

# AI-Based Underwater Image Enhancement System

Trupti Belure, Dr. Anita Pisote, Shivam Bawage, Sayukta Bhagwat

Department of Computer Applications

JSPM University, Wagholi, Pune

ayuktabhagwat23.ca@jspmuni.ac.in, truptibelure23.ca@jspmuni.ac.in

shivambawage23.ca@jspmuni.ac.in

**Abstract:** Underwater images play a crucial role in marine exploration, environmental monitoring, archaeological studies, and autonomous underwater robotics. However, underwater images often suffer from severe quality degradation due to light absorption, scattering, colour distortion, low contrast, and haze caused by suspended water particles. These factors significantly reduce the visibility and accuracy of visual information captured in underwater environments. To address these challenges, this research proposes an AI-Based Underwater Image Enhancement System that utilizes deep learning techniques to automatically restore image clarity, improve colour balance, and enhance contrast. The proposed system leverages Convolutional Neural Networks (CNNs) and advanced image restoration algorithms to learn complex underwater degradation patterns and reconstruct high-quality images from low-visibility inputs. A dataset of paired degraded and ground-truth underwater images is used to train the model, enabling it to correct colour casts, remove haze, and recover fine details while preserving natural textures. The model performance is evaluated using standard image quality metrics such as Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Underwater Image Quality Measure (UIQM). Experimental results demonstrate that the proposed CNN-based method achieves a PSNR of 31.6 dB, SSIM of 0.91, and UIQM of 3.54, significantly outperforming traditional image enhancement approaches on standard underwater benchmark datasets.

**Keywords:** Underwater Image Enhancement, Deep learning, Convolutional Neural Networks, Image Restoration, Colour Correction, Dehazing, PSNR, SSIM, UIQM, Encoder-Decoder, Marine Imaging

## I. INTRODUCTION

Underwater imaging has become an essential tool across multiple scientific and industrial disciplines including marine biology, oceanographic research, underwater archaeology, environmental monitoring and autonomous underwater vehicle (AUV) navigation. High-quality underwater images are indispensable for accurately observing marine ecosystems, detecting submerged objects, and supporting autonomous robotic vision systems. However, capturing clear and visually faithful images in underwater environments presents unique and persistent challenges that distinguish underwater photography fundamentally from terrestrial image capture.

The primary source of underwater image degradation is the selective absorption and scattering of light by water molecules and suspended particulates. As light propagates through water, different wavelengths are attenuated at different rates: red wavelengths (620–750 nm) are absorbed most rapidly within the first few meters of depth, while shorter blue and green wavelengths penetrate significantly deeper. This wavelength-selective absorption produces the characteristic bluish or greenish colour cast universally observed in unprocessed underwater images.

Simultaneously, particulate matter suspended in the water column—sediment, plankton, organic detritus, and air bubbles—acts as a scattering medium that redirects light and produces haze-like effects reducing contrast, obscuring fine detail, and blurring object boundaries. The combined effects of absorption and scattering create underwater images typically characterized by low contrast, reduced sharpness, significant colour distortion, and spatially varying haze that intensifies with the distance between camera and subject.



Traditional underwater image enhancement approaches including histogram equalization, white balance correction, gamma correction, and spatial frequency filtering represent decades of effort to mitigate degradation through principled but inherently limited signal processing. These methods treat each image independently and rely on hand-crafted models that fail to capture the full complexity of underwater degradation, which varies with water depth, turbidity, ambient illumination, and the spectral composition of artificial lighting.

The emergence of deep learning and convolutional neural networks has opened fundamentally new possibilities for image restoration. Unlike traditional methods, CNNs can learn the complex and nonlinear mapping from degraded inputs to enhanced outputs directly from paired training data without requiring explicit physical models of the degradation process. This research presents the design, implementation, and evaluation of an AI-Based Underwater Image Enhancement System employing a CNN-based encoder-decoder architecture trained on large paired underwater image datasets.

### **A. Motivation**

The growing deployment of underwater imaging systems in conservation monitoring, offshore inspection, marine archaeology, and autonomous robotics creates an urgent need for automated high-quality enhancement that does not require manual parameter tuning per image. The inability of existing methods to produce consistently high-quality results across diverse underwater conditions represents a critical bottleneck that automated deep learning approaches can overcome.

### **B. Contributions**

The primary contributions of this paper are:

- A complete AI-based underwater image enhancement pipeline using a CNN encoder-decoder architecture trained end-to-end on paired underwater image data comprising over 14,000 image pairs.
- A comprehensive quantitative evaluation against four traditional enhancement baselines using PSNR, SSIM, and UIQM metrics on standard benchmark datasets.
- A practical web-based deployment interface enabling real-time interactive enhancement through an upload- and-process workflow without requiring user expertise in image processing.
- Detailed analysis of failure cases and architectural limitations providing direction for future research.

## **II. RELATED WORK**

Research on underwater image enhancement spans three broad methodological paradigms: physics-based image formation model approaches, traditional image processing techniques, and data-driven deep learning methods.

### **A. Physics-Based Methods**

Physics-based approaches model the underwater image formation process explicitly using the Jaffe-McGlamery model, which expresses the observed image as  $I(x) = J(x) \cdot t(x) + B \cdot (1 - t(x))$ , where  $J(x)$  is the scene radiance,  $t(x)$  is the transmission map, and  $B$  is the global background light from backscattering. He et al. adapted the Dark Channel Prior (DCP), originally developed for atmospheric haze removal, to underwater dehazing. However, physics-based methods require accurate estimation of water-type-specific parameters that are difficult to obtain in practice, and performance degrades significantly when underlying physical assumptions are violated by actual water conditions.

### **B. Traditional Image Processing**

Histogram equalization and its adaptive variant (CLAHE) redistribute intensity values to improve perceived contrast, providing rapid improvement but often producing oversaturated or unnatural-looking results. White balance correction addresses colour cast but provides no improvement in contrast or dehazing. Retime - based methods decompose images into illumination and reflectance components, providing the best traditional results but requiring careful per-image parameter tuning and falling substantially below CNN performance across all metrics.



### **C. Deep Learning Approaches**

Li et al. (2019) introduced the UIEB benchmark dataset of 950 paired underwater images that has become the standard evaluation benchmark. Ranneberger et al. (2015) proposed the U-Net encoder-decoder architecture with skip connections, which has since been widely adopted for underwater enhancement. Liu et al. (2019) demonstrated that a deep residual framework for underwater enhancement achieves PSNR improvements of 3–5 dB over traditional methods. Zhang et al. (2020) combined physics-based image formation models with deep learning, achieving state-of-the-art performance. Islam et al. (2020) proposed FUIE- GAN for fast underwater image enhancement using generative adversarial training, enabling near-real-time performance. The proposed system builds on these foundations with a complete, practically deployable end-to-end system.

## **III. METHODOLOGY**

The proposed AI-Based Underwater Image Enhancement System employs a structured multi-stage methodology encompassing dataset preparation, preprocessing, model architecture design, training procedure, and evaluation protocol.

### **A. Dataset Collection**

The training dataset is assembled from three publicly available underwater image benchmarks. The primary source is the UIEB dataset (Li et al., 2019), providing 890 paired training images of raw degraded photographs and human-selected reference enhanced versions. Additional data is sourced from the EUVP dataset (Islam et al., 2020), which provides over 12,000 paired images from oceanic explorations and scientific dives, and the UFO-120 dataset providing 1,500 paired images of underwater objects with known ground truth. The combined training dataset of approximately 14,000 paired samples provides sufficient diversity of degradation types, water conditions, and subject matter for effective generalization. The dataset is partitioned into training (80%), validation (10%), and testing (10%) splits using stratified random sampling.

### **B. Data Preprocessing**

All images are resized to  $256 \times 256$  pixels using bicubic interpolation and pixel values are normalized to the  $[0, 1]$  range. Data augmentation applied exclusively to the training set includes:

- Random horizontal flipping ( $p = 0.5$ ) for orientation invariance
- Random rotation within  $\pm 15$  degrees for camera tilt robustness
- Random crops of  $224 \times 224$  to introduce translational variation
- Random brightness and contrast perturbation within  $\pm 10\%$
- Random Gaussian noise addition ( $\sigma \in [0, 0.02]$ ) for sensor noise robustness

### **C. Model Architecture**

The proposed model employs a U-Net-inspired encoder-decoder architecture with skip connections. The Encoder Path consists of four convolutional blocks, each comprising two  $3 \times 3$  convolutional layers with Batch Normalization and ReLU activation, followed by  $2 \times 2$  max-pooling. Channel counts double at each stage:  $64 \rightarrow 128 \rightarrow 256 \rightarrow 512$ . The Bottleneck layer at  $16 \times 16$  spatial resolution contains 1,024 channels capturing global colour context. The Decoder Path mirrors the encoder with four up sampling blocks performing  $2 \times$  bilinear up sampling, concatenation with encoder skip connections, and two  $3 \times 3$  convolutional layers. Skip connections preserve fine spatial detail lost through bottleneck compression. The Output Layer is a  $1 \times 1$  convolution with Sigmoid activation mapping to the 3-channel enhanced image. The architecture contains approximately 7.8 million trainable parameters.

### **D. Loss Function**

The model is trained using a composite loss function:



Total =  $0.5 \cdot L_{MSE} + 0.3 \cdot L_{perceptual} + 0.2 \cdot L_{SSIM}$  The MSE loss provides pixel-level reconstruction supervision. The perceptual loss computes feature-space distances using pre-trained VGG-16 (relu2\_2 and relu3\_3 layers) to encourage perceptually natural textures. The SSIM loss ( $1 - SSIM$ ) directly optimizes structural similarity. Loss weights were determined through ablation experiments on the validation set.

### **E. Training Configuration**

The model is trained for 100 epochs with batch size 16 using the Adam optimizer ( $\text{lr} = 1 \times 10^{-4}$ ,  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ ). A cosine annealing learning rate schedule with warm restarts ( $T_0 = 10$  epochs,  $T_{\text{mult}} = 2$ ) is applied. Early stopping (patience = 15) monitors validation PSNR, and the best checkpoint is retained. Training uses PyTorch 2.0 with automatic mixed precision on an NVIDIA Tesla T4 GPU (16 GB VRAM), requiring approximately 6.5 hours.

## **IV. SYSTEM WORKING**

The end-to-end working of the AI-Based Underwater Image Enhancement System is structured as a sequential six-stage pipeline that transforms raw degraded underwater input images into high-quality enhanced outputs.

### **1. Input Image Acquisition**

Input images may originate from diverse sources including underwater cameras on ROVs or AUVs, scientific diving operations, fixed seabed monitoring cameras, or publicly available datasets. The system accepts images in JPEG, PNG, BMP, and TIFF formats through a web-based upload interface with image preview functionality.

### **2. Image Preprocessing**

The preprocessing module resizes input images to  $256 \times 256$  pixels using bicubic interpolation, normalizes pixel values to  $[0, 1]$ , and formats the image tensor as a  $1 \times 3 \times 256 \times 256$  batch tensor for model input. These steps are identical to training preprocessing, ensuring consistent training- inference distribution.

### **3. Feature Extraction**

The preprocessed tensor passes through the encoder pathway. The four successive convolutional blocks progressively extract increasingly abstract representations: the first block captures edges, color gradients, and fine textures; deeper blocks capture structural patterns and global color statistics. Feature maps at each resolution level are stored for skip connection use.

### **4. Image Enhancement Using AI Model**

The bottleneck feature representation is decoded by the up sampling pathway. At each resolution, decoder features are concatenated with encoder skip connection features before convolutional processing, combining global context with preserved local detail. The AI model performs:

- **Colour Correction:** Redistributes channel intensities to restore warm tones attenuated by wavelength- selective absorption, reducing the artificial blue-green cast.
- **Contrast Enhancement:** Improves local and global contrast through learned feature transformations, increasing perceptual distinction between boundaries and background.
- **Dehazing:** Suppresses scattering-induced haze through learned inversion of the backscattering component, restoring visibility at depth.
- **Detail Recovery:** Recovers fine structural details including coral textures, fish scale patterns, and seafloor features blurred by scattering.

### **5. Output Generation**

The final  $1 \times 1$  convolutional layer with Sigmoid activation maps the decoder output to a 3-channel enhanced image in  $[0, 1]$ . The output is converted to uint8, saved in the requested format, and displayed side-by-side with the original in the web interface with download capability.



## 6. Performance Analysis

Enhanced output quality is assessed using three complementary objective metrics: PSNR for pixel-level reconstruction fidelity, SSIM for structural similarity, and UIQM for underwater-specific perceptual quality across colourfulness, sharpness, and contrast dimensions.

## V. RESULTS AND DISCUSSION

The proposed system was evaluated on the held-out test partition of the UIEB and EUVP datasets comprising 1,439 paired image samples. Quantitative performance was assessed using PSNR, SSIM, and UIQM, with comparative evaluation against four traditional baselines.

### A. Quantitative Results

Table I presents quantitative comparison of the proposed CNN against traditional enhancement approaches. The proposed method achieves substantial improvements over all baselines across every metric.

TABLE I: Quantitative Comparison of Enhancement Methods

Method	PSNR (dB)	SSIM	UIQM	Time (s)
Hist. Equal.	19.4	0.71	2.31	0.02
White Bal.	21.8	0.76	2.58	0.03
CLAHE	23.5	0.81	2.73	0.05
Retime	24.9	0.83	2.87	0.12
Proposed CNN	31.6	0.91	3.54	0.48

The proposed CNN achieves PSNR of 31.6 dB—a 6.7 dB improvement over histogram equalization and 2.1 dB over the strongest traditional baseline (Retinex). The SSIM improvement from 0.83 to 0.91 (9.6% relative) demonstrates superior structural preservation through skip connections. The UIQM improvement from 2.87 to 3.54 confirms perceptually superior colourfulness, sharpness, and contrast in the enhanced outputs.

### B. Evaluation Metrics

Table II summarizes the three primary evaluation metrics used in this study.

TABLE II: Image Quality Evaluation Metrics

Metric	Full Name	Description
PSNR	Peak Signal-to- Noise Ratio	Pixel-level fidelity vs. reference (dB)
SSIM	Structural Similarity Index	Luminance, contrast and structure comparison
UIQM	Underwater Image Quality Measure	Perceptual colorfulness, sharpness, contrast

### C. Visual Results

Qualitative evaluation confirms quantitative findings. On shallow turbid water test images, the proposed method produces visually natural colour restoration with warm tones recovered in the 600–700 nm range absent from raw captures. Object boundaries are sharper, and fine coral and seafloor textures are visible at distances showing only undifferentiated haze in originals.

On deep-water test images (>30 m depth), where red and orange wavelengths are completely attenuated, the model produces plausible colour restoration with appropriately muted rather than artificially saturated colours. In high-turbidity scenarios, dehazing substantially improves visibility, recovering objects at 2–3 meter distances that are effectively invisible in unprocessed inputs.



#### D. Comparison with Traditional Methods

Traditional methods exhibit characteristic failure modes the proposed CNN avoids. Histogram equalization produces over-enhanced images with extreme local contrast and amplified noise. White balance correction improves color cast but provides no dehazing or sharpness improvement. CLAHE improves local contrast but can produce edge halos and does not address colour distortion. Retinex-based methods provide the best traditional results but require careful per-image parameter tuning and fall substantially below the CNN on all metrics.

#### E. Discussion

The results demonstrate that deep learning-based enhancement substantially outperforms all evaluated traditional approaches. The large PSNR margin reflects the CNN's direct pixel-level supervision advantage over unsupervised traditional methods. The SSIM improvement confirms skip connection effectiveness in structural preservation. The primary limitation is computational cost: CPU inference latency of 0.48 seconds per image is acceptable for most applications but insufficient for real-time video processing, which requires GPU inference ( $\approx 0.08$  s/frame).

### VI. CONCLUSION

This paper presented an AI-Based Underwater Image Enhancement System employing a CNN encoder-decoder architecture trained end-to-end on paired underwater image datasets. The proposed system addresses the fundamental challenges of underwater image degradation—wavelength-selective absorption, scattering-induced haze, low contrast, and color distortion—through a learned mapping that generalizes across diverse water conditions and imaging scenarios.

Quantitative evaluation demonstrated PSNR of 31.6 dB, SSIM of 0.91, and UIQM of 3.54, representing consistent and substantial improvements over all evaluated traditional enhancement methods. The practical web-based deployment interface enables real-world use by marine researchers, conservation practitioners, and underwater robotics operators without requiring deep learning expertise or GPU hardware during inference.

### VII. LIMITATIONS AND FUTURE SCOPE

#### Limitations

- **Dependence on Training Data Diversity:** Model generalization is bounded by the diversity of underwater conditions in training data. Water types, depth ranges, or turbidity levels not adequately represented may produce suboptimal results.
- **High Training Computational Requirements:** Training requires GPU hardware with at least 8 GB VRAM and several hours of computation, representing a barrier for practitioners without specialized hardware.
- **Limited Performance on Extremely Degraded Images:** In extreme turbidity, very low illumination, or severe overexposure, information irreversibly lost in capture cannot be recovered by any algorithmic approach.
- **Inference Latency for Video:** CPU inference latency of 0.48 s/frame is insufficient for real-time video at 25–30 fps; GPU inference is required for video applications.

#### Future Scope

- **Advanced Generative Architectures:** Conditional GANs with patch discriminators and Transformer-based architectures can model long-range spatial dependencies more effectively, improving global colour correction coherence.
- **Self-Supervised and Unsupervised Enhancement:** Methods using cycle-consistency constraints on unpaired data would enable training on the much larger collections of unlabelled underwater images, reducing dependence on expensive paired annotations.
- **Real-Time Video Enhancement:** Network pruning, INT8 quantization, and Tensors deployment on NVIDIA Jetson edge devices would enable real-time enhancement for AUV visual navigation and live ROV video streaming.
- **Multi-Task Learning Integration:** Incorporating object detection or semantic segmentation as auxiliary training tasks could guide the enhancement network to preserve features most critical for downstream computer vision analysis.



- Domain Adaptation: Few-shot fine-tuning frameworks for rapid adaptation to specific deployment water conditions would enable practitioners to achieve optimized performance without full model retraining.

#### REFERENCES

- [1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in Proc. IEEE CVPR, pp. 770–778, 2016.
- [2] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao, "An Underwater Image Enhancement Benchmark Dataset and Beyond," IEEE Trans. Image Process., vol. 29, pp. 4376–4389, 2020.
- [3] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in Proc. ICLR, 2015.
- [4] C. Ledig et al., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," in Proc. IEEE CVPR, pp. 4681–4690, 2017.
- [5] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in Proc. MICCAI, pp. 234–241, 2015.
- [6] J. Liu, J. Zhang, and N. Barnes, "Underwater Image Enhancement with a Deep Residual Framework," IEEE Signal Process. Lett., vol. 26, no. 6, pp. 908–912, 2019.
- [7] S. Zhang, T. Peng, and J. Chen, "Underwater Image Enhancement Using Deep Learning and Image Formation Model," IEEE Access, vol. 8, pp. 125473–125485, 2020.
- [8] M. Z. Islam, Y. Xia, and J. Sattar, "Fast Underwater Image Enhancement for Improved Visual Perception," IEEE Robot. Autom. Lett., vol. 5, no. 2, pp. 3227–3234, 2020.
- [9] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and P. Bekaert, "Color Balance and Fusion for Underwater Image Enhancement," IEEE Trans. Image Process., vol. 27, no. 1, pp. 379–393, 2018.
- [10] K. He and J. Shi, "Guided Image Filtering," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 6, pp. 1397–1409, 2013.

