission guidance modulation module is developed to investigate a prior

information-oriented feature representation. Thus, the underwater scene prior is progressively embedded into a multi-scale residual feature stream-based enhancer to recover high-frequency details and improve feature-level contrast. In addition, a novel bilevel adaptive learning strategy is introduced, based on bilevel optimization reformulation, which utilizes an early-stopping policy and finite difference approximation to learn the loss function coefficient instead of the traditional empirical loss-dependent selection. The effectiveness and influence of key components are thoroughly analyzed and measured in the ablation study. In future work, we aim to address the limitations of the proposed method and enhance its robustness in extremely harsh environments.

CRediT authorship contribution statement

Sihan Xie: Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **Peiming Li:** Investigation, Writing – original draft, Software. **Jiaxin Gao:** Supervision, Visualization, Writing – original draft. **Ziyu Yue:** Software. **Xin Fan:** Project administration, Supervision. **Risheng Liu:** Funding acquisition, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work is partially supported by the National Natural Science Foundation of China (Nos. U22B2052, 62027826), and the LiaoNing Revitalization Talents Program (No. 2022RG04).

References

[1] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, R. Timofte, Swinir: Image restoration using swin transformer, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 1833–1844.

[2] C. Ma, J. Zhang, J. Zhou, J. Lu, Learning series-parallel lookup tables for efficient image super-resolution, in: European Conference on Computer Vision, Springer, 2022, pp. 305–321.

[3] J. McMahon, E. Plaku, Autonomous data collection with timed communication constraints for unmanned underwater vehicles, IEEE Robot. Autom. Lett. 6 (2) (2021) 1832–1839, http://dx.doi.org/10.1109/LRA.2021.3060709.

[4] B. Kim, S.-C. Yu, Imaging sonar based real-time underwater object detection utilizing AdaBoost method, in: 2017 IEEE Underwater Technology, UT, IEEE, 2017, pp. 1–5.

[5] Z. Wang, L. Shen, Z. Wang, Y. Lin, Y. Jin, Generation-based joint luminance-chrominance learning for underwater image quality assessment, IEEE Trans. Circuits Syst. Video Technol. (2022).

[6] R. Liu, Z. Jiang, S. Yang, X. Fan, Twin adversarial contrastive learning for underwater image enhancement and beyond, IEEE Trans. Image Process. 31 (2022) 4922–4936.

[7] Z. Jiang, Z. Li, S. Yang, X. Fan, R. Liu, Target oriented perceptual adversarial fusion network for underwater image enhancement, IEEE Trans. Circuits Syst. Video Technol. 32 (10) (2022) 6584–6598.

[8] Z. Zhang, Z. Jiang, J. Liu, X. Fan, R. Liu, Waterflow: Heuristic normalizing flow for underwater image enhancement and beyond, in: Proceedings of the 31st ACM International Conference on Multimedia, 2023, pp. 7314–7323.

[9] J. Zhou, Q. Liu, Q. Jiang, W. Ren, K.-M. Lam, W. Zhang, Underwater camera: Improving visual perception via adaptive dark pixel prior and color correction, Int. J. Comput. Vis. (2023) 1–19.

[10] R. Liu, X. Fan, M. Zhu, M. Hou, Z. Luo, Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light, IEEE Trans. Circuits Syst. Video Technol. 30 (12) (2020) 4861–4875, http://dx.doi.org/10.1109/TCSVT.2019.2963772.

[11] Y. Kim, J.-S. Choi, M. Kim, A real-time convolutional neural network for super-resolution on FPGA with applications to 4K UHD 60 fps video services, IEEE Trans. Circuits Syst. Video Technol. 29 (8) (2019) 2521–2534, http://dx.doi.org/10.1109/TCSVT.2018.2864321.

[12] J. Lei, Z. Zhang, X. Fan, B. Yang, X. Li, Y. Chen, Q. Huang, Deep stereoscopic image super-resolution via interaction module, IEEE Trans. Circuits Syst. Video Technol. 31 (8) (2021) 3051–3061, http://dx.doi.org/10.1109/TCSVT.2020.3037068.

[13] Z.-S. Liu, W.-C. Siu, Y.-L. Chan, Photo-realistic image super-resolution via variational autoencoders, IEEE Trans. Circuits Syst. Video Technol. 31 (4) (2021) 1351–1365, http://dx.doi.org/10.1109/TCSVT.2020.3003832.

[14] J. Zhou, J. Sun, C. Li, Q. Jiang, M. Zhou, K.-M. Lam, W. Zhang, X. Fu, HCLR-Net: Hybrid contrastive learning regularization with locally randomized perturbation for underwater image enhancement, Int. J. Comput. Vis. (2024) 1–25, http://dx.doi.org/10.1007/s11263-024-01987-y.

[15] P. Drews, E. Nascimento, F. Moraes, S. Botelho, M. Campos, Transmission estimation in underwater single images, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2013, pp. 825–830.

[16] Y.-T. Peng, P.C. Cosman, Underwater image restoration based on image blurriness and light absorption, IEEE Trans. Image Process. 26 (4) (2017) 1579–1594.

[17] H. Zhao, X. Kong, J. He, Y. Qiao, C. Dong, Efficient image super-resolution using pixel attention, in: European Conference on Computer Vision, Springer, 2020, pp. 56–72.

[18] M.J. Islam, P. Luo, J. Sattar, Simultaneous enhancement and super-resolution of underwater imagery for improved visual perception, 2020, arXiv preprint arXiv:2002.01155.

[19] C. Li, S. Anwar, J. Hou, R. Cong, C. Guo, W. Ren, Underwater image enhancement via medium transmission-guided multi-color space embedding, IEEE Trans. Image Process. 30 (2021) 4985–5000.

[20] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, 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, 2017, pp. 4681–4690.

[21] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, C. Change Loy, Esrgan: Enhanced super-resolution generative adversarial networks, in: Proceedings of the European Conference on Computer Vision (ECCV) Workshops, 2018.

[22] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin transformer: Hierarchical vision transformer using shifted windows, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 10012–10022.

[23] M.J. Islam, S.S. Enan, P. Luo, J. Sattar, Underwater image super-resolution using deep residual multipliers, in: 2020 IEEE International Conference on Robotics and Automation, ICRA, IEEE, 2020, pp. 900–906.

[24] P.L. Drews, E.R. Nascimento, S.S. Botelho, M.F.M. Campos, Underwater depth estimation and image restoration based on single images, IEEE Comput. Graph. Appl. 36 (2) (2016) 24–35.

[25] C.-Y. Li, J.-C. Guo, R.-M. Cong, Y.-W. Pang, B. Wang, Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior, IEEE Trans. Image Process. 25 (12) (2016) 5664–5677.

[26] C. Ancuti, C.O. Ancuti, T. Haber, P. Bekaert, Enhancing underwater images and videos by fusion, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2012, pp. 81–88.

[27] X. Fu, Z. Fan, M. Ling, Y. Huang, X. Ding, Two-step approach for single underwater image enhancement, in: 2017 International Symposium on Intelligent Signal Processing and Communication Systems, ISPACS, IEEE, 2017, pp. 789–794.

[28] J. Perez, A.C. Attanasio, N. Nechyporenko, P.J. Sanz, A deep learning approach for underwater image enhancement, in: International Work-Conference on the Interplay Between Natural and Artificial Computation, Springer, 2017, pp. 183–192.

[29] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, D. Tao, An underwater image enhancement benchmark dataset and beyond, IEEE Trans. Image Process. 29 (2020) 4376–4389, http://dx.doi.org/10.1109/TIP.2019.2955241.

[30] C. Li, S. Anwar, F. Porikli, Underwater scene prior inspired deep underwater image and video enhancement, Pattern Recognit. 98 (2020) 107038.

[31] J. Zhou, B. Li, D. Zhang, J. Yuan, W. Zhang, Z. Cai, J. Shi, UGIF-Net: An efficient fully guided information flow network for underwater image enhancement, IEEE Trans. Geosci. Remote Sens. 61 (2023) 1–17, http://dx.doi.org/10.1109/TGRS.2023.3293912.

[32] J. Zhou, Q. Gai, D. Zhang, K.-M. Lam, W. Zhang, X. Fu, IACC: Cross-illumination awareness and color correction for underwater images under mixed natural and artificial lighting, IEEE Trans. Geosci. Remote Sens. 62 (2024) 1–15, http://dx.doi.org/10.1109/TGRS.2023.3346384.

[33] J. Li, K.A. Skinner, R.M. Eustice, M. Johnson-Roberson, WaterGAN: Unsupervised generative network to enable real-time color correction of monocular underwater images, IEEE Robot. Autom. Lett. 3 (1) (2017) 387–394.

[34] R. Liu, J. Gao, J. Zhang, D. Meng, Z. Lin, Investigating bi-level optimization for learning and vision from a unified perspective: A survey and beyond, IEEE Trans. Pattern Anal. Mach. Intell. (2021).

[35] R. Liu, J. Gao, X. Liu, X. Fan, Learning with constraint learning: New perspective, solution strategy and various applications, IEEE Trans. Pattern Anal. Mach. Intell. (2024) 1–18.

[36] K. Sabarika, S. Selvan, Image denosing and deblurring using framelet decomposition, in: 2017 Third International Conference on Sensing, Signal Processing and Security, ICSSS, 2017, pp. 469–473, http://dx.doi.org/10.1109/SSPS.2017.8071642.
[37] M. Nachaoui, L. Afraites, A. Laghrib, A regularization by denoising super-resolution method based on genetic algorithms, Signal Process., Image Commun. 99 (2021) 116505, http://dx.doi.org/10.1016/j.image.2021.116505, URL https://www.sciencedirect.com/science/article/pii/S0923596521002460.
[38] J. Xie, R.S. Feris, S.-S. Yu, M.-T. Sun, Joint super resolution and denoising from a single depth image, IEEE Trans. Multimed. 17 (9) (2015) 1525–1537, http://dx.doi.org/10.1109/TMM.2015.2457678.
[39] M.T. Rasheed, D. Shi, LSR: Lightening super-resolution deep network for low-light image enhancement, Neurocomputing 505 (2022) 263–275, http://dx.doi.org/10.1016/j.neucom.2022.07.058, URL https://www.sciencedirect.com/science/article/pii/S092523122200916X.
[40] A. Aakerberg, K. Nasrollahi, T.B. Moeslund, RELIEF: Joint low-light image enhancement and super-resolution with transformers, in: Scandinavian Conference on Image Analysis, Springer, 2023, pp. 157–173.
[41] J.Y. Chiang, Y.-C. Chen, Underwater image enhancement by wavelength compensation and dehazing, IEEE Trans. Image Process. 21 (4) (2011) 1756–1769.
[42] Y.-T. Peng, K. Cao, P.C. Cosman, Generalization of the dark channel prior for single image restoration, IEEE Trans. Image Process. 27 (6) (2018) 2856–2868.
[43] J. Zhou, Y. Wang, C. Li, W. Zhang, Multicolor light attenuation modeling for underwater image restoration, IEEE J. Ocean. Eng. 48 (4) (2023) 1322–1337, http://dx.doi.org/10.1109/JOE.2023.3275615.
[44] D. Zhang, J. Zhou, W. Zhang, Z. Lin, J. Yao, K. Polat, F. Alenezi, A. Alhudhaif, ReX-Net: A reflectance-guided underwater image enhancement network for extreme scenarios, Expert Syst. Appl. (2023) 120842.
[45] A. Ignatov, N. Kobyshev, R. Timofte, K. Vanhoey, L. Van Gool, Dslr-quality photos on mobile devices with deep convolutional networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 3277–3285.
[46] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[47] C. Guo, C. Li, J. Guo, C.C. Loy, J. Hou, S. Kwong, R. Cong, Zero-reference deep curve estimation for low-light image enhancement, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2020, pp. 1777–1786, http://dx.doi.org/10.1109/CVPR42600.2020.00185.
[48] M. Feurer, A. Klein, K. Eggensperger, J. Springenberg, M. Blum, F. Hutter, Efficient and robust automated machine learning, Adv. Neural Inf. Process. Syst. 28 (2015).
[49] R. Elshawi, M. Maher, S. Sakr, Automated machine learning: State-of-the-art and open challenges, 2019, arXiv preprint arXiv:1906.02287.
[50] S. Falkner, A. Klein, F. Hutter, BOHB: Robust and efficient hyperparameter optimization at scale, in: International Conference on Machine Learning, PMLR, 2018, pp. 1437–1446.
[51] R. Liu, J. Gao, X. Liu, X. Fan, Revisiting gans by best-response constraint: Perspective, methodology, and application, 2022, arXiv preprint arXiv:2205.10146.
[52] J. Gao, X. Liu, R. Liu, X. Fan, Learning adaptive hyper-guidance via proxy-based bilevel optimization for image enhancement, Vis. Comput. 39 (4) (2023) 1471–1484.
[53] H. Liu, K. Simonyan, Y. Yang, Darts: Differentiable architecture search, 2018, arXiv preprint arXiv:1806.09055.
[54] L. Franceschi, M. Donini, P. Frasconi, M. Pontil, Forward and reverse gradient-based hyperparameter optimization, 2017,
[55] R. Liu, P. Mu, X. Yuan, S. Zeng, J. Zhang, A general descent aggregation framework for gradient-based bi-level optimization, 2021, arXiv:2102.07976.
[56] H. Liang, S. Zhang, J. Sun, X. He, W. Huang, K. Zhuang, Z. Li, Darts+: Improved differentiable architecture search with early stopping, 2019, arXiv preprint arXiv:1909.06035.
[57] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, D. Tao, An underwater image enhancement benchmark dataset and beyond, IEEE Trans. Image Process. 29 (2019) 4376–4389.
[58] M.J. Islam, Y. Xia, J. Sattar, Fast underwater image enhancement for improved visual perception, IEEE Robot. Autom. Lett. 5 (2) (2020) 3227–3234.
[59] D. Kingma, J. Ba, Adam: A method for stochastic optimization, Comput. Sci. (2014).
[60] A. Horé, D. Ziou, Image quality metrics: PSNR vs. SSIM, in: 2010 20th International Conference on Pattern Recognition, 2010, pp. 2366–2369, http://dx.doi.org/10.1109/ICPR.2010.579.
[61] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, Image quality assessment: From error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612, http://dx.doi.org/10.1109/TIP.2003.819861.
[62] K. Panetta, C. Gao, S. Agaian, Human-visual-system-inspired underwater image quality measures, IEEE J. Ocean. Eng. 41 (3) (2016) 541–551, http://dx.doi.org/10.1109/JOE.2015.2469915.
[63] P.M. Uplavikar, Z. Wu, Z. Wang, All-in-one underwater image enhancement using domain-adversarial learning, in: CVPR Workshops, 2019, pp. 1–8.
[64] M.J. Islam, Y. Xia, J. Sattar, Fast underwater image enhancement for improved visual perception, IEEE Robot. Autom. Lett. 5 (2) (2020) 3227–3234, http://dx.doi.org/10.1109/LRA.2020.2974710.
[65] C. Fabbri, M.J. Islam, J. Sattar, Enhancing underwater imagery using generative adversarial networks, in: 2018 IEEE International Conference on Robotics and Automation, ICRA, IEEE, 2018, pp. 7159–7165.

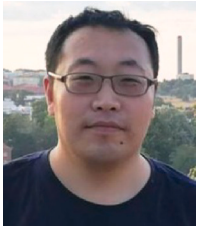
Sihan Xie received the B.E. degree in software engineering from Dalian University of Technology, China, in 2022, where she is currently pursuing the master's degree. Her research interests include super-resolution, underwater enhancement, and action recognition.

Peiming Li is an undergraduate majoring in digital media technology at the International School of Information Science&Engineering, Dalian University of Technology. His research interests include super-resolution, underwater enhancement, and human pose generation.

Jiaxin Gao received the B.S. degree in Applied Mathematics from Dalian University of Technology, China, in 2018. She is currently pursuing the Ph.D. degree in software engineering at Dalian University of Technology, Dalian, China. She is with the Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, Dalian University of Technology, Dalian, China. Her research interests include computer vision, machine learning and optimization.

Ziyu Yue graduated from Dalian University of Technology in 2017 with a B.S. degree in Information and Computing Science. He is currently pursuing his Ph.D. degree in Computer and Graphic Imaging at Dalian University of Technology. He works in the Liaoning Key Laboratory of Computational Mathematics and Data Intelligence at Dalian University of Technology. His research interests include super-resolution, low-light enhancement, and nerf.

Xin Fan (Senior Member, IEEE) was born in 1977. He received the B.E. and Ph.D. degrees in information and communication engineering from Xi'an Jiaotong University, Xi'an, China, in 1998 and 2004, respectively. He was with Oklahoma State University, Stillwater, OK, USA, from 2006 to 2007, as a Post-Doctoral Research Fellow. He joined the School of Software, Dalian University of Technology, Dalian, China, in 2009. He is also the Duty Dean of the DUTRU International School of Information and Science Technology. His current research interests include computational geometry and machine learning, and their applications to low-level image processing and DTI-MR image analysis.



Risheng Liu (Member, IEEE) was born in 1984. He received the B.E. degree in mathematics and the Ph.D. degree in computational mathematics from the Dalian University of Technology, Dalian, China, in 2007 and 2012, respectively. He was involved in joint training research at the Robotics Institute, Carnegie Mellon University, from 2010 to 2012. He has been a Lecturer and also an Associate Professor at the School of Software Technology, Dalian University of Technology, since 2012. His current research interests include machine learning, deep learning, computer vision, multimedia, and optimization.