Scene-adaptive Underwater Image Enhancement With State-Space Model

Scene-adaptive Underwater Image Enhancement With State-Space Model

Zhu Kou*, **Yu Wang, Ziying Li**

Dalian Naval Academy, Dalian, Liaoning Province, 116018, P.R. China

*Corresponding author's e-mail: 735272740@qq.com

Abstract. The combination of light absorption and scattering by underwater impurities and inhomogeneous illumination leads to the complex and diverse degradation of underwater environmental images, which inevitably affects the effectiveness of underwater visualization tasks. In recent years, underwater image enhancement techniques, especially neural network-based methods, have made great progress. However, since acquiring high-quality ground truth images is a major challenge, existing methods can only generate approximate reference maps, thus limiting the processing performance of their networks. Also, it is difficult to create global pixel-to-pixel dependencies since augmented networks often face limited perceptual regions. To address these issues, we propose a novel probability distribution framework, Mamba-SAU, which decomposes UIE into two phases: augmented distribution estimation and consensus refinement, effectively mitigating the approximation error induced by the reference. Drawing inspiration from the selective structured state space model (SSM), which can capture global dependencies with less complexity, we integrate the Mamba module into the underwater image enhancement network to establish global context awareness while maintaining computational efficiency. Experimental results show that our proposed Mamba-SAU approach outperforms state-of-the-art underwater image contrast enhancement algorithms and is validated by subjective and objective evaluation metrics.

1. Introduction

As a challenging technology in the field of underwater vision, underwater image enhancement is of great value in application scenarios such as marine exploration, underwater archaeology, and underwater robotics, and its core objective is to acquire high-quality underwater image data and to support underwater environment research [1][2]. Due to the complex lighting conditions underwater and the combined effects of dissolved impurities and suspended particulate matter, underwater images often show significant degradation. This degradation is mainly caused by a combination of wavelength-dependent light scattering [3] and light absorption [4]: the light absorption process triggers a sharp decrease in colour saturation and colour artefacts, while the light scattering effect enhances the overall brightness value of the image. At the same time, the non-uniform light distribution further reduces the light intensity in localized areas, and this multiple effect ultimately leads to the loss of image detail information and overall contrast degradation. Underwater image restoration typically suffers from different types of image noise compared to a single noise in other tasks, such as low light enhancement and super-resolution. Therefore, clear underwater images are crucial in vision-related underwater tasks. The main goal of underwater image enhancement is to obtain high-quality images by removing scattering and correcting colour distortions in degraded images.

Early underwater image enhancement methods relied on physical models like Retinex and degradation imaging models [5][6][7], which performed adequately in simple scenarios (e.g., shallow water) but struggled with complex conditions (e.g., low illumination, heavy blur) due to inherent limitations in modelling coupled optical-physical mechanisms. Recent deep learning approaches achieve superior performance through neural networks' nonlinear mapping capabilities yet face critical dependency on high-quality training data. Current reference image construction involves generating candidate images via advanced algorithms followed by subjective manual selection, introducing uncertainties from human bias and parameter variability. This intrinsic uncertainty undermines supervised learning frameworks that establish deterministic degraded-to-clean image mappings, compromising model robustness and reliability.

Underwater image enhancement networks achieve performance gains through receptive field expansion, which enhances global context capture by integrating broader pixel regions and activating relevant spatial patterns. However, exponential computational growth from enlarged receptive fields creates deployment bottlenecks. The Mamba architecture [8] addresses this through discretized state-space equations and structured reparameterization, enabling efficient long-range pixel dependency modelling while maintaining global processing capacity, thus overcoming traditional models' complexity limitations.

To address the highly dynamic and scene heterogeneity characteristics of underwater environments, this study innovatively draws on the core idea of the style migration task to construct a probabilistic-driven generative network framework. Firstly, a probabilistic modelling network performs the multimodal prediction of potential enhancement results. While extracting the shallow texture features of the image, the architecture synchronously estimates the semantic distribution features of the reference image during the training process, utilizing the dynamic distributional alignment mechanism to establish a fictionalizable probabilistic mapping relationship between the degraded image and reference image and effectively narrow down the difference of the feature distributions between the two. To further cope with the problem of fuzzy degradation and noise interference unique to underwater images, we constructed a multilevel spatial state feature extraction module. The module can adaptively expand the effective perceptual range and realize the synergistic modelling of global contextual information and local detailed features in the deep feature space through the cross-scale sensory field fusion mechanism. This dual-enhancement mechanism improves the model's ability to characterize complex degradation patterns underwater. It significantly improves the recovery of image edge sharpness and texture details through multi-dimensional feature reconstruction. The core innovations of this work can be summarized as follows:

-Constructing distribution estimation using variational self-encoder and adaptive normalization to solve the complex and variable problems of underwater scenes.

-Applying State-Space Model to underwater image enhancement as a balanced solution to CNN and Transformer structures.

-Experimental results strongly indicate that our method achieves promising performance compared to state-of-the-art methods on two UIE datasets and exhibits excellent generalization capabilities.

Journal of Physics: Conference Series **3108** (2025) 012031 doi:10.1088/1742-6596/3108/1/012031

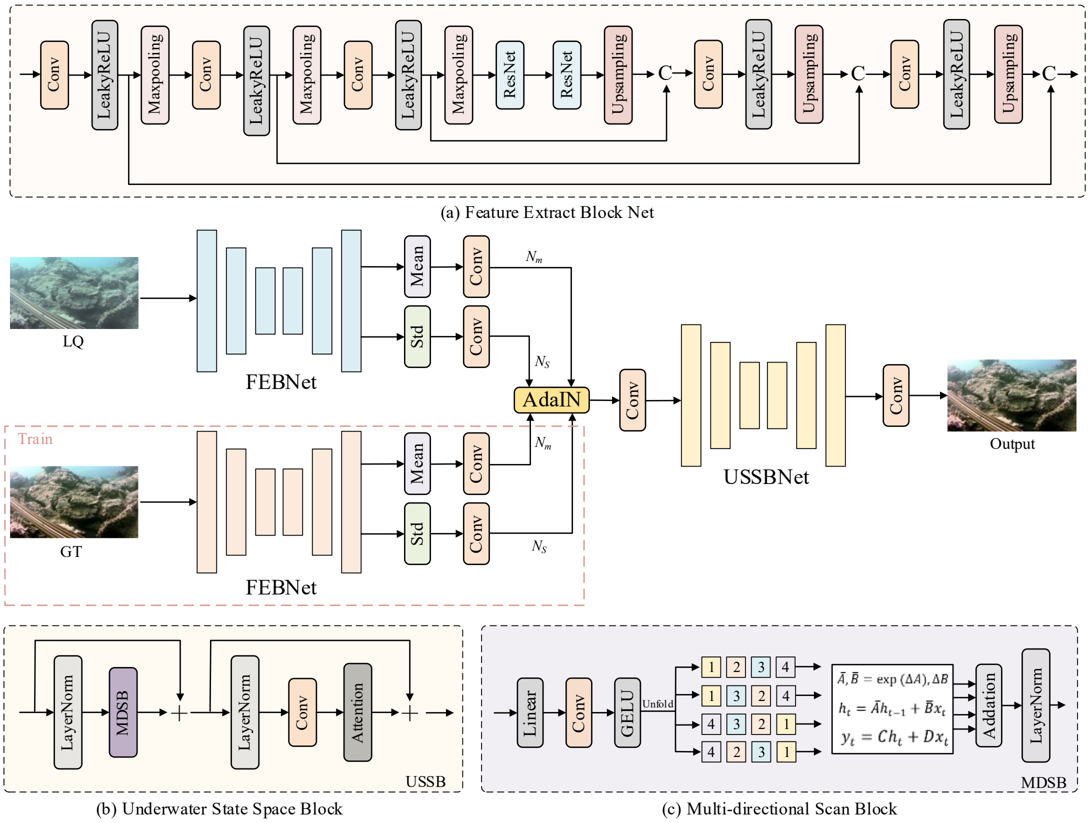

Figure 1. The network structure of Mamba-SAU.

2. Methods

2.1. *Feature Extract Block Net*

In this section, we elaborate the proposed scene-adaptive state-space network Mamba-SAU for underwater image enhancement. As shown in Figure 1. Our Mamba-SAU consists of two phases: domain-adaptive probabilistic modelling and high-quality image reconstruction. The domain-adaptive probabilistic modelling phase solves the uncertainty problem of reference map annotation through probabilistic modelling distribution estimation and consensus process. A variational self-encoder and adaptive instance normalization (AdaIN) are utilized to implement the probabilistic modelling process. A two-branch structure is used: a priori branch and a posteriori branch. A priori branch: input a single degraded image and estimate the prior distribution of the augmented global statistics (mean and standard deviation). Posterior branch: input its reference image and learn the posterior distribution by VAE. As shown in Figure 1.(a). Both branches use the U-Net structure to implement feature extraction, by embedding ResNet blocks to enhance feature representation. The mean and standard deviation of each channel of the extracted features are first computed, and then convolution is employed to obtain $\mu \in \mathbb{R}^{B\times N\times 1\times 1}$ and $\sigma \in \mathbb{R}^{B\times N\times 1\times 1}$, from the mean vector, and similarly, convolution is employed to obtain $m \in \mathbb{R}^{B\times N\times 1\times 1}$ and $v \in \mathbb{R}^{B\times N\times 1\times 1}$ from the standard deviation vector. The core module AdaIN implements the probabilistic transformation of the feature statistics by the following equation:

$$F = AdaIN(x) = a\left(\frac{x-\mu(x)}{\sigma(x)}\right) + b \tag{1}$$

Where $a \sim \mathcal{N}_s(m(y), v^2(y))$and $b \sim \mathcal{N}_m(\mu(y), \sigma^2(y))$ are sampled from the mean distribution ($\mathcal{N}_\mathrm{m}$) and the standard deviation distribution ($\mathcal{N}_\mathrm{s}$), respectively, where x stands for the *LQ* and y stands for the *GT*. This design extends $AdaIN$ into a probabilistic form to generate diversified augmentation results by random sampling. The above process can be formulated as:

$$x^l = LeakyReLU(Conv(x)) \tag{2}$$
$$x^{l+1} = LeakyReLU(Conv(MaxPool(x^l))) \tag{3}$$

2.2. Underwater State Space Block
The features after distribution probability estimation are processed in the high-quality image reconstruction stage. Referring to the flow of the previous classical architecture Norm → Attention → Norm → MLP, we consider the balanced relationship between maximum sensory field and computational complexity, and replace the Attention module with a structured state space sequence model, i.e., the Mamba architecture. It also adopts the U-Net structure, which consists of multiple Underwater State Space Block (USSB), and each group goes through Multi-directional Scan Block (MDSB) to capture the global receptive field. Since the state space sequence model is an operation done in one dimension, i.e., it only scans the transverse and longitudinal regions, but their diagonal positions are also similar in two dimensions, but cannot be scanned in the one-dimensional state space. Therefore, we add the convolution operation after the MDSB to compensate for the spatial neighbour pixel forgetting problem by using the properties of convolution. However, there is duplicity between these two operations, which will bring about the redundancy problem in the channel dimension. In order to improve the expressiveness of different channels, we introduce a channel attention mechanism in USSB to ensure that the state-space sequence model de-selectively learns the different channel features so as to avoid the channel redundancy problem. The above process can be formulated as:

$$Z = MDSB(LN(F^l)) + F^l \tag{4}$$
$$F^{l+1} = CA(Conv(LN(Z))) + Z \tag{5}$$

2.3. Multi-directional Scan Block
The Multi-directional Scan Block (MDSB) can utilize the features of the state-space sequence model to capture the global receptive field, and the structure of the MDSB is shown in Figure 1.(c). Firstly the feature channel is expanded to βC by linear layer, where β is a predefined expansion factor, and then after convolution and activation function operations, the features are processed by using the scanning method of the state-space sequence model, but since the original Mamba structure only scans the one-dimensional data, and it is not able to scan all the information well for the two-dimensional data such as the image, so we choose to flatten the 2D image features flattened into 1D features and scanned along different directions: top left to bottom right, bottom right to top left, top right to bottom left, and bottom left to top right. The global information scanned in different directions is merged using an addition operation. Finally, the channel is reprojected to C to generate an K_{out} with the same shape as the input. The above process can be expressed as:

$$K_1 = GELU(Conv(Linear(K)) \tag{6}$$
$$K_{out} = LN(Mamba(K_1)) \tag{7}$$

3. Experimental Results and Analysis
In our experiments, we evaluate the performance of the proposed method from two aspects. First, as an underwater image enhancement task, we perform qualitative and quantitative evaluations using relevant metrics, including visual reconstruction-related metrics (e.g., PSNR, SSIM) and underwater scene assessment metrics (e.g., UCIQE, UIQM). Finally, we performed ablation studies to analyse the role of different components.

3.1. Implementation Details and Datasets
The training and testing of the network model in this study are completed based on the NVIDIA GeForce RTX 4090 GPU platform, and a total of 500 rounds of iterative training are implemented. The experiments are carried out on two benchmark datasets, UIEB and UIQS, respectively: the UIEB dataset contains 890 fully labelled underwater images, divided into 800 training samples and 90 test samples according to the ratio of 8:1; the UIQS dataset is divided into five independent subsets according to the turbidity of the water body, and the experiments are carried out to evaluate the performance of the

turbidity levels respectively. Six mainstream algorithms, namely Water-Net, Ucolor, PUIE-Net, U-shape, Semi-UIR and HCLR, are selected as benchmark methods for the comparison experiments.

3.2. *Quantitative Evaluation*

Table 1. The comparison experiments on the UIEB dataset.

Method	**Pub**	**PSNR↑**	**SSIM↑**	**UIQM↑**	**UCIQE↑**
Water-Net[9]	TIP’19	16.305	0.797	2.875	0.606
Ucolor[10]	TIP’21	21.093	0.872	3.048	0.580
PUIE-Net[11]	ECCV’22	21.382	0.882	3.049	0.582
U-shape[4]	TIP’23	22.910	0.910	3.041	0.598
Semi-UIR[12]	CVPR’23	24.309	0.901	3.032	0.605
HCLR[13]	IJCV’24	**24.996**	0.925	3.033	0.607
Mamba-SAU		24.892	**0.931**	**3.052**	**0.609**

The quantitative assessment system in this study adopts a multi-dimensional metric design: for the UIEB benchmark test set, peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) are selected for the reference image-based quantitative comparison; the former measures the pixel-level reconstruction accuracy, and the latter assesses structural similarity, and the values of both are positively correlated with the image quality. For the UIQS dataset, which lacks reference images, the reference-free assessment metrics UCIQE (Underwater Colour Image Quality Evaluation) and UIQM (Underwater Image Quality Assessment) are used for quality metrics. The UIEB test set was supplemented with the above no-reference metrics analyses to build a comprehensive assessment framework. All metrics follow the evaluation criterion that the more significant value represents the better-perceived quality, and this strategy effectively guarantees the reliability of the assessment conclusions through multi-faceted validation.

As shown in Table 1, this paper's algorithm on the UIEB dataset demonstrates significant advantages in all four indicators; the UIQS cross-turbidity test results presented in Table 2 further confirm the method's universality. In terms of key quality indicators, the proposed algorithm comprehensively surpasses the existing optimal methods, and its comprehensive performance reaches the current state-of-the-art (SOTA) in the field of underwater image enhancement, which fully verifies the robustness and effectiveness of the Mamba-SAU architecture.

Table 2. The comparison experiments on the UIQS dataset.

Method	**Pub**	**UIQM↑**	**UCIQE↑**
Water-Net	TIP’19	2.098	0.590
Ucolor	TIP’21	3.079	0.611
PUIE-Net	ECCV’22	3.192	0.623
U-shape	TIP’23	2.953	0.617
Semi-UIR	CVPR’23	2.281	0.607
HCLR	IJCV’24	3.186	0.621
Mamba-SAU		**3.231**	**0.628**

3.3. *Qualitative Evaluation*

To qualitatively evaluate the reconstruction performance of the algorithm, we show some visualization results. Figure 2 shows a comparison of the qualitative results with different methods in the same scene. Our proposed Mamba-SAU method greatly improves the visibility and the color fidelity is significantly improved. Even under more turbid conditions, Mamba-SAU still achieves excellent restoration, whereas other methods do not have satisfactory scattering removal and color correction, and overexposure and distortion occur, especially under high turbidity conditions. Overall, our method greatly reduces the effect of scattering and presents superiority in underwater color correction, which is visually pleasing, whereas other methods suffer from bias and over-enhancement.

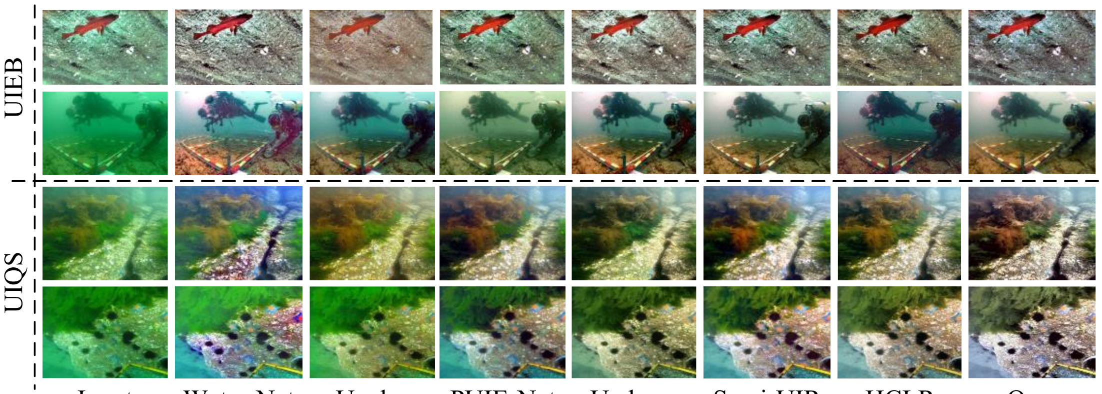

Figure 2. Visual comparisons on the UIEB and UIQS dataset.

3.4. *Ablation Studies*

Table 3. Ablation study with FEBN and USSB on the UIEB dataset.

#	Networks	PSNR↑	SSIM↑	UIQM↑	UCIQE↑
A	*w/o* AdaIN	22.077	0.873	2.988	0.598
B	*w/o* MDSB	21.133	0.886	2.996	0.599
C	*w/o* Attention in USSB	23. 432	0.896	3.043	0.601
D	*w/o* Conv in USSB	24.078	0.921	3.049	0.604
	Ours	24.892	**0.931**	**3.052**	**0.609**

Following the comparative experimental analysis, we conducted an ablation study to validate the effectiveness of the components in our Mamba-SAU network. As shown in Table 3, in this section, ablation studies are performed on probability distribution estimation with the help of AdaIN and capturing the global receptive field using the state space sequence model.

4. Conclusion

In this paper, we propose a scene-adaptive underwater image enhancement method using state-space modeling. The results are predicted by probabilistic network, which allows the network to extract to shallow features, and the semantic distribution of the reference image is obtained by distribution estimation, and the distribution distance between the low-quality image and the reference image is narrowed. Next, more information is extracted by expanding the sensory field through state space modeling as a way to enhance and reconstruct the image details. Both quantitative and qualitative experiments strongly demonstrate the effectiveness and superiority of our proposed method.

References

[1] McMahon, James, and Erion Plaku. "Autonomous data collection with timed communication constraints for unmanned underwater vehicles." *IEEE Robotics and Automation Letters* 6.2 (2021): 1832-1839.
[2] Kimball, Peter W., et al. "The ARTEMIS under-ice AUV docking system." *Journal of field robotics* 35.2 (2018): 299-308.
[3] Akkaynak, Derya, et al. "What is the space of attenuation coefficients in underwater computer vision?." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
[4] Peng, Lintao, Chunli Zhu, and Liheng Bian. "U-shape transformer for underwater image enhancement." *IEEE Transactions on Image Processing* 32 (2023): 3066-3079.
[5] Chiang, John Y., and Ying-Ching Chen. "Underwater image enhancement by wavelength compensation and dehazing." *IEEE transactions on image processing* 21.4 (2011): 1756-1769.
[6] Drews, Paulo LJ, et al. "Underwater depth estimation and image restoration based on single images." *IEEE computer graphics and applications* 36.2 (2016): 24-35.
[7] Li, Chong-Yi, et al. "Underwater image enhancement by dehazing with minimum information

loss and histogram distribution prior." *IEEE Transactions on Image Processing* 25.12 (2016): 5664-5677.

[8] Guo, Hang, et al. "Mambair: A simple baseline for image restoration with state-space model." *European conference on computer vision*. Cham: Springer Nature Switzerland, 2024.

[9] Li, Chong-Yi, et al. "An underwater image enhancement benchmark dataset and beyond." *IEEE transactions on image processing* 29 (2019): 4376-4389.

[10] Li, Chong-Yi, et al. "Underwater image enhancement via medium transmission-guided multi-color space embedding." *IEEE Transactions on Image Processing* 30 (2021): 4985-5000.

[11] Fu, Zhenqi, et al. "Uncertainty inspired underwater image enhancement." *European conference on computer vision*. Cham: Springer Nature Switzerland, 2022.

[12] Huang, Shirui, et al. "Contrastive semi-supervised learning for underwater image restoration via reliable bank." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2023.

[13] Zhou, Jingchun, et al. "HCLR-Net: hybrid contrastive learning regularization with locally randomized perturbation for underwater image enhancement." *International Journal of Computer Vision* 132.10 (2024): 4132-4156.