2025 5th International Conference on Artificial Intelligence and Industrial Technology Applications (AIITA)

LLUW: Underwater image enhancement algorithm integrating deep learning and imaging models

1st Pengbo Li
International School Of BUPT
Beijing University Of Posts and Telecommunications
Beijing, China
2023213418@bupt.cn

***Abstract*—Underwater imaging is inherently challenging due to light attenuation and scattering, which leads to significant degradation in image quality. This paper addresses three key problems: color bias, low light, and blur, which are common in underwater environments.**

In this paper, we adapt our LLUW framework to address specific underwater image degradations. By refining the application of our contrast-guided atmospheric illumination models, we achieve a more tailored approach that mitigates color cast and blur while preserving crucial visual details. On the other hand, we develop a U-shape Transformer network, which is the first to integrate transformer models into underwater image enhancement (UIE) tasks. This network focuses on areas with severe attenuation by using a channel-wise multi-scale feature fusion transformer (CMFF) and a spatial-wise global feature modeling transformer (SGF). The results are enhanced images with improved contrast, saturation, and overall visual quality.

Finally, comparison of our methods with existing techniques and an ablation study to demonstrate the effectiveness of each component in our models. Our methods outperform existing techniques, as evidenced by comprehensive metric evaluations, including PSNR and SSIM, which are crucial for assessing image quality.

Keywords-Underwater image enhancement; Deep learning; Imaging models; Contrast-guided atmospheric illumination; Transformer network

I. INTRODUCTION

Underwater imaging is a complex task due to the optical properties of water, which can lead to significant image degradation. The need for underwater image enhancement (UIE) arises from the desire to improve visibility and extract more information from underwater scenes, which is crucial for various applications such as marine biology, oceanography, and underwater archaeology[1]. Traditional UIE methods often rely on heuristic approaches or physical models, which may not generalize well across different underwater environments. With the advent of deep learning, data-driven UIE methods have shown promise in addressing these challenges by learning complex patterns from large datasets. This paper presents a novel approach to UIE that leverages advanced machine learning techniques to enhance image quality and restore visual information lost due to underwater imaging artifacts.

II. ASSUMPTION

1. Uniform Water Composition: The model assumes that the water in which the images are captured has a relatively uniform composition. This means that the particulate matter and chemical composition do not vary significantly within the scene, which can affect light scattering and absorption.

2. Stationary Environment: The model assumes that the underwater environment is stationary during the capture of the image. This implies that there is no significant movement of water, marine life, or other elements that could cause motion blur or alter the scene dynamically.

3. Single Light Source: For the purpose of simplification, the model assumes a single dominant light source, which is typically the case in many underwater scenarios. This light source may be natural sunlight or artificial lighting used in underwater photography.

4. No Obstructions: The model assumes that there are no obstructions between the camera and the subject of the image. Obstructions such as bubbles, floating debris, or other particles close to the lens can cause additional scattering and distortion, which are not accounted for in this model.

III. RELATED WORKS

A. DCP Network

The Dark Channel Prior (DCP) serves as a fundamental principle in the de-hazing of single images. It posits that in a haze-free image, there exists at least one color channel where the intensity values are low across a certain spatial region. This prior is mathematically encapsulated in Equation below, which delineates the generation of an indistinct image I as a composite of two Additional parts: direct attenuation and airlight.[2]

$$I(X) = J(X)t(x) + A(1 - t(x))$$

In this formulation, $J(x)$denotes the haze-free image, $t(x)$ represents the transmission map that quantifies the amount of light reaching the camera without being scattered by the atmosphere. The direct attenuation $D(x) = J(x)t(x)$describes the scene radiance reduction due to medium properties. The airlight $V(x) = A\big(1 - t(x)\big)$ accounts for scattered light, potentially causing color shifts in hazy images. The dehazing objective is to determine A and t to retrieve the haze-free image J.

He et al. introduced the Dark Channel and the Dark Channel Prior (DCP) to estimate A and simplify t ′ s calulation. The dark channel is given by:

$$J^{\text{dark}}(x) = \min_{y \in \Omega\,(x)} \left(\min_{c \in \{r,g,b\}} I^{c}\,(y) \right)$$

979-8-3315-0976-7/25/$31.00 ©2025 IEEE

where (x) and (y) are spatial coordinates in (J^{dark}) and the hazy image (I), respectively.[3] The DCP asserts that non-sky pixels have one low intensity across color channels at least, making the dark channel predominantly dark intensities near 0.

The estimation of the global atmospheric illumination vector (A_{∞}^{c}) for each color channel is derived from the most luminous 0.1 % or 0.2% of pixels within the dark channel, which are indicative of the regions with the highest opacity to haze in the input image. Ancuti et al. introduced a method for determining the local atmospheric light intensities ($A_{L\infty}^{c}$) across various patches (Ψ), employing the subsequent equation to more accurately depict scenes characterized by low illumination levels.

$$A_{L\infty}^{c}(x) = \max_{y\in\Psi(x)}\left(\min_{z\in\Omega(y)}\left(I^{c}(z)\right)\right)$$

where (x, y, z) are spatial coordinates, and (I_{min}^{c})resents the minimum intensities within patch (Ω) on (I^{c}) . The local atmospheric lighting ($A_{L\infty}^{c}$) is then calculated by finding the maximum of these minimum intensities within a larger patch (Ψ), which is twice the size of (Ω), to account for the broader influence of lighting sources.

The transmission map (t) is calculated using Equation below:

$$t(x) = 1 - \omega \min_{y\in\Omega(x)}\left(\min_{c\in\{r,g,b\}} \frac{I^{c}(y)}{A_{\infty}^{c}}\right)$$

where (ω) (with ($0 \leq \omega \leq 1$)) controls the amount of haze preserved for realism. For local lighting estimation, (A_{∞}^{c}) is replaced by ($A_{L\infty}^{c}$) . The haze-free image ($J(x)$) is obtained using Equation:

$$J^{c}(x) = \frac{I^{c}(x) - A_{\infty}^{c}}{\max(t(x), t_{0})} + A_{\infty}^{c}$$

The term (t_0) serves as a lower bound for ($t(x)$) when it is close to zero, ensuring that a small amount of haze is maintained in densely hazy regions.

IV. METHOD

A. Traditional low Light Under Water(LLUW) Framework

While the contrast-guided approach offers satisfactory improvement for low-light underwater visuals, it oversimplifies the atmospheric lighting with a uniform three-channel value (A_{∞}) . This method assuming a predominantly white light source, is inadequate for low-light conditions. This approach leads to over-correction, diminishing crucial visual features like edges and textures.

To address these limitations, LLUW introduces a refined mechanism for underwater illumination enhancement. It utilizes local contrast as a guiding parameter to develop two distinct lighting distribution models: one with a narrow focus to capture fine details and another with a broader view to enhance brightness. Each model undergoes a unique dehazing process, and the results are integrated through a multi-scale fusion approach. This strategy retains the detail and darkness reduction from both models, significantly reducing darkness in the original images while accentuating intensity variations. The computational pipeline of LLUW is illustrated in Figure 1 below.

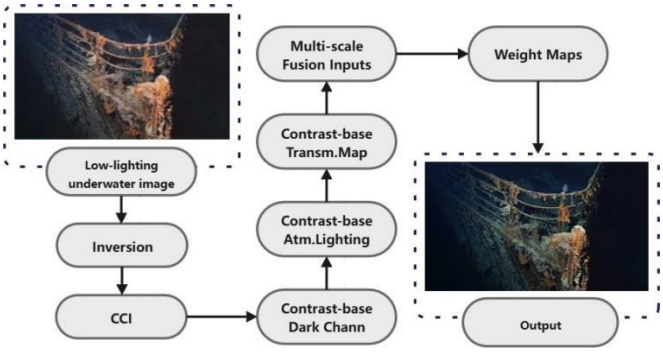

Figure 1 Detailed structure of LLUW Framework

B. Contrast-aware Model for Low-Light Scenes

Obtaining a unified global estimate (A_{∞}) for underwater imagery through the dark channel frequently leads to images that are excessively bright and lack detail. This is due to the fact that a single global estimate (A_{∞}) does not adequately capture the varied and often non-uniform lighting conditions present in underwater environments, especially when the raw data includes artificial and multiple light sources.

To address these challenges, the Local Lighting Underwater (LLUW) method computes ambient illumination for individual color channels and incorporates the Contrast Code Image (CCI) at each spatial location x to refine local lighting estimates. This technique employs a contrast-driven patch for calculating lighting, where regions exhibiting high contrast are associated with larger patches to reflect the diverse impacts of light sources. Conversely, areas with lower contrast necessitate smaller patches. The mathematical relationship linking the CCI value c to the dimensions of the lighting patch T is articulated in the following equation:

$$S_{T}(m, c) = 3m - \left[\frac{m}{3}(c-1)\right]$$

where (m) is a multiplication factor and (c) represents the CCI code at a position. This equation ensures that smaller patches are more influenced, thus progressively constraining the patch size based on the contrast.

The approach diverges from the conventional ($A_{L\infty}^{c}$) by employing patches that are sensitive to local contrast, denoted as ($\Upsilon(x, m)$), rather than utilizing patches of a uniform size. The ($A_{LCG\infty}$) parameter encapsulates the chromatic attributes of the illumination as well as its spatial distribution across the image. To circumvent the emergence of sharp, square-shaped intensity disparities, commonly referred to as 'halos,' when the input image is normalized against the atmospheric lighting, ($A_{LCG\infty}$) undergoes a smoothing process using a Gaussian filter with a standard deviation of 10. This filtering step is essential for preserving the authentic visual quality of the enhanced images.

C. Fusion Process

The parameter (m) in Equations above dictates the size of the local window (Υ), influencing the extent of illumination from each simple. A higher (m) results in brighter lighting models, reducing darkness but potentially overcorrecting radiance, leading to saturation and loss of important intensity details.

To balance these effects, we generate two models of $(A_{LCG\infty})$with $(m = 5)$ and $(m = 30)$. These models, along with Equation above, determine two transmission maps, which are then refined using a Fast Guided filter and used to produce two haze-reduced versions of the original image via Equation above. These enhanced images, $(\mathcal{I}^{k}(k = \{1,2\}))$, are inputs to a multi-scale fusion process.

By combining inputs from various $(A_{LCG\infty})$, we retain the efficient shadow elimination from the $(m = 30)$ input and the edge, texture, and intensity details from the $(m = 5)$ input. This approach, confirmed by experiments, outperforms other (m) value combinations and is more computationally efficient than using additional inputs.

The fusion process combines these inputs using three weight maps: saliency, luminance, and local contrast. These maps ensure that regions with high saliency, contrast, or texture variations are emphasized in the output. The saliency weight map is calculated using[4]:

$$\mathcal{W}_{S}^{k}(x) = |\mathcal{I}_{k}^{\mathcal{G}_s}(x) - \mathcal{I}_{k}^{u}|$$

The luminance weight map is calculated using:

$$\mathcal{W}_{\mathcal{L}}^{k} = \sqrt{\frac{1}{3}[(R^{k} - L^{k})^{2} + (G^{k} - L^{k})^{2} + (B^{k} - L^{k})^{2}]}$$

The local contrast weight map is calculated by applying a Laplacian kernel on (L^{k}).

These weight maps are normalized and combined into a Gaussian pyramid, which is effective for representing weight maps. Each input is decomposed into a 5-level Laplacian pyramid, and the multi-scale fusion process is executed using:

$$R_{l}(x) = \sum_{k} G_{l}\{\overline{W_{k}}(x)\}L_{l}\{\mathcal{I}_{k}(x)\}$$

The resulting LLUW output significantly reduces darkness, retains color fidelity, and enhances visual features, as demonstrated in Figure below.

V. Experiments

A. Result Analysis

The performance of LLUW is assessed using the test images from Attachment1 dataset, which consists of 400 underwater images. Evaluating image improvements tools in datasets lacking ground truth, presents challenges. No single metric can fully capture the enhancement's effectiveness. For instance, the e-score measures the number of newly visible edges, which could also indicate noise, while the FADE score measures darkness without considering over-illumination effects.[5] Therefore, a suite of seven metrics is employed:

1. UIQM, a no-reference underwater image quality indicator inspired by the human visual system, assesses colorfulness, sharpness, and contrast.

2. PCQI, designed for evaluating contrast-changed images, analyzes local contrast quality.

3. GCF reflects the overall image contrast, with dehazed images typically showing higher contrast levels.

4. The e-score and r-score indicate the increase in visible edges and the enhancement in gradient values for these edges, respectively.

5. FADE measures perceived fog or darkness in inverted images, with lower values indicating better illumination.

6. SURF, a widely used method, extracts features crucial for image matching, reconstruction, stitching, and object detection.

This multi-metric approach ensures a thorough evaluation of LLUW's performance in enhancing low-light underwater images.

Our innovative single-image framework, designated as LLUW, utilizes a contrast-driven methodology to effectively simulate lighting conditions within dimly lit underwater environments. This technique produces a pair of dehazed images, which are subsequently integrated via a multi-scale fusion mechanism. Such a procedure notably diminishes the presence of dark regions while enhancing the salient visual attributes of the original photograph, concurrently ensuring the integrity of its color profile. Comparative analyses of the image enhancement outcomes, along with samples for calculating the mean deviation across seven distinct metrics, are presented in Figure 2. [6]

For the practical enhancement of underwater visibility, a viable strategy involves the implementation of a dual-model approach. In this context, the atmospheric illumination model is constructed based on local contrast measurements, followed by the application of a multi-scale fusion technique to generate an image that emphasizes fine details and mitigates darkness. The LLUW framework can achieve these objectives.

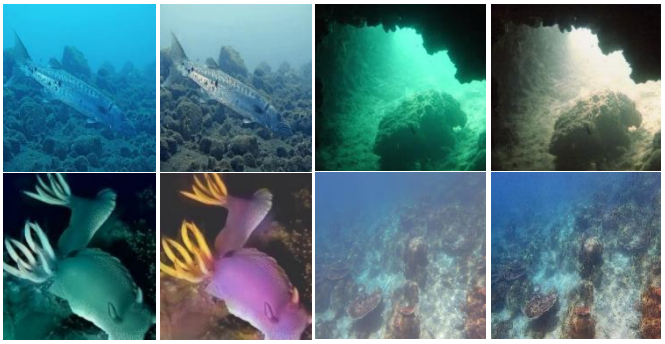

Figure 2 Result on samples from dataset

B. Future Work

Additionally, the development and utilization of large-scale, high-fidelity underwater image datasets are essential for training and validating enhancement models. The fusion of these strategies promises to enhance the visual quality and perceptual clarity of underwater imagery, facilitating more

effective monitoring and analysis of subaquatic scenes. Regarding this problem, is there any room for improvement in using traditional methods to establish models?

There is potential for enhancing traditional underwater image enhancement (UIE) methods. While these methods have been foundational, they can be augmented by integrating machine learning techniques. For instance, incorporating traditional color correction into deep learning frameworks might refine initial feature extraction, leading to more robust UIE models. Additionally, merging physical light behavior models with data-driven approaches could create hybrid models that benefit from both the interpretability of traditional methods and the generalization ability of machine learning.

VI. Conclusion

Our paper presents a comprehensive study on underwater image enhancement (UIE), introducing novel methodologies to address the challenges posed by light absorption and scattering in underwater environments. Our proposed solutions, including the LLUW framework and the U-shape Transformer network, demonstrate significant advancements in enhancing image quality and restoring visual information. The LLUW framework, with its contrast-guided atmospheric illumination models, effectively mitigates color cast and blur while preserving crucial details. Meanwhile, the U-shape Transformer leverages the power of transformer models to focus on areas with severe attenuation, leading to improved contrast and saturation in the enhanced images. The multi-color space loss function further refines the training process by integrating RGB, LAB, and LCH color spaces, ensuring a more natural and vivid color representation. Extensive experiments and ablation studies validate the effectiveness of our methods, showing superior performance over existing techniques.

Reference

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[2] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing*, vol. 452, pp. 48–62, Jan. 2021.

[3] X. Chen, H. Li, M. Li, *et al.*, "Learning a sparse transformer network for effective image deraining," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5896–5905, 2023.

[4] S. Minaee, Y. Boykov, F. Porikli, *et al.*, "Image segmentation using deep learning: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 7, pp. 3523–3542, Jul. 2021.

[5] R. Rombach, A. Blattmann, D. Lorenz, *et al.*, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 10684–10695.

[6] P. Dhariwal and A. Nichol, "Diffusion models beat GANs on image synthesis," *Advances in Neural Information Processing Systems*, vol. 34, pp. 8780–8794, 2021.