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

Jun Huang, Zongze Li, Ruihao Zheng
School of System Design and Intelligent Manufacturing,
Southern University of Science and Technology,
Shenzhen, China
{12111119, 12111802, 12132686}@mail.sustech.edu.cn

Zhenkun Wang
School of System Design and Intelligent Manufacturing
& Department of Computer Science and Engineering,
Southern University of Science and Technology,
Shenzhen, China
wangzhenkun90@gmail.com

***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.**

***Index Terms*—semi-supervised learning, mixture density network, reduced dataset, color distortion**

I. Introduction

Underwater images encounter unique issues (*e.g.*, poor contrast, skewed color balance, light attenuation, and blurred detail) due to the scattering and absorption of light in aquatic environments [1]. Therefore, underwater image enhancement (UIE) plays an important role in underwater image processing. Conventional methods, such as histogram equalization [2], gamma correction [3], and retinex theory [4] [5], have been developed to mitigate poor visibility and color distortion. Recent developments shift towards deep-learning-based methods, offering more sophisticated and accurate solutions. Novel approaches, including leveraging local color distributions [6] and semantic-aware knowledge guidance [7], are introduced to UIE. Emerging techniques like zero-reference learning [8] and unsupervised learning [9] show promise for UIE, as they can enhance image quality with fewer labeled datasets.

This work is supported by the National Natural Science Foundation of China (Grant No. 62106096), the Characteristic Innovation Project of Colleges and Universities in Guangdong Province (Grant No. 2022KTSCX110), the Shenzhen Technology Plan, China (Grant No. JCYJ20220530113013031), and the 2023 Graduate Innovation Practice Fund Project of Southern University of Science and Technology. (*Jun Huang and Zongze Li contribute equally.*) (*Corresponding author: Ruihao Zheng and Zhenkun Wang.*)

However, existing methods depend on large-scale and paired data for training. Obtaining these datasets is challenging and costly, restricting the generalization of the model to diverse underwater conditions [10]. To fill this gap, we propose an Underwater Mixture Density Network (UWM-Net), combining deep learning techniques with advanced image processing methods to improve the performance of UIE. UWM-Net employs a semi-supervised framework with the Mixture Density Network (MDN) [11], requiring a smaller size of paired data without performance deterioration and overcoming the scarcity of high-quality underwater datasets. In the experimental studies, UWM-Net effectively enhances underwater images, improving clarity and color accuracy. 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.

II. Proposed Method

UWM-Net first employs the MDN pre-trained on a minimal set of paired underwater images, facilitating the initial transformation of standard images into synthetic underwater conditions. Then, a U-Net architecture [12] is conducted for UIE tasks, utilizing the transformed images as an expanded training set. The MDN, a low-light image enhancement algorithm, is an essential component. Inspired by Dimma [13], it serves as a generative model to simulate underwater image features. This simulation creates a bridge between limited real-world underwater image pairs and a more expansive training regime for the U-Net model. Consequently, UWM-Net can efficiently learn to correct color distortions, improve image clarity, and adapt to a variety of underwater conditions.

979-8-3503-5409-6/24/$31.00 ©2024 IEEE
DOI 10.1109/CAI59869.2024.00269

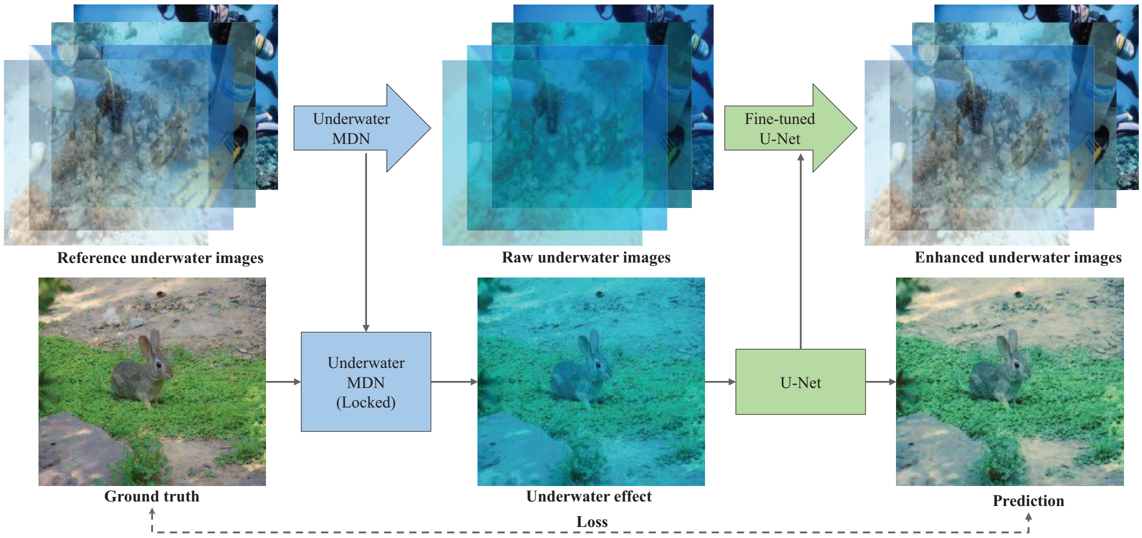


Fig. 1: The overall framework of UWM-Net.

A. MDN

The MDN implements a novel approach to image processing, commencing with the application of Retinex theory for image decomposition. This procedure decomposes an input image into two separate features: reflectance, which suggests the color information (independent of lighting conditions), and luminance, which represents the brightness aspect of the image. Such decomposition is essential for isolating the color distortions induced by underwater environments. Thereby, the MDN employs a series of convolutional layers, each followed by a non-linear activation function, to process the decomposed image. The network architecture, comprising a sequence of Conv2D and ReLU layers, can capture the intricate patterns and nuances necessary for the subsequent reconstruction phase. The network utilizes the estimated Gaussian mixture model parameters to reconstruct the color channels of the underwater image. And we achieve the regularization by multiplying it by a random adjustment factor.

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

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

where $r_{D,i,j,k}^{(n)}$ represents the pixel intensity of the k_{th} color channel at position (i, j) in the n_{th} image of the transformed (underwater-like) dataset; $x_{i,j}^{(n)}$ denotes the input feature vector for the corresponding pixel, extracted from the decomposed components of the image. Based on the structure of Dimma [13], Our formula incorporates a normalization term (*i.e.*, $L_{i,j}^{(n)}$) within the equation. $L_{i,j}^{(n)}$ is derived from the luminance component of the original image, representing the normalized luminance factor for the corresponding pixel. It adjusts the loss weight of each pixel to compensate for typical underwater lighting conditions such that the variation of regions distant from the camera can be effectively captured.

B. U-Net

The U-Net architecture, renowned for its efficacy in image segmentation tasks, is used in UWM-Net to enhance images together with the MDN. The U-Net model of UWM-Net employs a series of convolutional and up-convolutional layers to process input images, aiming to recover the color and clarity commonly degraded in underwater settings. Besides, the network architecture includes attention mechanisms to focus on important features, further enhancing the ability to restore precise details and true colors. The training of the network considers multiple loss functions, including Mean Squared Error (MSE), Structural Similarity Index Measure (SSIM) [14], and a novel color histogram loss. Specifically, the training loss L_{total} is a weighted sum of the three loss components:

$$L_{\text{total}} = \omega_{\text{MSE}} \cdot L_{\text{MSE}} + \omega_{\text{SSIM}} \cdot L_{\text{SSIM}} + \omega_{\text{hist}} \cdot L_{\text{hist}}, \quad (2)$$

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 $\omega_{\text{hist}} = 1$.

1) MSE Loss: The MSE loss is defined as the average of the squares of the differences between the predicted and target pixel values, which is mathematically represented as:

$$L_{\mathbf{MSE}} = \frac{1}{N} \sum_{i=1}^{N} \left(Y_i - \hat{Y}_i\right)^2, \quad (3)$$

where Y_i is the ground truth, $\hat{Y}_i$ is the prediction for the i_{th} pixel, and N denotes the total number of pixels. This loss emphasizes pixel-level accuracy and maintains image integrity.

2) SSIM Loss: The SSIM loss evaluates the perceived quality of the predicted image by comparing its structural similarity with the target image. It is defined as:

$$L_{\mathbf{SSIM}} = 1 - \mathrm{SSIM}(Y, \hat{Y}), \tag{4}$$

where $\mathrm{SSIM}(Y, \hat{Y})$ is the structural similarity metric between the target image Y and the predicted image $\hat{Y}$. The SSIM metric is defined as:

$$\mathrm{SSIM}(Y, \hat{Y}) = \frac{(2\mu_Y \mu_{\hat{Y}} + C_1)(2\sigma_{Y\hat{Y}} + C_2)}{(\mu_Y^2 + \mu_{\hat{Y}}^2 + C_1)(\sigma_Y^2 + \sigma_{\hat{Y}}^2 + C_2)}, \tag{5}$$

where μ_Y, $\mu_{\hat{Y}}$ are the average pixel values; σ_Y^2, $\sigma_{\hat{Y}}^2$ are the variances of the target image Y and the predicted image $\hat{Y}$ respectively; $\sigma_{Y\hat{Y}}$ is the covariance of Y and $\hat{Y}$; C_1 and C_2 are constants to stabilize the division.

3) Color Histogram Loss: The color histogram loss aims to align the color distributions between the predicted and target images. It calculates the L_1 distance between their normalized color histograms:

$$L_{\text{hist}} = \frac{1}{C} \sum_{c=1}^{C} \left| \frac{\text{hist}_c(\hat{Y}, B)}{\sum_{b=1}^{B} \text{hist}_c(\hat{Y}, B)} - \frac{\text{hist}_c(Y, B)}{\sum_{b=1}^{B} \text{hist}_c(Y, B)} \right|, \tag{6}$$

where C is the number of color channels, B denotes the number of histogram bins, and $\text{hist}_c(Y, B)$ represents the histogram of the c-th channel.

C. Fine-Tune

In the final phase of UWM-Net's procedure, we fine-tune the U-Net network to refine the model's performance in real-world underwater scenarios. For this purpose, the same U-Net architecture, inherited from the previous phase, is employed. However, instead of relying on the MDN-generated synthetic images, the network is fine-tuned using a small but highly representative dataset comprising the initial 10 pairs of real-world underwater images. These images, which were instrumental in the initial training of the MDN, now serve as a more authentic and challenging dataset for fine-tuning the U-Net. This method ensures that the network proficiently manages the simulated underwater conditions produced by the MDN while also being finely tuned to the details and complexities of real-world underwater environments.

III. Experiments

A. Experiment Settings

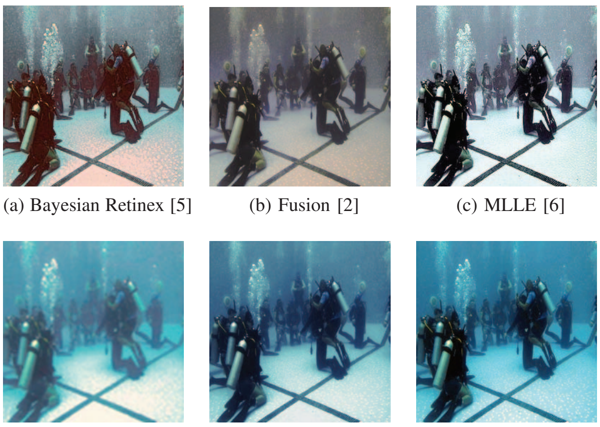
Our experiments run on a hardware platform consisting of an Intel(R) Xeon(R) Gold 5320 CPU @ 2.20GHz and an NVIDIA A30 GPU with 32GB memory. We train our MDN model using the Adam optimizer with a learning rate of 1e-3 and weight decay of 5e-4 for 1000 epochs. Then we train U-Net employing the Adam optimizer with a cosine annealing scheduler and a learning rate adjusted to 5e-4 for 2000 epochs.

We utilize the Underwater Image Enhancement Benchmark (UIEB) [15] and the MixHQ dataset [13] as training datasets for UWM-Net. From the UIEB dataset, we select 18 pairs of images: 10 for training, 4 for validation, and 4 for testing. The images across the dataset are standardized to a resolution of 256x256 pixels with three color channels.

To evaluate the models, we adopt two full-reference metrics: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) [14]. These metrics are widely recognized for their effectiveness in assessing the quality of enhanced images against their original counterparts. Additionally, we also include a no-reference image quality assessment metric called the Underwater Color Image Quality Evaluation (UCIQE) [16]. This metric offers a comprehensive evaluation of image quality by focusing on crucial aspects such as chromaticity, contrast, and sharpness, which are particularly susceptible to degradation in underwater imaging conditions.

B. Comparison with Representative Models

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.



(a) Bayesian Retinex [5] (b) Fusion [2] (c) MLLE [6]

(d) UGAN [17] (e) Ours (f) Ground Truth

Fig. 2: The qualitative comparison shows that our methods provide more details and a better color enhancement.

C. Ablation Study

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 fine-tuning step, and the final enhanced images after fine-tuning

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

(*i.e.*, the results of UWM-Net). Results are shown in Table II. We can find that both the color histogram loss and the fine-tuning 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.

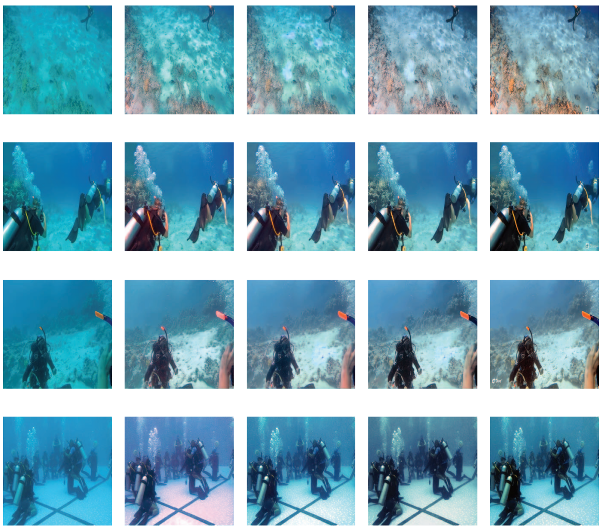


Fig. 3: Illustrations of the ablation study on the color histogram loss and fine-tuning step, where (a) Raw, (b) w/o Histogram Loss, (c) w/o Fine-Tuning, (d) UWM-Net, (e) Reference.

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**

IV. Conclusions

This study addresses the requirement for effective UIE with limited paired data. The proposed UWM-Net leverages a combination of the MDN and the modified U-Net, enabling significant improvements in image quality with limited training data (18 pairs of underwater images). Experimental results validate the efficiency of our method. Particularly, significantly high SSIM and PSNR scores have been achieved by our method, showing superior clarity and fidelity of enhanced underwater images.

Note that the UCIQE metric value obtained by UWM-Net only secures the second place. In the future, we will concentrate on refining the color correction capability of UWM-Net to achieve greater adaptability and accuracy across diverse underwater conditions.

References

[1] H. Lu, Y. Li, Y. Zhang, M. Chen, S. Serikawa, and H. Kim, "Underwater optical image processing: a comprehensive review," *Mobile Networks and Applications*, vol. 22, pp. 1204–1211, 2017.

[2] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and P. Bekaert, "Color balance and fusion for underwater image enhancement," *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 379–393, 2017.

[3] S. Raveendran, M. D. Patil, and G. K. Birajdar, "Underwater image enhancement: a comprehensive review, recent trends, challenges and applications," *Artificial Intelligence Review*, vol. 54, pp. 5413–5467, 2021.

[4] M. Li, J. Liu, W. Yang, X. Sun, and Z. Guo, "Structure-revealing low-light image enhancement via robust retinex model," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 2828–2841, 2018.

[5] P. Zhuang, C. Li, and J. Wu, "Bayesian retinex underwater image enhancement," *Engineering Applications of Artificial Intelligence*, vol. 101, p. 104171, 2021.

[6] W. Zhang, P. Zhuang, H.-H. Sun, G. Li, S. Kwong, and C. Li, "Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement," *IEEE Transactions on Image Processing*, vol. 31, pp. 3997–4010, 2022.

[7] Q. Qi, K. Li, H. Zheng, X. Gao, G. Hou, and K. Sun, "SGUIE-Net: Semantic attention guided underwater image enhancement with multi-scale perception," *IEEE Transactions on Image Processing*, vol. 31, pp. 6816–6830, 2022.

[8] Y. Huang, F. Yuan, F. Xiao, J. Lu, and E. Cheng, "Underwater image enhancement based on zero-reference deep network," *IEEE Journal of Oceanic Engineering*, 2023.

[9] B. Sun, Y. Mei, N. Yan, and Y. Chen, "UMGAN: Underwater image enhancement network for unpaired image-to-image translation," *Journal of Marine Science and Engineering*, vol. 11, no. 2, p. 447, 2023.

[10] S. Raveendran, M. D. Patil, and G. K. Birajdar, "Underwater image enhancement: a comprehensive review, recent trends, challenges and applications," *Artificial Intelligence Review*, vol. 54, pp. 5413–5467, 2021.

[11] C. M. Bishop, "Mixture density networks," 1994.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer, 2015, pp. 234–241.

[13] W. Kozłowski, M. Szachniewicz, M. Stypułkowski, and M. Zięba, "Dimma: Semi-supervised low light image enhancement with adaptive dimming," *arXiv preprint arXiv:2310.09633*, 2023.

[14] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[15] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao, "An underwater image enhancement benchmark dataset and beyond," *IEEE Transactions on Image Processing*, vol. 29, pp. 4376–4389, 2019.

[16] M. Yang and A. Sowmya, "An underwater color image quality evaluation metric," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 6062–6071, 2015.

[17] D. Zhu, S. Liu, W. Jiang, C. Gao, T. Wu, Q. Wang, and G. Guo, "Ugan: Untraceable gan for multi-domain face translation," *arXiv preprint arXiv:1907.11418*, 2019.