

SimpleEnhanceNet: A Deep Learning-Based Underwater Image Enhancement Method

Chenyu Yu

*Department of Automation, Hohai University, Changzhou, China
yaircy@163.com*

Abstract. Underwater images typically suffer from low contrast, color distortion, and blurred details caused by light absorption and scattering, which severely limit the performance of visual perception tasks such as marine ecosystem monitoring, shipwreck inspection, and autonomous underwater vehicle (AUV) navigation. Conventional physics-based restoration methods are highly sensitive to water types and illumination conditions, thus lacking robustness in practical scenarios. To overcome these limitations, we propose a deep learning-based underwater image enhancement framework termed SimpleEnhanceNet, which adopts an encoder–decoder convolutional neural network (CNN) with skip connections. The network is trained on the public UIEB and EUVP datasets in a supervised manner, where the optimization objective combines pixel-wise mean squared error with perceptual loss to jointly preserve structural fidelity and perceptual quality. Extensive experiments demonstrate that our method achieves superior performance over traditional approaches on the UIEB benchmark, yielding improvements of 5.82 dB in peak signal-to-noise ratio (PSNR) and 0.21 in structural similarity index (SSIM). Moreover, the no-reference metrics UIQM and UCIQE also exhibit substantial gains. Qualitative comparisons confirm that SimpleEnhanceNet effectively restores natural colors and enhances scene clarity across diverse underwater conditions. These results highlight its potential for real-time deployment in AUV navigation, environmental monitoring, and marine exploration.

Keywords: Underwater Image Enhancement, Lightweight Convolutional Network, Skip Connections, Multi-scale Feature Fusion, Perceptual Loss

1. Introduction

Underwater images serve as a crucial source of information for marine science and engineering applications, playing an essential role in tasks such as marine ecosystem monitoring, shipwreck and coral reef inspection, environmental protection, and autonomous underwater vehicle (AUV) navigation. However, due to the absorption and scattering of light in water, underwater images often suffer from degradations such as low contrast, color distortion, and blurred details. Such quality degradation not only hinders the visual perception and interpretation for human observers but also limits the reliability and robustness of computer vision algorithms in underwater scenarios. Therefore, enhancing the quality of underwater images remains a fundamental and challenging problem in underwater visual perception.

Traditional underwater image enhancement approaches are mostly based on physical imaging models or color correction priors, such as histogram equalization, white balance adjustment, dark channel prior, and optical propagation-based restoration methods [1]. However, these approaches are highly sensitive to water conditions and illumination environments, often leading to over-enhancement or under-enhancement in complex underwater scenarios, thereby lacking robustness. With the rapid development of deep learning, convolutional neural networks (CNNs) have been widely applied to underwater image enhancement tasks [2]. Among them, encoder–decoder architectures represented by U-Net can effectively integrate local and global features. Nevertheless, existing U-Net-based methods still suffer from several limitations when applied to diverse underwater environments: on the one hand, their feature modeling capacity is insufficient to adapt to degradations caused by varying water depths and turbidity; on the other hand, most methods rely solely on pixel-level loss functions, making the enhanced results prone to unnatural color reproduction or missing details.

To overcome the aforementioned limitations, several improved approaches have been proposed. Chen et al. enhanced the Water-Net model by refining the enhancement units and applying Softmax to calibrate the confidence map output, thereby improving the model’s robustness [3]. Yang et al. introduced the F-GAN model based on generative adversarial networks for underwater image color correction, employing a multi-objective function to supervise the training process and evaluate image quality [4]. Jamieson et al. proposed the DeepSeeColor model, which combines depth estimation with physical modeling to correct degradations under varying water depth conditions [5]. Peng et al. incorporated the Transformer model into underwater image enhancement, leveraging multi-scale feature fusion and global feature modeling to improve the network’s attention to severely attenuated color channels [6].

Although existing methods have achieved progress in various aspects, several challenges remain. First, cross-dataset generalization is insufficient, and the model’s performance varies significantly under different water quality conditions. Second, it is difficult to balance structural fidelity with perceptual visual quality; some methods produce natural colors but lack detail sharpness. Third, certain network architectures are complex and fail to meet real-time requirements, limiting their deployment on embedded platforms such as AUVs [7]. Therefore, designing lightweight, efficient, and generalizable deep models remains a critical research direction in the field of underwater image enhancement.

To address the aforementioned issues, this paper proposes a deep learning-based method for underwater image enhancement. The method employs an encoder–decoder convolutional network with skip connections and is trained in a supervised manner on the publicly available UIEB and EUVP datasets[8–9]. The loss function integrates pixel-wise mean squared error and perceptual loss to simultaneously improve structural fidelity and perceptual visual quality. Experimental results demonstrate that the proposed method significantly outperforms traditional approaches in both full-reference metrics such as PSNR and SSIM, and no-reference metrics including UIQM and UCIQE, achieving more natural color restoration and enhanced clarity across various complex underwater scenarios.

The main contributions of this work are summarized as follows: First, a lightweight encoder–decoder network is designed to address the diverse degradation characteristics of underwater images, employing skip connections to achieve multi-scale feature fusion. Second, an optimization objective combining pixel-wise loss and perceptual loss is constructed to enhance perceptual visual quality while maintaining structural consistency. Third, systematic experiments are conducted on the UIEB and EUVP datasets, showing that the proposed method outperforms existing traditional

approaches in both quantitative and qualitative evaluations, demonstrating strong generalization and application potential. The remainder of this paper is organized as follows: Section 2 reviews underwater image enhancement and related research progress; Section 3 presents the network architecture and training strategy of the proposed method; Section 4 details the experimental setup and results analysis; Section 5 concludes the paper and provides discussion.

2. Related work

2.1. CNN-based underwater image enhancement networks

With the application of deep learning in underwater image enhancement, convolutional neural networks (CNNs) based on encoder–decoder structures have become the mainstream paradigm. By progressively downsampling to extract semantic features and then upsampling to restore spatial resolution, these networks can simultaneously improve local details and global structures in an end-to-end training manner. Based on this approach, Li et al. proposed Water-Net, which generates multiple candidate enhanced images through preprocessing methods such as white balancing and histogram equalization, and learns corresponding confidence maps for weighted fusion, achieving robust color restoration on real underwater images [8]. The Ucolor network integrates multiple color-space encoders with a medium-transmission-based decoder, adaptively fusing different features via an attention mechanism, significantly improving enhancement for color casts and low-contrast images [10]. In addition, lightweight CNN designs have been proposed to reduce computational cost while maintaining enhancement quality, making them suitable for embedded platforms such as underwater robots. For example, Yang et al. proposed LU2Net, which employs axial depthwise convolutions and channel attention modules to significantly reduce computational demands and model parameters, thereby improving processing speed [11]. However, despite their advantages in local detail recovery and efficient inference, CNN-based methods still face challenges due to the limited receptive field of convolutional operators, making it difficult to model long-range color shifts and non-uniform degradations, and their generalization across varying water quality and depth conditions remains limited.

2.2. Skip connections and feature fusion

Feature fusion and skip connection mechanisms are core designs in underwater image enhancement networks. Regarding skip connections, the low-level features from the encoder are directly passed to the corresponding scales of the decoder, preventing information loss through layer-by-layer processing and enhancing the detail recovery of high-level semantic features during reconstruction. Gao et al. proposed a U-shaped network with skip connections to hierarchically capture multi-scale information for underwater polarimetric dehazing, reducing data pre-processing and floating-point computations, while improving the preservation of edge and texture details during reconstruction [12]. For multi-scale feature fusion, information from different feature levels is combined to balance local details and global color distribution. Xu et al. proposed a GAN-based underwater image enhancement network that leverages multi-scale feature fusion within the generator to integrate hierarchical representations from different layers. This design allows the network to simultaneously capture global color consistency and local texture details, thereby improving the overall perceptual quality of enhanced underwater images [13]. Additionally, attention mechanisms have been increasingly incorporated, with channel and spatial attention enhancing the network's focus on key regions, effectively restoring color and contrast in complex environments. For example, Peng et al.

proposed the U-shape Transformer, which incorporates channel attention modules at each layer of the encoder and decoder to assign weights according to the importance of different channels. Simultaneously, spatial attention modules generate saliency maps to highlight foreground regions and edge structures. During the multi-scale feature fusion process, these attention mechanisms guide the network to more accurately restore regions with severe color deviation while preserving image texture details, thereby significantly enhancing both color correction and visual clarity of the enhanced images [6]. Although these fusion strategies significantly improve enhancement performance, they also increase model complexity and training difficulty, posing challenges for resource-constrained embedded platforms and real-time applications.

2.3. Loss functions and training strategies

In underwater image enhancement, the design of network training objectives has a critical impact on enhancement performance. Traditional pixel-wise reconstruction losses ensure structural consistency but often result in overly smooth outputs with insufficient texture details. By introducing perceptual loss, adversarial loss, and multi-objective training strategies, both structural fidelity and perceptual visual quality can be improved. Perceptual loss is typically computed based on high-level feature differences extracted from pre-trained networks such as VGG [14], measuring the discrepancy between enhanced and reference image features to optimize the network's ability to restore textures and colors. Adversarial loss, leveraged within a generative adversarial network framework, constrains the generated images to resemble real distributions, further enhancing perceptual realism [9]. Multi-objective training strategies integrate pixel-wise reconstruction loss, perceptual loss, adversarial loss, and color consistency constraints, enabling the network to maintain structural fidelity and visual perception under various degradation conditions. During training, both full-reference and no-reference quality metrics are used for evaluation and hyperparameter tuning, quantifying color naturalness, sharpness, and contrast [15]. In summary, careful design of loss functions and training strategies is essential for improving perceptual quality and cross-domain generalization. The proposed method introduces an improved skip-fusion module and perceptually guided training objectives within a lightweight encoder-decoder network, effectively balancing structural fidelity and visual quality.

3. Method

3.1. Overall network architecture

The proposed underwater image enhancement network employs a lightweight encoder-decoder convolutional neural network backbone to meet real-time processing requirements on resource-constrained platforms such as AUVs and ROVs. The encoder progressively downsamples the input image to extract multi-level semantic representations, with shallow layers focusing on edges and textures, and deeper layers capturing global semantics and degradation patterns. The decoder progressively upsamples to restore spatial resolution, integrating key information from the encoder at each scale during reconstruction, ultimately producing the enhanced image.

To prevent detail loss caused by unidirectional upsampling, skip connections are established between aligned scales of the encoder and decoder. Traditional symmetric skip connections directly concatenate encoder and decoder features at the same scale, which can supplement high-frequency details but often introduces unstable color casts and noise into the decoder in underwater scenarios. To address this issue, the proposed model inserts a lightweight channel-spatial attention module

before each skip connection. Specifically, in the channel dimension, global average pooling is applied to aggregate channel statistics, followed by a 1×1 convolution and Sigmoid activation to generate channel weights. In the spatial dimension, a local convolution-based saliency map suppresses background and regions with strong scattering, highlighting structural edges and foreground areas. Features processed by this module are then fused with decoder features, enhancing detail consistency and reconstructability while reducing artifacts and oversaturation in severely degraded regions.

All convolutional layers in the network use 3×3 kernels with ReLU activation to ensure nonlinear representation capability. Upsampling is performed through a combination of bilinear interpolation and convolution to reduce artifacts. This lightweight design controls the number of parameters and computational complexity, ensuring network adaptability and robustness while facilitating deployment on embedded platforms.

3.2. Multi-scale feature fusion

For non-uniform illumination attenuation and channel-dependent color shifts in underwater images, single-scale features are insufficient to balance global consistency and local detail recovery. To address this, the proposed model introduces a weighted fusion strategy based on multi-level outputs from the encoder. Specifically, feature maps are extracted from shallow, intermediate, and deep layers, and a lightweight channel attention module is used to estimate response weights for each scale. The weighted multi-scale features are then concatenated along the channel dimension to form fused features, which are subsequently injected into the decoder branch to participate in progressive reconstruction.

This process implements a stepwise information flow of selection–aggregation–reconstruction: selection suppresses channels heavily affected by scattering noise; aggregation integrates global and local cues within a unified representation space; reconstruction explicitly guides the joint recovery of edges and colors during upsampling. Based on the multi-scale perception and multi-dimensional spatial fusion concept [13], the model further emphasizes lightweight implementation: the attention module consists of 1×1 convolutions and per-channel normalization without introducing global self-attention computation, thereby reducing inference cost while maintaining enhancement quality.

3.3. Loss function design

To balance structural fidelity and perceptual quality, this work adopts a dual-objective optimization strategy combining pixel reconstruction loss and perceptual loss. The overall objective can be expressed as:

$$L = \lambda_1 L_{\text{pixel}} + \lambda_2 L_{\text{perc}} \quad (1)$$

where λ_1 and λ_2 are the weights for pixel and perceptual losses, respectively.

The pixel reconstruction loss uses mean squared error (MSE) to constrain the difference between the enhanced image and the reference image at the pixel level:

$$L_{\text{pixel}} = \frac{1}{3HW} \|\hat{I} - I^*\|_2^2 \quad (2)$$

where $\hat{I}$ and I^* denote the enhanced and reference images, respectively, and H and W denote the height and width of the image, respectively.

To overcome over-smoothing and texture loss associated with pixel-level metrics, perceptual loss is introduced, based on high-level feature representations extracted from a pre-trained VGG network:

$$L_{\text{perc}} = \sum_{\ell} \frac{1}{C_{\ell}H_{\ell}W_{\ell}} \left\| \phi_{\ell}(\hat{\mathbf{I}}) - \phi_{\ell}(\mathbf{I}^*) \right\|_1 \quad (3)$$

where $\phi_{\ell}()$ denotes the feature map of the ℓ -th VGG layer, and $C_{\ell}, H_{\ell}, W_{\ell}$ are the corresponding channel, height, and width of the feature map.

The loss weights λ_1 and λ_2 are determined via grid search on the validation set to ensure that both full-reference and no-reference metrics benefit simultaneously.

4. Experiments

4.1. Datasets

To evaluate the performance of the proposed underwater image enhancement network in terms of structural fidelity, texture detail recovery, and color naturalness, the UIEB [8] and EUVP [9] datasets are employed for training and testing.

The UIEB dataset contains 890 real underwater images, of which 800 images are used for training and 90 images with expert-provided reference images are used for testing and evaluation. This dataset covers various water bodies and lighting conditions and is one of the most commonly used benchmarks for supervised learning scenarios. The EUVP dataset consists of over 20k real and synthetic underwater images, including paired low- and high-quality images, enabling evaluation of model generalization under diverse water conditions.

In the experiments, the UIEB dataset is used for training and quantitative evaluation, while the EUVP dataset serves as an additional cross-dataset test to validate the model's generalization capability.

4.2. Experimental setup

All training and testing experiments were conducted on a local workstation equipped with a 13th Gen Intel(R) Core(TM) i9-13900HX CPU and an NVIDIA GeForce RTX 4060 Laptop GPU, using the PyTorch deep learning framework. During training, the Adam optimizer was employed with an initial learning rate of 1×10^{-4} , decayed at fixed intervals during mid-training. The batch size was set to 4, and training proceeded for a total of 100 epochs. Input images were uniformly resized to 256×256 pixels and normalized to the $[0,1]$ range. To enhance network robustness, data augmentation strategies such as random horizontal flipping and random cropping were applied. The training loss function followed the joint objective of pixel reconstruction loss and perceptual loss described in Section 3.3.

4.3. Experimental results

Quantitative evaluation was conducted using a combination of full-reference and no-reference metrics.

Full-reference metrics include peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) [15]. PSNR measures the pixel-level closeness between the enhanced and reference images,

with higher values indicating lower distortion. SSIM assesses structural and textural similarity, ranging from 0 to 1, where higher values indicate better structural fidelity.

No-reference metrics include the Underwater Image Quality Measure (UIQM) and the Underwater Color Image Quality Evaluation (UCIQE). UIQM evaluates image quality comprehensively from color fidelity, contrast, and sharpness, with higher values indicating better quality. UCIQE, based on color deviation, saturation, and contrast, effectively reflects the perceptual visual quality of underwater images.

This combination of full-reference and no-reference metrics ensures that the evaluation captures both structural accuracy and perceptual improvement.

To validate the effectiveness of the proposed method, several traditional enhancement methods were selected for comparison. In the experiments, the groups are defined as follows: Raw (RAW) refers to the original unenhanced images for visual comparison; Histogram Equalization (HE) enhances contrast by stretching the grayscale histogram but may cause color distortion and over-enhancement in complex underwater environments; the proposed method (SimpleEnhanceNet) employs a lightweight encoder-decoder network with multi-scale feature fusion for image enhancement; Reference denotes high-quality images manually processed by experts, representing the theoretical upper bound.

The comparative results of different methods on the UIEB test set are presented in Table 1.

Table 1. Comparison of different methods

Method	PSNR↑	SSIM↑	UIQM↑	UCIQE↑
Raw	14.32	0.45	2.41	48.5
HE	15.21	0.47	2.53	50.2
SimpleEnhanceNet	20.14	0.68	3.02	56.7

The experimental results show that the proposed method achieves approximately 5 dB and 0.21 improvements over the traditional histogram equalization approach in PSNR and SSIM, respectively, and also demonstrates significant advantages in UIQM and UCIQE. This indicates that the proposed method outperforms conventional techniques in both structural fidelity and perceptual visual quality.

As shown in Table 2, further testing on the EUVP dataset shows that the method maintains strong enhancement performance in unsupervised scenarios, particularly for images with bluish-green color casts and low illumination, demonstrating good generalization. These results suggest that the proposed lightweight network architecture and loss design possess notable cross-domain adaptability.

Table 2. Cross-dataset generalization results on EUVP

Method	PSNR↑	SSIM↑	UIQM↑	UCIQE↑
Raw	13.85	0.42	2.36	47.9
HE	14.72	0.45	2.48	49.6
SimpleEnhanceNet	18.94	0.64	2.91	55.1
Reference	23.50	0.80	3.42	57.3

5. Conclusion and discussion

To address the common issues of illumination attenuation and color shift in underwater images, this paper proposes a lightweight underwater image enhancement network, aiming to achieve a balance between quality restoration and computational efficiency. The method is based on an encoder-decoder convolutional architecture, incorporates skip connections and lightweight channel attention modules, and employs joint optimization of pixel reconstruction loss and perceptual loss, effectively addressing the shortcomings of traditional methods in texture preservation and color restoration. Extensive experiments comparing the proposed method with other approaches show that it achieves excellent performance on the publicly available UIEB and EUVP underwater image enhancement benchmark datasets. The method not only outperforms traditional techniques such as histogram equalization in full-reference metrics (PSNR, SSIM) but also achieves significant improvements in no-reference metrics (UIQM, UCIQE), further validating its applicability and robustness in real-world scenarios. Despite these advances, several directions remain for further exploration. For example, the current loss functions mainly combine pixel reconstruction and perceptual constraints, and adversarial loss could be introduced to further enhance optimization of perceptual quality. In addition, the model's training relies on paired reference data, and its cross-domain generalization still requires further improvement. Future research will focus on the following three aspects: structural optimization by integrating multi-scale fusion and attention mechanisms to enhance the model's representation of complex degradation patterns; reducing reliance on reference data through unsupervised and self-supervised learning to improve adaptability under different water quality and lighting conditions; and multi-task and multi-modal extensions, combining underwater object detection, segmentation, and recognition tasks to explore cross-modal information fusion and enhance overall perceptual capability.

References

- [1] Wang Y, Song W, Fortino G, et al. An experimental-based review of image enhancement and image restoration methods for underwater imaging [J]. IEEE access, 2019, 7: 140233-140251.
- [2] Vijayalakshmi M, Sasithradevi A. A comprehensive review on deep learning architecture for pre-processing of underwater images [J]. SN Computer Science, 2024, 5(5): 472.
- [3] Chen Y, Li H, Yuan Q, et al. Underwater image enhancement based on improved water-net [C]//2022 IEEE International Conference on Cyborg and Bionic Systems (CBS). IEEE, 2023: 450-454.
- [4] Yang F H, Guo R R Z, Cheung R C C, et al. F-GAN: Real-Time Color Correction Model of Underwater Images [C]//TENCON 2022-2022 IEEE Region 10 Conference (TENCON). IEEE, 2022: 1-6.
- [5] Jamieson S, How J P, Girdhar Y. Deepseecolor: Realtime adaptive color correction for autonomous underwater vehicles via deep learning methods [J]. arXiv preprint arXiv: 2303.04025, 2023.
- [6] Peng L, Zhu C, Bian L. U-shape transformer for underwater image enhancement [J]. IEEE transactions on image processing, 2023, 32: 3066-3079.
- [7] Zhang S, Zhao S, An D, et al. LiteEnhanceNet: A lightweight network for real-time single underwater image enhancement [J]. Expert Systems with Applications, 2024, 240: 122546.
- [8] Li C, Guo C, Ren W, et al. An underwater image enhancement benchmark dataset and beyond [J]. IEEE transactions on image processing, 2019, 29: 4376-4389.
- [9] Islam M J, Xia Y, Sattar J. Fast underwater image enhancement for improved visual perception [J]. IEEE robotics and automation letters, 2020, 5(2): 3227-3234.
- [10] Li C, Anwar S, Hou J, et al. Underwater image enhancement via medium transmission-guided multi-color space embedding [J]. IEEE Transactions on Image Processing, 2021, 30: 4985-5000.
- [11] Zhang Z, Jiang Y, Jiang J, et al. Star: A structure-aware lightweight transformer for real-time image enhancement [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 4106-4115.
- [12] Gao J, Wang G, Chen Y, et al. Mueller transform matrix neural network for underwater polarimetric dehazing imaging [J]. Optics Express, 2023, 31(17): 27213-27222.

- [13] Xu B, Zhou D, Li W. Image enhancement algorithm based on GAN neural network [J]. IEEE Access, 2022, 10: 36766-36777.
- [14] Tammina S. Transfer learning using vgg-16 with deep convolutional neural network for classifying images [J]. International Journal of Scientific and Research Publications (IJSRP), 2019, 9(10): 143-150.
- [15] Hore A, Ziou D. Image quality metrics: PSNR vs. SSIM [C]//2010 20th international conference on pattern recognition. IEEE, 2010: 2366-2369.