Algerian Journal of Research and Technology
https://doi.org/10.58681/ajrt.23070205
A.J.R.T Volume 7 (N°2) (2023) 43 - 51

Deep Neural Network model for Underwater Image Enhancement

O.Benaida[*,a], A. Loukil[b], A. Ali Pacha[a]

[a]LACOSI (Laboratoire de Codage et de la Sécurité de l'Information)
[b]LARESI (Laboratoire de Recherche en Systèmes Intelligents)
Department of Electronics, Faculty of Electrical Engineering. University of Science and Technology of Oran, Algeria

***Corresponding author:** *ouafa.benaida@univ-usto.dz**

Received. June 03, 2023. Accepted. August 02, 2023. Published August 30, 2023

DOI: https://doi.org/10.58681/ajrt.23070205

Abstract. In recent years, there has been a growing interest in the field of underwater image enhancement, driven by its significance in underwater robotics and ocean engineering. Initially, research efforts focused on physics-based approaches, but with advancements in technology, the utilization of deep convolutional neural networks (CNNs) and generative adversarial networks (GANs) has become prevalent.

These state-of-the-art algorithms have shown impressive results; however, their computational complexity and memory requirements pose challenges to their practical implementation on portable devices used for underwater exploration tasks. Furthermore, these models are often trained on either synthetic or limited real-world datasets, limiting their applicability in real-world scenarios.

In this paper, we propose a novel deep neural network architecture that maintains high performance while reducing the number of parameters compared to existing state-of-the-art models. Our approach aims to address the computational and memory limitations associated with underwater image enhancement algorithms. By leveraging the strengths of our architecture, we demonstrate its generalization capability by evaluating its performance on a combination of synthetic and real-world datasets. This approach enhances the practicality and applicability of our model in real-world underwater scenarios.

The findings presented in this paper lay the foundation for further exploration and development in this field.

Keywords. Underwater image enhancement, Convolutional neural network CNNs, Generative adversarial network GANs, Deep learning.

INTRODUCTION

The field of underwater robotics is experiencing rapid growth and is being driven by extensive research efforts. One prominent development in this area is the advancement of autonomous underwater vehicles (AUVs), which are deployed to address various challenging engineering problems. These include tasks such as underwater surveillance, seabed mapping, underwater archaeological exploration, garbage collection, underwater rescue operations, and military applications. To effectively tackle these challenges, real-time interpretation of images and videos is crucial for AUVs to intelligently perceive the underwater environment and take appropriate actions.

However, underwater images inherently suffer from degradation caused by wavelength-dependent absorption and forward/backward scattering due to particles present in the water. This degradation leads to reduced visibility, decreased contrast, color deviations, and even the introduction of color casts. These limitations significantly hinder the usability of underwater images in downstream tasks that rely on underwater vision, such as tracking, classification, and detection. Moreover, underwater images tend to have a dominant green or blue hue, as the red wavelengths are absorbed more profoundly in deep water.

These inherent challenges emphasize the need for effective techniques to enhance and improve the quality of underwater images. By addressing issues such as visibility, contrast, color accuracy, and color casts, it becomes possible to enhance the applicability of underwater vision for various tasks performed by AUVs. Consequently, developing robust algorithms and methodologies for underwater image enhancement is crucial to overcome the limitations imposed by underwater conditions and optimizing the performance of AUVs in real-world scenarios.

Overall, the advancements in underwater robotics and the challenges associated with underwater image quality pave the way for innovative research and technological breakthroughs. By overcoming the limitations imposed by underwater environments, we can unlock the full potential of AUVs and enable them to operate effectively in diverse underwater applications.

To address the aforementioned challenges, the initial step before undertaking any downstream tasks involving underwater image interpretation is image improvement. This process encompasses tasks such as image enhancement, image restoration, and specific methods tailored to particular tasks. In our study, our primary focus is on image enhancement, as it plays a vital role in the real-time alleviation of the problems associated with underwater images by restoring their perceptual and statistical qualities.

Underwater image enhancement methods have the advantage of extracting image information without relying on prior knowledge about the specific underwater environment. This characteristic makes them more generalized compared to image restoration methods. The existing literature in this field mainly revolves around very deep Convolutional Neural Networks (CNNs) and Generative Adversarial Network (GAN)-based models (Anwar and Li, 2020). These approaches address various aspects such as noise removal, contrast stretch, combined improvement using multiple information sources, and deep learning techniques for image dehazing.

However, the major drawback of these advanced models lies in their high computational and memory requirements, rendering them unsuitable for real-time underwater image enhancement tasks. Consequently, to enhance the deployability of machine learning models and reduce their compute and memory demands while maintaining comparable performance to state-of-the-art models, we propose our novel model.

By developing our model, we aim to strike a balance between computational efficiency and performance. Our model addresses the limitations posed by heavy computational and memory requirements, making it more suitable for real-time underwater image enhancement applications. Through our research, we strive to improve the practicality and usability of machine learning models in this domain, paving the way for more efficient and deployable solutions in underwater image enhancement.

In summary, our proposed model aims to optimize the trade-off between computational demands and performance, enabling real-time underwater image enhancement with comparable results to existing state-of-the-art models. By reducing the computational burden while maintaining effectiveness, we seek to contribute to the development of more practical and deployable solutions in this field.

BACKGROUND

In recent years, numerous methods have been proposed to tackle the task of image enhancement, which can be broadly categorized into three groups: non-physical models, physical model-based methods, and deep learning methods.

Non-physical models (Yang et al., 2022) primarily improve image quality by adjusting pixel values without relying on a specific mathematical equation. On the other hand, physical model-based (Wang et al., 2021) methods aim to formulate the degradation process of the image by estimating the parameters of a model. However, both of these approaches alone are not sufficient for effective underwater image enhancement, as they often overlook the specific properties associated with underwater conditions.

Deep learning methods (Almutiry et al., 202) have shown promising results in image enhancement, particularly in addressing color correction challenges. These methods primarily utilize generative adversarial networks (GANs) and convolutional neural networks (CNNs) to achieve notable improvements. In our study, we have conducted a detailed review of the current state-of-the-art GAN-based models and CNN-based models, which have demonstrated impressive performance in underwater image enhancement.

By focusing on color correction, deep learning methods tend to outperform non-physical and physical model-based methods in the context of underwater image enhancement. The utilization of GANs and CNNs enables these models to learn complex representations and capture specific underwater image characteristics more effectively.

Through our comprehensive review, we aim to provide a thorough understanding of the advancements in GAN-based and CNN-based models for underwater image enhancement. By evaluating and analyzing the state-of-the-art approaches, we can gain insights into their strengths, limitations, and potential for further improvement.

In summary, deep learning methods, particularly GANs and CNNs have emerged as powerful tools for addressing color correction challenges in underwater image enhancement. By examining and scrutinizing the current state-of-the-art models, we can gain valuable knowledge and contribute to the ongoing progress in this field.

Funie-gan

The proposed approach (Islam et al., 2020) addresses image blurring by formulating it as an image-to-image translation problem, assuming the presence of a non-linear mapping between distorted and enhanced images. Through adversarial training on a large-scale dataset, a conditional GAN-based model learns this mapping. However, an issue arises as the model incorrectly models sunlight and amplifies background noise, resulting in over-saturated or under-saturated images.

Water-net

The architecture under consideration (Li et al., 2019) is a gated fusion CNN, trained on the UIEB dataset (Li et al., 2017), specifically designed for underwater image enhancement. In order to align with the unique characteristics of degraded underwater images, the Water-Net model incorporates three enhanced inputs: White Balance (WB), Gamma Correction (GC), and Histogram Equalization (HE). However, the complex nature of this CNN architecture presents a challenge in dealing with the adverse effect of backscatter.

UNDERWATER DATASET

In our study, we extensively examined a diverse range of underwater image datasets, encompassing both synthetically generated and real-world underwater images. We specifically selected three datasets, which we will describe in detail below, to evaluate and benchmark the generalization capabilities of our model against state-of-the-art models trained on these datasets.

EUVP Dataset

The EUVP Dataset (Enhancement of Underwater Visual Perception) (Islam et al., 2020) is a comprehensive collection consisting of 10,000 paired images and 25,000 unpaired images. These images were captured by the authors Islam et al (2020) during oceanic explorations under various visibility conditions. The paired images were created by applying a distortion model based on Cycle GAN to real-world images, thus simulating underwater distortions. For the training-validation phase of our model, we utilize a subset of the Image Net paired images from the EUVP dataset. Additionally, we employ another subset called EUVP-Dark, which consists of a pair of highly degraded images, for testing our model's performance.

UIEB Dataset

The UIEB dataset (Li et al., 2019) (Underwater Image Enhancement Benchmark) comprises a collection of 890 real underwater images captured under varying lighting conditions. These images exhibit diverse color ranges and degrees of contrast. The authors of the dataset have provided corresponding reference images that are free from color casts and demonstrate improved visibility and brightness compared to the original source images. We evaluate the performance of our model on this dataset as it represents a real-world underwater dataset with reference images obtained without the use of synthetic techniques.

EVALUATION METRICS

In our study, we performed a quantitative evaluation of the output images generated by our model using standard metrics proposed by Yang et al. (2022). Including Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM). These metrics provide an assessment of the structural similarity and reconstruction quality of the enhanced output image compared to the corresponding reference image.

Furthermore, we analyzed the quality of the generated output images using a non-reference underwater image quality measure (UIQM). The UIQM incorporates three underwater image attribute measures: image colorfulness (UICM), sharpness (UISM), and contrast (UIConM). Each attribute evaluates a specific aspect of underwater image degradation. The UIQM is computed using the formula:

$$\mathrm{UIQM} = c_1 \times \mathrm{UICM} \;\; c_2 \times \mathrm{UISM} \quad c_3 \times\times \mathrm{UIConM} \quad (1)$$

The values of the parameters c1, c2, and c3 are set to 0.0282, 0.2953, and 3.5753, respectively, as specified in the paper by Panetta et al. (2015).

To assess the quality of model compression and acceleration, we calculated compression rates and speed-up rates, as proposed by Cheng et al. (2017). The compression rate (M, M_*) is determined by the ratio of the number of parameters of the original model M (α(M)) to the number of parameters of the compressed model M_* (α(M_*)). Similarly, the speed-up rate (M,

M_*) is calculated by dividing the testing time for one image for the original model M ($\beta(M)$) by the testing time for one image for the compressed model M_* ($\beta(M_*)$).
In summary:

$$\text{Compression rate } (M, M*) = \frac{\alpha(M)}{\alpha(M*)}$$

$$\text{Speed-up rate } (M, M*) = \frac{\beta(M)}{\beta(M*)}$$

M represents the original model
M_* represents the compressed model
These metrics provide valuable insights into the performance and efficiency of our model, both in terms of image quality and computational resources.

OUR PROPOSED APPROACH

In this section, we will begin by providing a detailed explanation of the proposed architecture, including its components and the calculations performed within the network. Following that, we will discuss the evaluation metrics used to assess the performance of the model.

Network ARCHITECTURE

Figure 1 illustrates the architecture diagram of our proposed model. The model consists of a fully connected convolution network that is connected to three densely connected convolutional blocks in a sequential manner. A skip connection is employed, where the input image is concatenated with the output of each block. The input to our model is an RGB underwater image with dimensions of 256x256 pixels.
The raw input image is initially passed through the first convolutional layer with a kernel size of 3x3, resulting in the generation of 64 feature maps. This is followed by a Rectified Linear Unit (ReLU) activation layer. Subsequently, the image flows through three convolution blocks. Finally, a convolution layer with 3 kernels is applied to generate the enhanced underwater image.

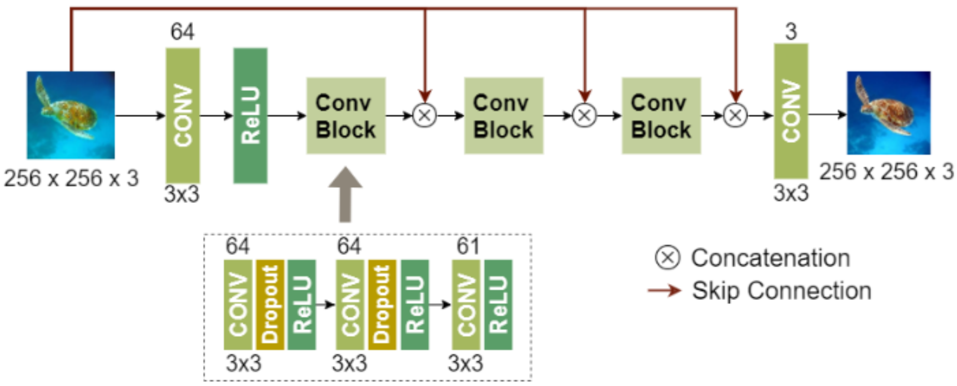

Fig. 1. Our proposed model architecture.

The convolutional blocks (convblocks)

Within our architecture, there are two sets of convolution layers, each followed by a dropout layer and a Rectified Linear Unit (ReLU) activation function. After passing through the ConvBlocks, the output is then fed into an additional Convolutional Layer with a Rectified Linear Unit (ReLU) activation function. This setup allows for the concatenation of the raw input image from the skip connection.
The purpose of these ConvBlocks, in combination with the skip connection, is to prevent overfitting of the network to the training data. By incorporating these blocks, we promote the

generalization ability of the network, allowing it to perform well on unseen data beyond the training set. This serves as a measure to mitigate the risk of overfitting and enhance the overall performance of the model.

Skip connections

In our model, we employ concatenation to merge the raw input image with the output of each residual block. This approach is specifically designed to tackle the issue of vanishing gradients by giving more significance to the channels associated with the raw input image when compared to the channels generated by the ConvBlocks.

By incorporating these skip connections, our model ensures that feature learning occurs in each block while also preserving crucial characteristics from the original raw image. This design allows the network to combine the learned features from the ConvBlocks with the intrinsic information contained in the base raw image, resulting in a more comprehensive and enriched representation of the input data.

The network loss function

Used to train the model encompasses multiple terms to achieve different objectives. It aims to preserve the sharpness of edges, enforce structural and texture similarity in the enhanced image, and account for pixel-wise differences. The loss function consists of two main components:

- **a- Mean Squared Error (MSE) Loss**: The MSE loss calculates the pixel-wise mean squared differences between the estimated image I and the clear ground truth image I*.

  $L_{MSE} = \frac{1}{N} \sum_{i=1}^{N} (I_i - I_i^*)^2$ (1)
- **b- VGG Perceptual Loss:** The perceptual loss, based on the work of Johnson, Alahi, and Fei-Fei (Johnson et al., 2016), is defined using the ReLU activation of the last convolutional layer in a pretrained VGG Network. Both the enhanced image I and the ground truth clear image I* are passed through this layer to obtain their respective feature representations. The perceptual loss, denoted as LVGG, measures the distance between these feature representations. L_{VGG} = distance(feature(I), feature(I*))

By incorporating these two loss components, the model learns to minimize both the pixel-wise differences and the perceptual differences between the enhanced image and the ground truth clear image, leading to improved image quality and fidelity.

The final loss L, is calculated as the summation of the two losses. $L_{TOTAL} = L_{MSE} + L_V$ (2)

EXPERIMENTAL EVALUATIONS

In order to assess the performance of our model, we conducted both qualitative and quantitative comparisons with state-of-the-art underwater image enhancement methods on synthetic and real-world underwater images. The methods we compared against include WaterNet (Li et al., 2019), FunIE-GAN (Islam et al., 2020), and Deep SESR (Islam et al., 2020). To ensure fairness in evaluation, we utilized the model checkpoints provided by the respective authors to obtain their best results.

Training and Validation Data For training and validation, we utilized the EUVP Underwater ImageNet dataset. This dataset consists of a diverse collection of images captured under various visibility conditions using different cameras such as GoPros and low-light USB cameras. Paired images with ground truth clear versions were generated using CycleGAN (Islam et al., 2020).The EUVP dataset offers location and perceptual quality diversity, making it suitable for training a model that can generalize to other underwater datasets. We trained our model on 6,128 images and validated it on 515 images. The input images had various resolutions, including 800 × 600, 640 × 480, 256 × 256, and 224 × 224, which were resized to 256 × 256 before training.

Network IMPLEMENTATION AND TRAINING

Our model was implemented using the PyTorch deep learning framework. We utilized the ADAM optimizer with a learning rate of 0.0002 for the training process. To prevent overfitting, dropout layers were applied with a rate of 0.2. The batch size was set to 1, meaning that one image was processed in each iteration during training.

Training the model over 50 epochs required approximately ten hours of computational time. The training process was performed on a system equipped with an Intel(R) Core(TM) i7-8750H CPU, 16GB RAM, and an Nvidia GTX 1060 GPU.

Testing data sets

To evaluate the transferability of our model to different datasets, we tested it on a variety of synthetic and real-world images. Specifically, we used the following datasets:

(a) EUVP Dark (Islam et al., 2020): Under the EUVP dataset, a separate dataset was created comprising 5,500 paired images with a dark underwater background. We chose a subset of 1,000 images from this dataset to evaluate our model.

(b) UIEB (Li et al., 2019): For simulating real-world underwater scenes, we utilized the Underwater Image Enhancement Benchmark Dataset (UIEBD). This dataset comprises 890 paired underwater images that were captured under various light conditions, exhibiting a diverse color range and degrees of contrast. To generate the reference images, meticulous pairwise comparisons were performed.

By evaluating our model on these various datasets, we aimed to assess its transferability and performance in different underwater image enhancement scenarios.

RESULTS

In this section, we provide an analysis of the experimental results, both quantitatively and qualitatively. Our proposed model exhibits comparable performance across all three test datasets, as illustrated in Table 1.

Table 1.Underwater image enhancement performance metric.

Metric	Dataset	Waternet	FunIEGAN	Deep SESR	Ours
PSNR	EUVP	25.33	26.36	26.11	**27.39**
	UIEB	19.11	**20.13**	19.36	18.99
SSIM	EUVP	0.80	0.82	0.81	**0.83**
	UIEB	**0.79**	0.73	0.73	0.69
UIQM	EUVP	**3.02**	2.84	2.66	2.77
	UIEB	2.69	2.69	2.65	**2.98**

Table 2.Model compression performance metric.

Models	# Parameters	Compression Ratio	Testing per image (secs)	Speed-Up
Our model	2,19,560	1	0.02	1
Waternet	10,60,636	3.96	0.60	25
FunIEGAN	25,62,023	11.17	0.18	7
Deep SESR	43,12,707	18.57	0.20	9

When evaluating model compression and acceleration, the proposed model exhibits a lower number of trainable parameters compared to all three state-of-the-art models, making it more lightweight for on-device deployment in different locations and conditions, as indicated in table 2.

Furthermore, our model demonstrates faster processing of test images, making it suitable for real-time underwater image enhancement applications.

Analyzing the results for each test dataset, as presented in Table 1, the following observations are made:

EUVP-Dark: our model outperforms all other models in terms of PSNR, SSIM, and UIQM metrics. Despite being trained on images with better lighting conditions, Shallow-UWnet effectively enhances color hues and sharpens the images. This highlights the generalization capabilities of Shallow-UWnet.

UIEB Dataset: the model also performs well on this near real-world dataset. WaterNet, trained on the UIEBD, achieves better performance than our model. Despite being trained on synthetic datasets, both our model and Deep SESR exhibit comparable performance ranges for this dataset.

A comparative visual analysis of the performance of these models on the three datasets can be observed in figure 2.

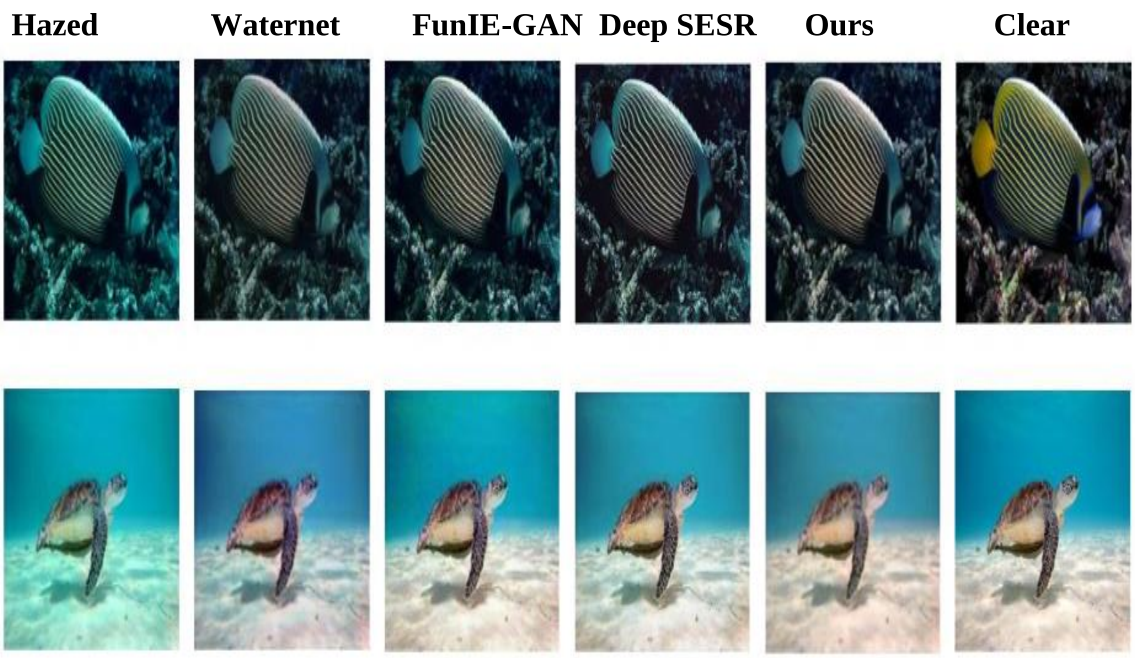

Fig. 2. Underwater image enhancement for two datasets and multiple models.

CONCLUSION

The application of convolutional neural networks in computer vision extends to underwater images. Our proposed model maintains comparable quantitative performance while requiring 18 times fewer trainable parameters and achieving 10 times faster testing. It is noteworthy that our model demonstrates generalization capabilities on diverse datasets, highlighting its potential for real-world applications.

REFERENCES

Almutiry, O., Iqbal, K., Hussain, S., Mahmood, A., Dhahri, H. (2021). Underwater images contrast enhancement and its challenges: a survey. *Multimedia Tools and Applications*, 1-26.

Anwar, S., Li, C. (2020). Diving deeper into underwater image enhancement: A survey. *Signal Processing: Image Communication*, *89*, 115978.

Cheng, Y., Wang, D., Zhou, P., Zhang, T. (2017). A survey of model compression and acceleration for deep neural networks. *arXiv preprint arXiv:1710.09282*.

Islam, M. J., Xia, Y., Sattar, J. (2020). Fast underwater image enhancement for improved visual perception. *IEEE Robotics and Automation Letters*, *5*(2), 3227-3234.

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

Johnson, J., Alahi, A., Fei-Fei, L. (2016). Perceptual losses for real-time style transfer and super-resolution. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14* (pp. 694-711). Springer International Publishing.

Li, C., Guo, C., Ren, W., Cong, R., Hou, J., Kwong, S., Tao, D. (2019). An underwater image enhancement benchmark dataset and beyond. *IEEE Transactions on Image Processing*, *29*, 4376-4389.

Li, Y., Wang, N., Liu, J., Hou, X. (2017). Demystifying neural style transfer. *arXiv preprint arXiv:1701.01036*.

Panetta, K., Gao, C., Agaian, S. (2015). Human-visual-system-inspired underwater image quality measures. *IEEE Journal of Oceanic Engineering*, *41*(3), 541-551.

Wang, W., Lai, Q., Fu, H., Shen, J., Ling, H., Yang, R. (2021). Salient object detection in the deep learning era: An in-depth survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(6), 3239-3259.

Yang, S., Xiao, W., Zhang, M., Guo, S., Zhao, J., Shen, F. (2022). Image data augmentation for deep learning: A survey. *arXiv preprint arXiv:2204.08610*.