UWM-Net: A Mixture Density Network Approach with Minimal Dataset Requirements for Underwater Image Enhancement

2024 IEEE Conference on Artificial Intelligence

Jun Huang[1], Zongze Li[1], Ruihao Zheng[1], Zhenkun Wang[12]

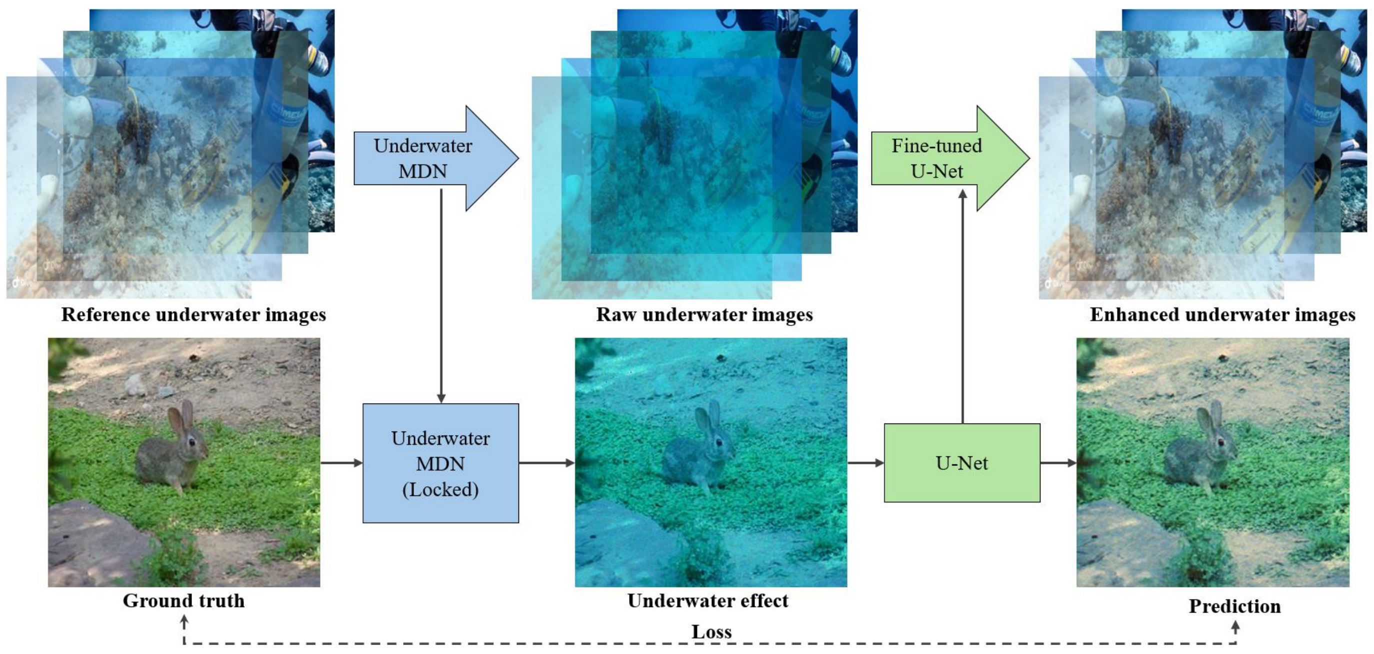
[1]School of System Design and Intelligent Manufacturing, Southern University of Science and Technology, Shenzhen, China

[2]Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, China

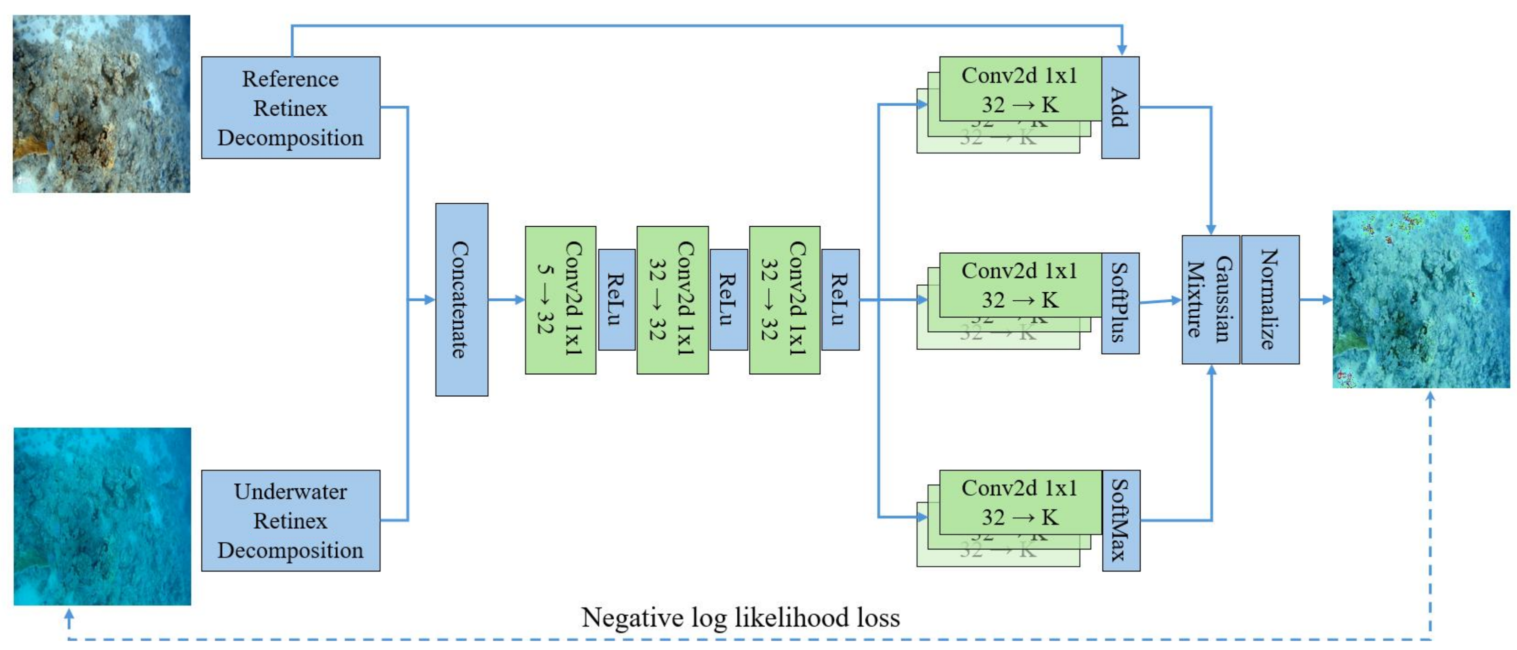
Jun Huang and Zongze Li contribute equally. Corresponding author: Ruihao Zheng and Zhenkun Wang.

Abstract: The learning-based underwater image enhancement, which is suitable for batch processing, is a pivotal research direction in underwater image processing. Extensive paired image data are required in existing learning-based methods, which necessitate considerable preprocessing and hinder the application of these methods. To address these limitations, we propose a semi-supervised approach called UWM-Net: firstly, we use a compact dataset of underwater image pairs to train the Mixture Density Network (MDN) with an underwater scene setting; subsequently, U-Net can learn underwater image enhancement more efficiently. The MDN can transform standard images into underwater scenes, reducing the reliance on paired data and making much smaller training datasets. In experimental studies, UWM-Net using only 18 pairs of underwater image data achieves highly competitive results in terms of 3 metrics compared with advanced models.

The overall framework of UWM-Net



MDN



Our main contributions are summarized as follows:

- **We propose UWM-Net, which only requires a small size of training datasets and effectively alleviates the shortage of underwater images.**
- **We adapt MDN to the light attenuation characteristics of underwater images.**
- **The experimental results indicate the images enhanced by UWM-Net improve the Structural Similarity Index Measure (SSIM) and Peak Signal-to-Noise Ratio (PSNR) by about 25% and 50%, respectively.**

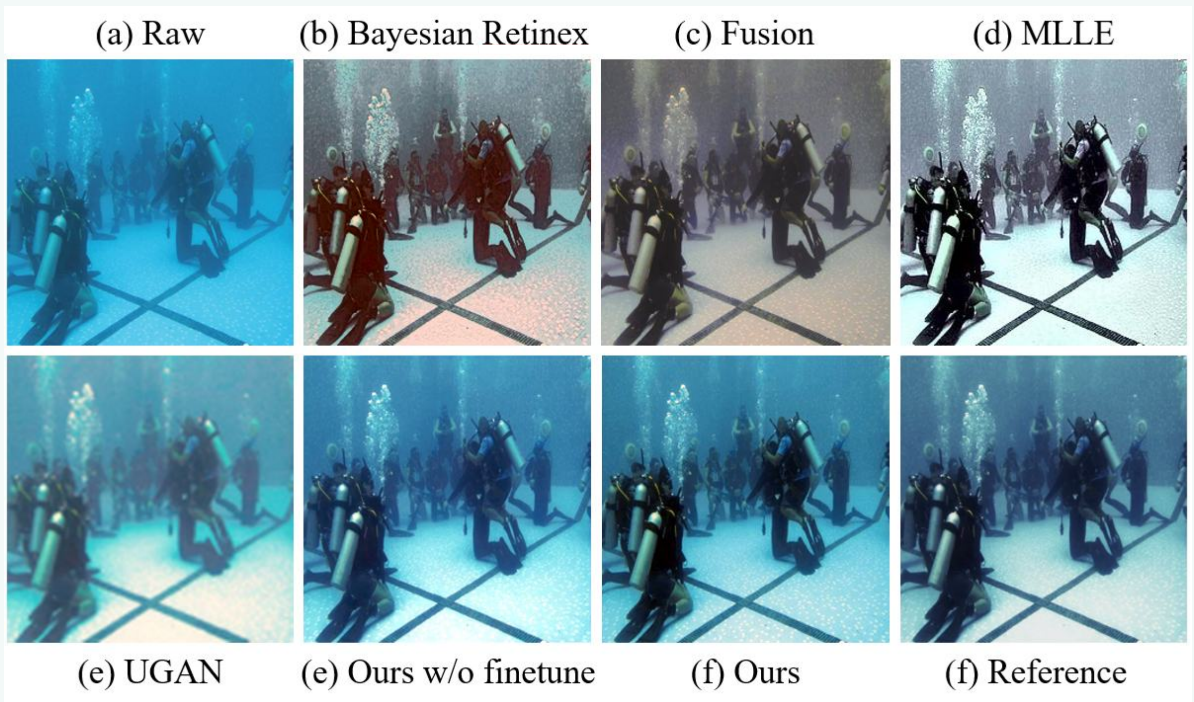


Fig. 2: The qualitative comparison

The mathematical formulation for optimizing the MDN parameters (encompassing the weights and biases of the MDN network), denoted as θ, is achieved by:

$$\theta^* = \arg\min_{\theta} \sum_{n} \sum_{i,j,k} -\frac{\log\, p\left(r_{D,i,j,k}^{(n)} \mid \mathbf{x}_{i,j}^{(n)}; \theta\right)}{L_{i,j}^{(n)}}, \quad (1)$$

TABLE I: Comparison results of existing methods. "**bold**" = the best score; "underlined" = the second-best score.

Image	SSIM	PSNR	UCIQE
Raw	N/A	N/A	0.4301
Bayesian Retinex	0.7552	16.42	0.7441
Fusion	0.7760	15.59	0.6722
MLLE	0.6158	13.81	**2.7160**
UGAN	0.5563	16.32	0.4881
UWM-Net	**0.9265**	**23.75**	0.7655
Reference	N/A	N/A	1.1119

We compare our methods with some advanced models, and comparison is shown in Fig. 2. Results are summarized in Table I. UWM-Net consistently shows superiority, achieving top scores in terms of the SSIM and PSNR metrics, thereby confirming its effectiveness in restoring image quality. It also exhibits competitive UCIQE metric values, demonstrating its robustness across different underwater scenarios. These results highlight the model's proficiency in learning pixel-level improvements, catering to the intricate details necessary for high-fidelity underwater image reconstruction.

The training of the network considers multiple loss functions, including Mean Squared Error (MSE), Structural Similarity Index Measure (SSIM), and a novel color histogram loss. Specifically, the training loss is a weighted sum of the three loss components:

$$L_{\text{total}} = \omega_{\text{MSE}} \cdot L_{\text{MSE}} + \omega_{\text{SSIM}} \cdot L_{\text{SSIM}} + \omega_{\text{hist}} \cdot L_{\text{hist}}$$

$$L_{\text{SSIM}} = 1 - \text{SSIM}(Y, \hat{Y})$$

$$L_{\text{MSE}} = \frac{1}{N}\sum_{i=1}^{N} \left(Y_i - \hat{Y}_i\right)^2$$

$$\text{SSIM}(Y, \hat{Y}) = \frac{(2\mu_Y\mu_{\hat{Y}} + C_1)(2\sigma_{Y\hat{Y}} + C_2)}{(\mu_Y^2 + \mu_{\hat{Y}}^2 + C_1)(\sigma_Y^2 + \sigma_{\hat{Y}}^2 + C_2)}$$

$$L_{\text{hist}} = \frac{1}{C}\sum_{c=1}^{C} \left| \frac{\text{hist}_c(\hat{Y}, B)}{\sum_{b=1}^{B} \text{hist}_c(\hat{Y}, B)} - \frac{\text{hist}_c(Y, B)}{\sum_{b=1}^{B} \text{hist}_c(Y, B)} \right|$$

where ω_{MSE}, ω_{SSIM}, and ω_{hist} are the weights assigned to the MSE, SSIM, and color histogram loss, respectively.
We choose the following weights to balance the contribution of each component: $\omega_{\text{MSE}} = 1$, $\omega_{\text{SSIM}} = 0.05$, and $\omega_{\text{hist}} = 1$.

In the ablation study, shown in Fig. 3, we systematically investigate the contribution of each component in UWM-Net. The study compares the following results: raw images, results without applying color histogram loss, results without the finetuning step, and the final enhanced images after fine-tuning (i.e., the results of UWM-Net). Results are shown in Table II. We can find that both the color histogram loss and the finetuning step have positive effects on UWM-Net. Incorporating color histogram loss significantly improves color correction and balance, as demonstrated by UCIQE metrics. Furthermore, the fine-tuning phase leverages limited underwater images to further refine the enhancement by sharpening details and enriching colors. Although SSIM marginally decreases, the PSNR score shows a substantial improvement.

TABLE II: Ablation study on key components of UWM-Net. "**bold**" = the best score; "underlined" = the second-best score.

Condition	SSIM	PSNR	UCIQE
Reference	N/A	N/A	0.7292
w/o Histogram Loss	0.8886	21.34	0.5613
w/o Fine-Tuning	**0.9210**	24.32	0.5841
UWM-Net	0.9197	**24.91**	**0.6461**

(a) (b) (c) (d) (e)

Fig. 3: Illustrations of the ablation study on the color histogram loss and fine-tuning step, where (a) Raw, (b) w/o Histogram Loss, (c) w/o Fine-Tuning, (d) UWM-Net, (e) Reference.