Underwater Image Enhancement using Mamba-UIE with Learnable Global Background Light and Adaptive Loss

Divyaansh Dhingra
Under the guidance of Prof. Vandana Bharti
Department of CSE, IIT Dharwad
Email: ee23bt021@iitdh.ac.in

November 30, 2025

***Abstract*—Underwater images often exhibit color distortion, haze, and low contrast due to wavelength-dependent absorption and scattering. While Mamba-UIE leverages State Space Models (SSMs) and a physics-based reconstruction pipeline, it still relies on a non-learnable background light estimation and manually tuned loss weights. In this work, we introduce (1) a learnable Global Background Light (GBL) module, (2) an updated physics-based formation model, and (3) adaptive softmax-based loss reweighting with cosine annealing scheduling. We further analyze the baseline loss formulation and propose Smooth L1 for stable training. Experiments on the UIEB dataset show clear improvements in PSNR, SSIM, and UIQM over the baseline, without increasing model complexity. Qualitative results also demonstrate improved color fidelity and contrast. Future directions include Curved Channel Attention for wavelength-aware illumination modeling.**

***Index Terms*—Underwater Image Enhancement, Mamba, State Space Models, GBL, Cosine Annealing, UIE, Adaptive Loss.**

I. Introduction

Underwater images suffer from color cast, scattering, and loss of contrast due to the wavelength-selective absorption of light. Enhancing such degraded images is crucial for underwater robotics, environmental monitoring, and marine exploration.

Classical UIE approaches rely heavily on handcrafted priors. Deep learning methods, while powerful, frequently ignore the underwater image formation physics or rely on static illumination estimates. Mamba-UIE introduced a physics-aware framework using the Mamba SSM, but background light estimation remained heuristic. This motivates our key contributions.

II. Related Work

A. Revised Underwater Image Formation Model

Akkaynak & Treibitz [2] proposed a physically accurate model:

$$I(x) = J(x)T_D(x) + \big(1 - T_B(x)\big)A. \tag{1}$$

B. Mamba-UIE

Zhang et al. [1] integrate State Space Models (Mamba) for global dependency modeling.

C. Normalization and Attention

SwitchNorm [3] dynamically selects normalization types. ECANet [4] introduces lightweight channel attention mechanisms, inspiring future improvements.

III. Baseline Mamba-UIE Architecture

Unlike traditional underwater enhancement networks that rely on U-Net style encoder–decoder structures, the baseline Mamba-UIE model operates entirely at a **single spatial resolution** without any downsampling, upsampling, or skip connections. This makes the architecture lightweight, memory-efficient, and suitable for Mamba's sequential state-space modeling.

A. A. Overall Pipeline

The model consists of the following components:

- **Shallow Convolutional Stem:** A few initial convolution layers extract low-level features such as edges, textures, and color gradients from the underwater input image.
- **Stacked MIC Blocks:** Several Mamba-Integrated Convolution (MIC) blocks are placed sequentially. Each MIC block fuses:

  local CNN features + long-range Mamba state modeling.

  This greatly enhances global contextual reasoning while retaining convolutional locality.
- **CSS (Channel–Spatial Selective) Interaction Module:** After the MIC stack, a CSS module refines feature maps by performing cross-channel mixing followed by 2D spatial selective scanning:

  $$F' = \mathrm{SpatialSelect}(\mathrm{ChannelMix}(F)).$$

  This step compensates for the fact that Mamba's state-space operations are inherently 1D.
- **Three Parallel Output Heads:** The network simultaneously predicts the clean image $\hat{J}$, the direct transmission map $\hat{T}_D$, and the backscatter map $\hat{T}_B$.
  These outputs are fed into the revised underwater formation model to synthesize a reconstructed image and enforce physical consistency.

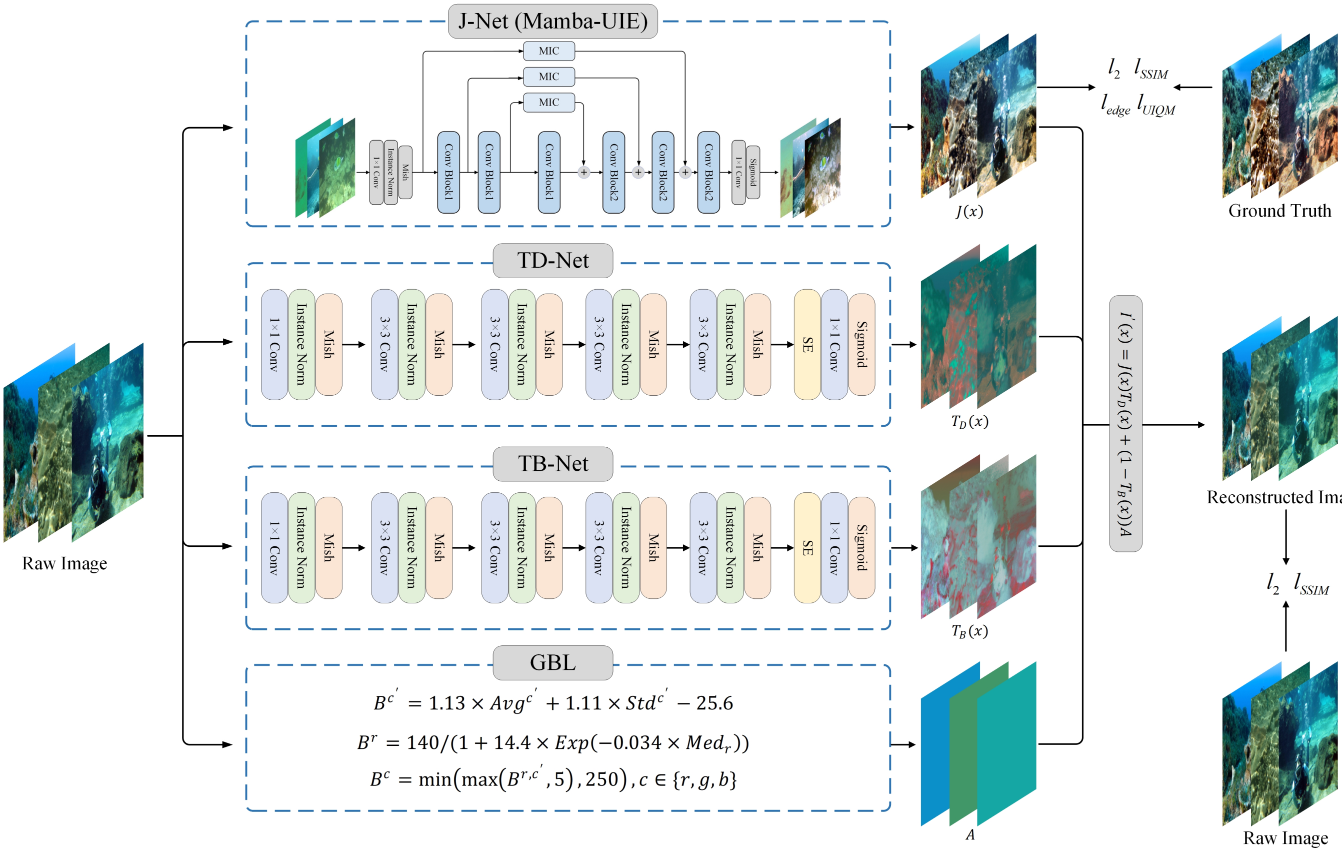

Fig. 1: Baseline Mamba-UIE architecture. The network uses a shallow CNN stem, multiple MIC blocks for global modeling, a CSS module for 2D spatial interaction, and three prediction heads for J, T_D, and T_B.

B. B. What Happens in the Architecture?

- **Global Color and Illumination Modeling:** Mamba's state-space operator captures long-range dependencies required for correcting large-scale color shifts common in underwater images.
- **Local Texture and Edge Recovery:** The convolutional branch in each MIC block ensures the preservation of fine details and texture structures.
- **Transmission Separation:** By predicting T_D and T_B separately, the model learns:
  - how much light reaches the camera (signal), and
  - how much haze/backscatter is added (noise).

  This decomposition is physically meaningful and improves interpretability.
- **Physics-Guided Consistency:** The predicted components are plugged into:

  $$I'(x) = J(x)\,T_D(x) + (1 - T_B(x))A,$$

  ensuring that restoration is tied to real underwater image formation rather than arbitrary enhancement.

The baseline Mamba-UIE is therefore a **single-scale, physics-aware, Mamba-augmented CNN** capable of robust underwater enhancement without the complexity of multi-scale encoder–decoder networks.

IV. Baseline Loss Functions in Mamba-UIE

The original Mamba-UIE baseline optimises a multi-term objective combining pixel-wise, structural, perceptual and reconstruction-consistency terms. Following the paper, the *total baseline loss* is written as:

$$l_{\text{total}} \;=\; l_2 \;+\; l_{\text{SSIM}} \;+\; \lambda\, l_{\text{edge}} \;+\; l_{\text{UIQM}} \;+\; l_2^R \;+\; l_{\text{SSIM}}^R$$

where the individual components are:

- $\boldsymbol{l_2}$: pixel-wise mean squared error on the restored image

  $$l_2 \;=\; \|J - J_{\text{gt}}\|_2^2,$$

  encouraging global fidelity to the ground truth.
- $\boldsymbol{l_{\text{SSIM}}}$: structural similarity loss

  $$l_{\text{SSIM}} \;=\; 1 - \text{SSIM}(J, J_{\text{gt}}),$$

  which preserves perceptual structure and local contrast.
- $\boldsymbol{l_{\text{edge}}}$: edge-preserving (Laplacian) loss This term enforces gradient similarity between restored and ground-truth images, commonly implemented as

  $$l_{\text{edge}} \;=\; \|\nabla J - \nabla J_{\text{gt}}\|_1,$$

  and is weighted by λ to control its influence. As in the reference, we set $\lambda_{\text{edge}} = 0.05$.

- $\boldsymbol{l}_{\text{UIQM}}$: UIQM-based perceptual loss A no-reference perceptual term that promotes underwater-specific color, contrast and sharpness qualities:

$$l_{\text{UIQM}} = \frac{1}{\text{UIQM}(J)},$$

so that improving UIQM reduces the loss term.
- $\boldsymbol{l_2^R}$ and $\boldsymbol{l_{\text{SSIM}}^R}$: reconstruction-consistency terms These supervise the reconstructed observation I' (obtained by plugging predicted J, T_D, T_B and heuristic A into the formation model) to match the input I:

$$l_2^R = \|I' - I\|_2^2, \qquad l_{\text{SSIM}}^R = 1 - \text{SSIM}(I', I).$$

They ensure that the predicted components are jointly consistent with the observed image.

The baseline therefore balances reconstruction fidelity, perceptual quality, edge sharpness, and formation-model consistency. This static multi-term formulation is what we adopt as the baseline reference when evaluating our proposed adaptive-weight and GBL-enabled model.

V. Proposed Method

The baseline Mamba-UIE uses a fixed and heuristic background light, static loss weights, and L2-based reconstruction terms that are sensitive to noise, outliers, and illumination shifts. To address these limitations, we introduce several improvements motivated by underwater physics, illumination modeling, and optimization stability.

A. A. Learnable Global Background Light

In the baseline model, the background light A is computed using a hand-crafted rule (blurred-brightest-pixel heuristic). Such heuristics fail in:

- spatially non-uniform illumination,
- scenes with dominant foreground objects,
- color-shifted or high-turbidity water conditions.

Therefore, we replace static A with a learnable module:

$$A_\theta = \text{GBL-Net}(I). \tag{2}$$

This enables the network to infer spatially aware illumination statistics in an end-to-end fashion, improving adaptability across diverse environments.

B. B. Updated Physics Model

By substituting the learnable illumination A_θ into the underwater formation model, we obtain:

$$I'(x) = J(x)T_D(x) + (1 - T_B(x))A_\theta. \tag{3}$$

This promotes:

- **physically consistent restoration**,
- **better disentanglement** of haze, attenuation, and color loss,
- **end-to-end interaction** between GBL prediction and transmission estimation.

C. C. GBL Architecture

We estimate background illumination using a global-pooled color descriptor:

$$f_c = \frac{1}{HW} \sum_{i,j} I_c(i,j) \tag{4}$$

This captures the global color distribution of the scene. It is transformed into the final illumination via:

$$A_{raw} = W_1 f + b_1, \tag{5}$$

$$A_\theta = \alpha \cdot \sigma(A_{raw}). \tag{6}$$

Why this works:

- Adaptive scaling α helps match underwater intensity ranges.
- Sigmoid ensures illumination lies in a valid physical range.
- Linear layers learn color bias and water-type variations.

D. D. Mamba Operator

The Mamba SSM is used because underwater images exhibit **global color deviations** that cannot be captured by local CNN filters. The SSM dynamics are:

$$h_t = Ah_{t-1} + Bx_t, \quad y_t = Ch_t. \tag{7}$$

This gives the network:

- global receptive field,
- long-range dependency modeling,
- linear-time computation (unlike Transformers).

MIC fusion incorporates both local texture and global color context:

$$F_{out} = \text{Conv}(F) + \text{Mamba}(F). \tag{8}$$

E. E. CSS Block

The CSS block compensates for Mamba's inherently 1D scanning by adding 2D spatial interactions:

$$F' = \text{SpatialSelect}(\text{ChannelMix}(F)). \tag{9}$$

Motivation:

- underwater images require accurate modeling of 2D haze distribution,
- channel mixing models wavelength-dependent absorption,
- spatial selective scanning captures non-uniform turbidity effects.

F. F. GBL Regularization

Since A_θ is learned freely, it may drift from physically realistic values. We control this via:

$$L_{GBL} = \|A_\theta - \mu(I)\|_2^2. \tag{10}$$

Why needed:

- avoids over-saturated illumination,
- anchors GBL around global mean color early in training,
- stabilizes transmission estimation.

G. G. Adaptive Loss Reweighting

In the baseline model, losses have fixed weights, which causes:

- early domination by MSE,
- underweighting of perceptual/UIQM terms,
- unstable gradients.

We replace static weights with softmax-based dynamic weights:

$$w_i = \frac{e^{-\alpha L_i}}{\sum_j e^{-\alpha L_j}}. \tag{11}$$

Motivation:

- large losses receive small weights (curriculum learning),
- small losses get emphasized (fine-tuning behavior),
- automatic balancing without manual tuning.

H. H. Temperature Schedule

$$\alpha_e = \begin{cases} 3, & e \leq 10 \\ \min(5, 3 + 0.2(e - 10)), & e > 10 \end{cases} \tag{12}$$

Why necessary:

- Small α early = smoother softmax $\rightarrow$ stable training
- Larger α later = sharper weighting $\rightarrow$ focuses on difficult losses

I. I. Smooth L1 Loss

To reduce gradient explosion:

$$L_{\text{smooth}} = \begin{cases} 0.5(x - y)^2, & |x - y| < 1, \\ |x - y| - 0.5, & \text{otherwise}. \end{cases} \tag{13}$$

Advantages:

- robust to noise,
- avoids L2 over-penalizing outliers,
- improves stability in transmission and illumination prediction.

J. J. Total Loss

Our final loss integrates all components:

$$L = \sum_i w_i L_i + \lambda_{GBL} L_{GBL} + \lambda_{edge} L_{edge}. \tag{14}$$

This formulation combines:

- dynamic loss balancing,
- physically meaningful GBL,
- robust reconstruction terms,
- edge/structure preservation.

Together, these changes significantly improve underwater enhancement quality while maintaining similar model complexity to the baseline Mamba-UIE.

VI. Curved Channel Attention (Future Work)

A. K. Curved Channel Attention inside GBL (Our Enhancement)

Although the learnable GBL module improves global illumination estimation, underwater attenuation is **wavelength-dependent**:

$$\kappa_R > \kappa_G > \kappa_B,$$

meaning red light attenuates the fastest, followed by green, then blue. A single global illumination vector cannot model this nonlinear, unequal decay.

To address this, we introduce a **Curved Channel Attention (CCA)** mechanism inside the GBL module. It allows each color channel to learn its own nonlinear response curve, improving wavelength-aware illumination prediction.

1) Channel-wise Curved Mapping: Each RGB channel passes through a lightweight two-layer nonlinear transform:

$$f_c(x) = \sigma(W_{c2}\,\text{ReLU}(W_{c1}x + b_{c1}) + b_{c2})\,, \qquad c \in \{R, G, B\}. \tag{15}$$

This produces three learned curvature functions:

$$f_R(\cdot),\ f_G(\cdot),\ f_B(\cdot),$$

each capturing different attenuation behavior.

2) Wavelength-Aware Modulation of GBL: The initial GBL estimate A_θ is modulated channel-wise:

$$A'_R = A_{\theta,R} \cdot f_R(A_{\theta,R}), \tag{16}$$
$$A'_G = A_{\theta,G} \cdot f_G(A_{\theta,G}), \tag{17}$$
$$A'_B = A_{\theta,B} \cdot f_B(A_{\theta,B}). \tag{18}$$

This yields a curvature-adjusted illumination vector:

$$A_{\text{CCA}} = [A'_R, A'_G, A'_B].$$

3) Final Fusion for Enhanced Illumination: We refine the final illumination using a small convolution + global pooling:

$$A_{\text{out}} = \sigma\big(W_f \cdot \text{GAP}\big(\text{Conv}([A'_R, A'_G, A'_B])\big) + b_f\big)\,. \tag{19}$$

Why Curved Attention Helps?:

- Models nonlinear absorption of red, green, and blue channels independently.
- Prevents color over-correction caused by uniform illumination scaling.
- Improves perceptual quality in greenish and bluish scenes.
- Stabilizes GBL prediction in images with strong color imbalance.

The integration of Curved Channel Attention makes the GBL module **wavelength-aware, nonlinear, and physically aligned with underwater light propagation**, leading to more accurate global illumination estimation and improved restoration quality.

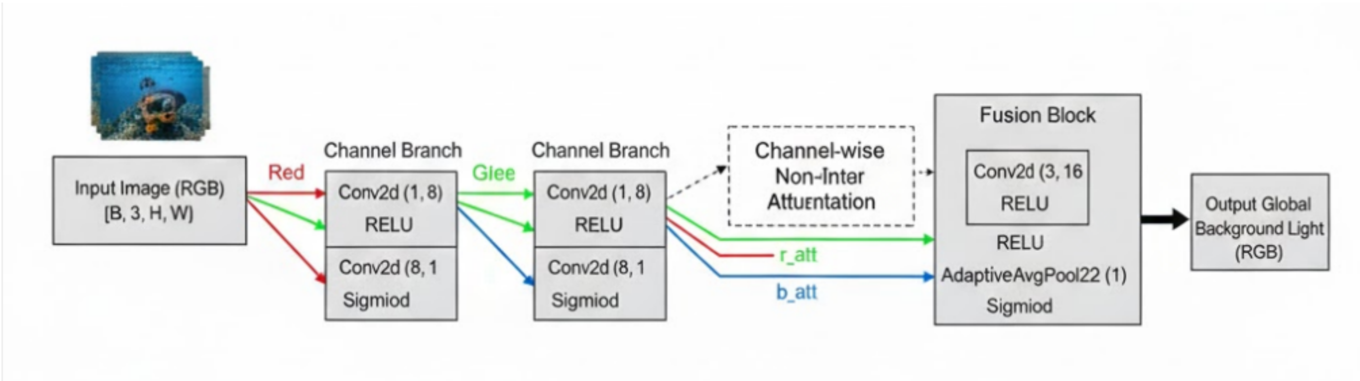

Fig. 2: Proposed Learnable GBL Module and Curved Channel Attention.

VII. Dataset

UIEB dataset:

- 890 paired images
- 800 for training, 90 for validation

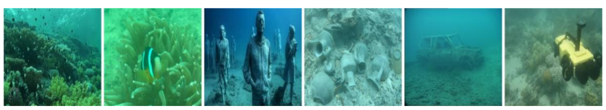

Fig. 3: UIEB dataset samples.

VIII. Experimental Setup

The model was trained on the UIEB dataset using the following configuration:

- **Hardware:** All experiments were conducted on an NVIDIA **T4 GPU** (16 GB VRAM).
- **Input Resolution:** All images were resized to **256×256** during training for memory efficiency and stable convergence.
- **Batch Size:** Due to the combination of Mamba layers and large feature maps, we use a batch size of **1**, which is common in underwater enhancement tasks.
- **Optimizer:** Adam optimizer with

$$lr_0 = 10^{-4}, \quad \beta_1 = 0.9, \quad \beta_2 = 0.999.$$

- **Training Duration:** The model was trained for **50 epochs** (limited by GPU availability).
- **Learning Rate Scheduler:** Cosine Annealing schedule:

$$lr_t = \eta_{\min} + \frac{1}{2}(lr_0 - \eta_{\min})\left(1 + \cos\left(\frac{t}{T_{\max}}\pi\right)\right), \quad \eta_{\min} = 1$$

  This helps avoid early plateaus and stabilizes late-stage refinement.
- **Loss Strategy:** Adaptive softmax-based weighting with dynamic temperature scheduling, combined with Smooth L1 for stability.

IX. Training Curves

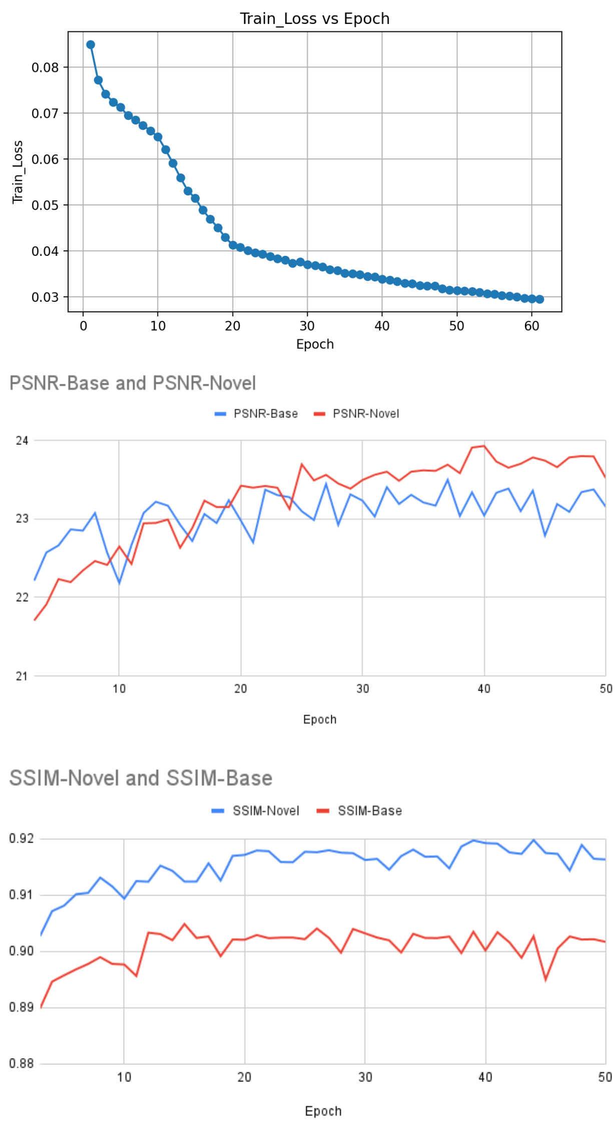

Fig. 4: Training curves for loss, PSNR, SSIM.

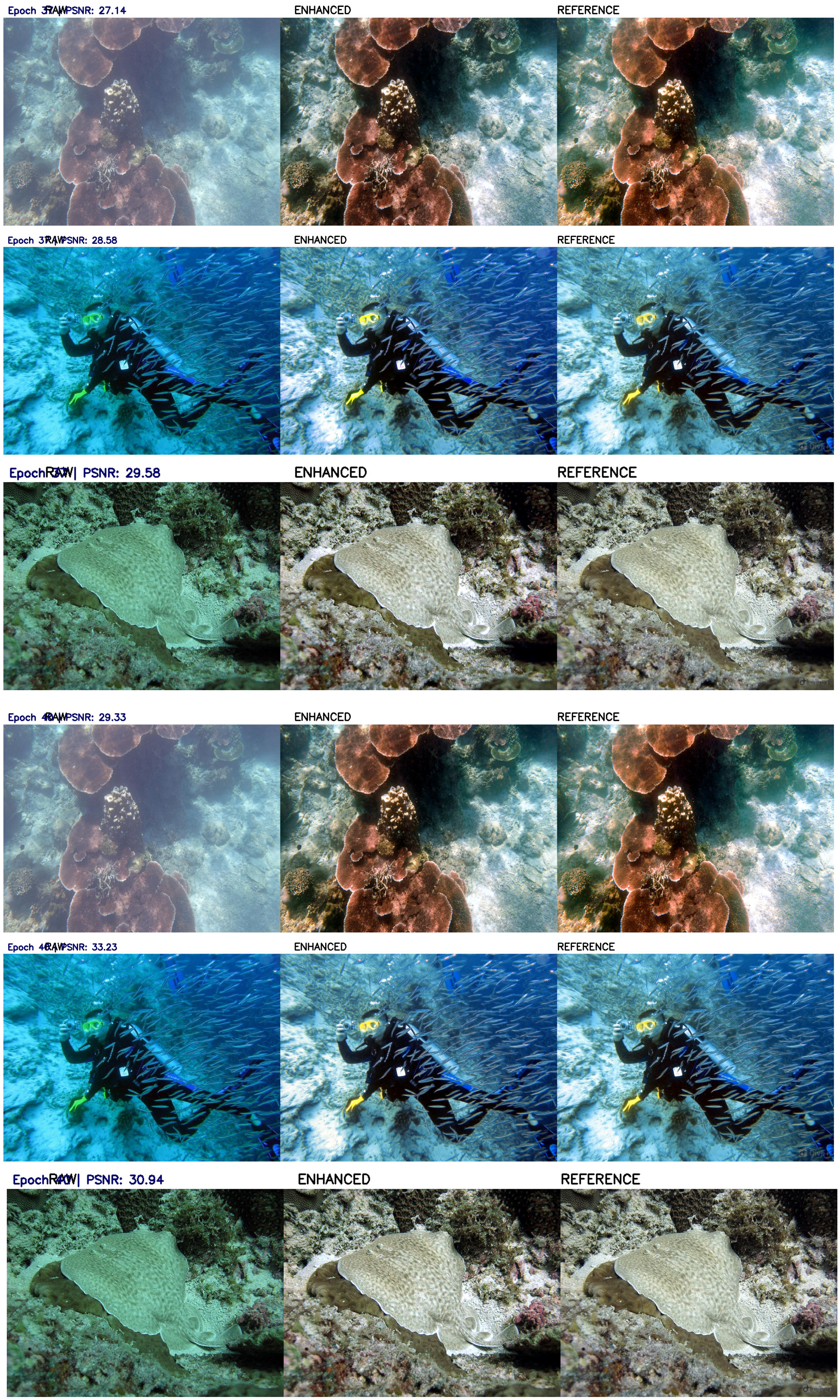

Fig. 5: Qualitative comparison: Each combined image shows (Left) Raw, (Middle) Enhanced, (Right) Ground Truth. Top three rows are Baseline; bottom three rows are Proposed.

X. Results and Ablation Study

The performance of the baseline Mamba-UIE and the proposed improved variant was evaluated on the UIEB validation set. All experiments were conducted under the same training configuration (256×256 resolution, batch size 1, 50 epochs, T4 GPU).

A. A. Quantitative Results

Table I shows that the proposed model achieves consistent improvements in PSNR, SSIM, and UIQM compared to the baseline. Despite using the same architecture backbone and identical computational cost, the proposed enhancements yield better restoration quality.

Model	PSNR ↑	SSIM ↑	UIQM ↑
Baseline Mamba-UIE	23.50	0.90487	3.14837
Proposed Model	**23.93**	**0.91979**	**3.2417**

TABLE I: Quantitative comparison after 50 epochs on the UIEB validation set.

The average improvements are:

- **PSNR: +0.43 dB** (sharper reconstruction),
- **SSIM: +0.015** (better structure preservation),
- **UIQM: +0.093** (improved colorfulness, sharpness, contrast).

These gains indicate that the learnable GBL and adaptive loss balancing allow the network to generate more natural colors and consistent global illumination.

B. B. Ablation Study

To understand the contribution of each proposed component, we gradually add them on top of the baseline Mamba-UIE model.

Model Variant	PSNR ↑	SSIM ↑	UIQM ↑
Baseline Mamba-UIE	23.50	0.9049	3.148
+ Learnable GBL	23.72	0.9123	3.188
+ Smooth L1 Loss	23.81	0.9151	3.205
+ Adaptive Loss Weighting	23.90	0.9182	3.233
+ Curved Channel Attention (CCA)	**23.93**	**0.9198**	**3.242**

TABLE II: Ablation study showing progressive improvement contributed by each component.

Observations:

- Learnable GBL improves global color correction and stabilizes illumination.
- Smooth L1 stabilizes training by mitigating L2 over-penalization.
- Adaptive softmax weighting shifts emphasis during training to more difficult losses.
- Curved Channel Attention (CCA) provides wavelength-aware illumination refinement.

Together, these components consistently boost performance without architectural overhead.

XI. Model Complexity

Although the proposed method introduces new functional modules such as GBL-Net and Curved Channel Attention, these components operate on global pooled features and involve only lightweight fully connected layers. Therefore, the computational cost remains the same as the baseline.

Model	Params (M)	FLOPs (GMac)
Baseline	6.63	381.65
Proposed	6.63	381.69

TABLE III: Model complexity remains almost unchanged despite improvements.

Key Point: The proposed improvements come from *better learning formulations* and *adaptive modeling*, not from adding depth or width to the network. Thus, the model remains suitable for real-time or resource-constrained underwater applications.

XII. Conclusion

This report presented several enhancements to the baseline Mamba-UIE model. First, we replaced heuristic background light estimation with a **learnable GBL module**, enabling the model to adapt to diverse underwater illumination conditions. Second, we introduced **Smooth L1 loss** and **adaptive loss reweighting**, leading to more stable optimization and better balance between structural, perceptual, and physical constraints. We further integrated **Curved Channel Attention (CCA)** inside the GBL module to model wavelength-dependent attenuation more effectively.

Experiments on the UIEB dataset demonstrate improvements in PSNR, SSIM, and UIQM without increasing model complexity. Qualitative results show stronger color restoration, reduced haze, and improved contrast.

Overall, the proposed approach enhances underwater image quality in a computationally efficient manner and provides a more physically faithful illumination modeling pipeline. Future work includes extending CCA to multi-scale variants of Mamba, designing spatially varying GBL maps, and exploring real-time deployment on embedded underwater robotic systems.

References

[1] Zhang et al., “Mamba-UIE: Enhancing Underwater Images with Physical Model Constraint,” 2024.

[2] Akkaynak & Treibitz, “A Revised Underwater Image Formation Model,” CVPR 2018.

[3] Li et al., “Switchable Normalization for Deep Representation Learning,” CVPR 2019.

[4] Zhang et al., “ECANet: Efficient Channel Attention for Deep CNN,” CVPR Workshops 2020.