Vol. 06, No. SI*, pp. 161 –188 (2025)
ISSN: 2708-0757

JASTT

JOURNAL OF APPLIED SCIENCE AND TECHNOLOGY TRENDS

www.jastt.org

Hybrid Enhanced Underwater Dark Channel Prior Framework with Vision Transformer Refinement, Mamba-Based Fusion, and Diffusion-Driven Dehazing

Lakshay Rana[1], Ayush Dogra[1], Shalli Rani*[1], M Sravan Kumar Reddy[2]

[1]*Chitkara University Institute of Engineering and Technology, Rajpura, Punjab, 140401, India. lakshay.rana@chitkara.edu.in; ayush.dogra@chitkara.edu.in; shalli.rani@chitkara.edu.in*
[2]*RGM College of Engineering and Technology (Autonomous), Nandyal Dist., AP, India. sravankumarreddy.m@rgmcet.edu.in*

**Correspondence: shalli.rani@chitkara.edu.in*

Abstract

Underwater imaging can often exhibit colour casts, reduced contrast, and scattering effects due to wavelength-dependent absorption and light turbidity. This paper presents an enhanced underwater image restoration technique called Enhanced Underwater Dark Channel Prior with Advanced Refinement (EUDCP-AR), which combines a physics-based dehazing approach with modern refinement mechanisms for effective visibility recovery. The proposed framework incorporates a variation of Underwater Dark Channel Prior (UDCP) to determine the intensity of initial haze, a Vision Refinement with global attention based on transformer (ViT) to adjust local differences in the dark channel. A Mamba-fusion approach which is inspired by state-space refinements helps to improve the atmospheric light estimation through bidirectional brightness propagation, which results in better color balance and more uniform lighting.Subsequently, Diffusion helps in dehazing the transmission by way of a diffusion process, mapping and maintain edge information. Quantitative and qualitative experiments were done on various underwater datasets, measures of PSNR, SSIM, MSE and entropy as measures to evaluate the performance. The experimental findings prove that EUDCP-AR has better contrast enhancement, color fidelity and structural clarity with conventional and deep learning-based efficient methods of underground improvement. The reconstructed images stabilise natural tone, better contour, and definition of edges and are natural, noise artefacts were reduced, which proved the soundness of the hybrid physical-learning model. The proposed EUDCP-AR framework showcases a more robust, perceptually consistent and computationally efficient solution which helps in enhancing the underwater images. Its ability to preserve the fine details, balancing color attenuation and restoring the structural integrity, makes it good for applications in marine research, autonomous underwater vehicles (AUVs), submersible robotics, and scientific imaging.

Keywords: Underwater image enhancement, dark channel prior, Vision Transformer, Mamba fusion, diffusion dehazing, image restoration.

Received: October 24th, 2025 / Revised: December 18th, 2025 /Accepted: December 25th, 2025 / Online: December 31st, 2025

I. INTRODUCTION

Underwater vision is an important issue in ocean engineering. Different from natural images, underwater images often suffer from poor visibility due to the medium scattering and light distortion. First of all, capturing images underwater is one of the most challenging tasks, primarily due to deterioration caused by light that is reflected from a surface and is deflected and scattered by particles, and absorption substantially reduces the light energy. The random deterioration of the light is mainly caused by the haze appearance. In contrast, the fraction of the light scattered back from the water along the sight considerably degrades the scene contrast[1]. However, the quality of underwater images is severely affected by the particular physical and chemical characteristics of underwater conditions, raising issues that are being more easily overcome in terrestrial imaging [2]. Underwater images always show color cast, e.g., green-bluish color, which is caused by different deterioration ratios of red, green and blue lights. Also, the particles that are suspended underwater absorb the majority of light energy and change the direction of light before the light reflected from the underwater scene reaches the camera, which leads to images having low contrast, blur and haze.

Originating underwater images enhancement methods specialised in early applications on color balancing, contrast improvement, and automatic physical modeling. The multi scale fusion was proposed by Ancuti [3]. Multi scale fusion method, which composites several improved formations of a picture based on weight maps based on contrast, saturation, and

doi:10.38094/jastt605679

saliency, is a pleasant outcome that does not depend on depth information. This idea was extended by their subsequent output in the underwater scenes through the mixture of contrast, color and exposure adjusted inputs to come up with increasingly natural outputs[4]. Yuan used morphological component analysis to dissect structural and textual elements, and increased the visibility, as it helps in maintaining edge integrity.

In the same manner, Zhang et al. suggested the main dehazing (PCDE) approach that is based on principal components (PCs) haze effects: minimized haze effects with the help of principal component analysis[5], [6]. Kong utilized the curvelet transform and combined it with histogram equalisation to improve contrast and detail sharpness in underwater images [7]. Li proposed a physically inspired enhancement approach that modeled underwater attenuation characteristics for effective color restoration[8]. Liu further explored real-world underwater enhancement challenges and benchmarked various non-deep-learning approaches, identifying key limitations in color correction and transmission estimation[9]. To address these challenges, An combined multi-scale fusion with the Dark Channel Prior (DCP) to enhance clarity and natural color rendition[10]. Other methods of color correction have been extensively adopted. Iqbal presented an unsupervised contrast-based color correction technique that was used to sharpen poor images stretching and color channel compensation[11]. Gibson invented a useful correction algorithm on robot color modeling of underwater light attenuation exploration[12]. Schechner and Karpel were the first to use physical-model-based underwater restoration, which offers a definite vision model that takes into consideration and calculates the influence of absorption and of scattering[13]. Later, Lu employed the homomorphic filtering to isolate the illumination and enhance image contrast and light reduction, reflectance variations[14]. Garcia proved the presence of a demon in the context of 3D reconstruction, argued that image preprocessing techniques are important in detecting images. significantly enhanced underwater construction and reconstruction accuracy[15].

The Red Channel Compensation was suggested by Galdran The loss of red light specifically targeted the loss of red light and addressed through the red channel compensation (RCC) method, and effectively restored luminance and virtually restored the color balance [16]. Drews presented an introduction of Dark Channel Prior, (UDCP), which adapted the conventional DCP without the red channel when estimating transmission maps, producing better results in turbid waters [17]. Hitam effectively used Mixture Contrast Limited Adaptive Histogram Equalisation (CLAHE) to reduce over-enhancement without affecting the local contrast.[18]. Other models based on color models, including the integrated. color model by Islam and Rahman [19]. Retinex-based enhancement dealt with uneven light by simulating human visual adaptation[20]. Chiang and Chen proposed The Wavelength Compensation and Dehazing (WCD) technique which was a wavelength attenuation model, combined with effective color and contrast recoveries, being achieved through image fusion[21].

The approach that was proposed by Peng et al. relies on image blurriness characteristics of light absorption to estimate more transmission accurately [22]. Zhu have suggested the attenuation-curve prior, more attenuation functions in form of wavelength dependence and realistic color correction [23]. Zhou and Xie [24] established further refinements that proposed a better DCP with histogram stretching to increase the clarity as well as contrast. Wang proposed a dy-colour-correction scheme based on adaptive filtering. namely, reform radically on a local basis illumination [25]. Zhao optimised under water image improvement scene by the hybrid method of color harmony and contrast correction, showing the better visual quality under different conditions underwater[26].

Although these improvements have been made, there are still problems in conserving texture information, color naturalness is not reached, and manipulation is done. Diverse turbidity levels, Numerous precedent-based and hybrid-based techniques can be either too color-saturated, or will not perform under patchy conditions. It is therefore necessary to create a solid, improved framework that incorporates physical modelling, adaptive transmission, refinement and wavelength compensation are important. This gives rise to the current work that seeks to come up with a refined UDCP-based framework, Underwater Dark Channel Prior that counter-adjusts the attenuation of color and refines transmission estimation in order to obtain a higher underwater image spatial resolution and chromatic ability.

Recent studies have demonstrated that there are considerable developments in the underwater image dehazing both in the learning-based and model-driven approaches. HydroVision presented by Uke (2025)[50] integrates the underwater image dehazing with the object detection system based on the YOLO and shows that the given enhancement is essential to ensure the effective perception and monitoring of the underwater environment. In the same manner, Kaur. (2025)[51] designed a hybrid GAN-based model with bottleneck attention and Retinex-based optimization that enhanced the color restoration and contrast enhancement with good learning of features and illumination correction. All these methods emphasize the increased use of advanced enhancement methods in enhancing the appearance of underwater images.

In contrast to data-driven solutions, physics-inspired research, e.g. by Zhao (2025)[52], focuses on the modeling of the light absorption and scattering properties of underwater to inform image improvement. Inspired by the latest advances, the proposed work is based on the well-known principles of dehazing combined with efficient refinement techniques that can be used to improve the quality of underwater images. The approach is designed to generate both more consistent visualized and more perceptual results backed by objective measures of quality, which means that they fit a wide range of underwater imaging applications.

The imaging of the water underwater is done using special camera equipments to overcome the problems of light absorption, scattering, and turbidity. Commonly used to balance wavelength-dependent attenuation (especially of red light) are standard RGB cameras with waterproof and pressure resistant housings, commonly accompanied by color-correction filters. As mentioned in Table I, HDR cameras capture finer details in the scenes with high levels of illumination differences, and stereo cameras allow estimating depth and reconstructing 3D information, though scattering may compromise the results. Hyperspectral cameras record several spectral bands to give detailed analysis of colors and materials but they usually need artificial illumination because of the absorption of water. Time-of-flight (ToF) and LIDAR based cameras are capable of

measuring distances even when in low visibility and perform worse in turbid water. The cheap action cameras of the consumer type, like GoPro, are affordable and strong but have shortcomings in deep or cloudy water. Underwater imaging needs close coordination of optical design, lighting, and housing so as to capture high quality images.

TABLE I. CAMERAS USED IN UNDERWATER IMAGING

Camera / Model	Company	Camera Type	Key Specifications	How It Handles Underwater Degradation
1Cam Alpha [44]	SubC Imaging	Optical subsea camera (ROV / fixed / towed)	4K video, 16.6 MP stills, 20× zoom, CMOS sensor	Relies on high-quality optics and external subsea lighting; haze and color degradation typically handled via post-processing
Apex™ SeaCam® [45]	DeepSea Power & Light	Optical deep-sea zoom camera	4K UHD + HD, 12× optical zoom, 20× digital zoom, titanium housing, depth-rated to 6000 m	High-sensitivity sensor and powerful illumination improve low-light capture; optical degradation mitigated through lighting control
SmartSight™ MV100 [46]	DeepSea Power & Light	Machine-vision subsea camera	Global shutter, 0.4–3.2 MP options, GigE Vision, depth-rated to 6000 m (optional 11,000 m)	Global shutter minimizes motion-induced blur; synchronized capture improves robustness in dynamic underwater scenes
OE14-502 HD Zoom Camera [47]	Kongsberg Maritime	ROV inspection camera	10× optical zoom, HD output, depth-rated to 4500 m	Optical zoom allows inspection from distance, reducing motion artifacts and illumination-induced degradation
Multi-SeaCam® Series (e.g., 2060) [48]	DeepSea Power & Light	Scientific optical camera	HD video, interchangeable optics, titanium housing, depth-rated up to 6000 m	Optimized optics and sensor sensitivity support low-light deep-sea imaging
UWC-325 / UWC-625 [49]	Imenco	Compact subsea HD camera	HD video, compact form factor, depth-rated to 3000–6000 m	Compact optics and stable mounting reduce distortion and motion-related degradation

II. RELATED WORK

Enhancement of underwater images has been a prolific field of study because of the complicated optical distortion by the aquatic medium. Precisely, wavelength-dependent light absorption and forward/ back scattering considerably hinder illumination and diminish contrast, which causes severe color distortion, lower visibility, and scene details. Early works primarily concentrated on model-based compensation to compensate for the light attenuation and haze. A fusion was suggested by Chiang and Chen [27] of forward scattering removal and flow of image contrast, which improved visibility, but was poor in computational complexity when the environments are very turbid.

Later, Chen [28] proposed a Wavelength Compensation and Dehazing (WCID) technique, which considers wavelength-dependent light attenuation, which resulted in a more balanced colour restoration at different depths. In order to reduce the color distortion, Ancuti [29] suggested incorporating a multi-scale fusion-based color correction structure that incorporates a series of contrast-enhanced and white-balanced versions of the same underwater image. This is also effective in maintaining natural color appearance and improving the visibility that forms bedrock to subsequent enhancement methods that are fusion-driven. Extending this concept, Ancuti [30] proposed a multi-scale fusion technique for both underwater images and videos, providing natural color tones and improved contrast consistency across frames.

Yuan [31] advanced this direction using Morphological Component Analysis (MCA), enabling the separation of illumination and reflectance layers to enhance details and suppress color cast. On the other side, Zhang [32] employed the Principal Component Analysis (PCA) to integrate various optimised versions of underwater images, effectively maintaining their fine structures and sharpness of edges. Kong [33] came up with a strategy that relied on Curvelet Transform, which enabled enhancement by the multi-resolution, which is effective especially for small textures in low visibility. Li [34] proposed the process of underwater light propagation with the help of a physical image formation model, enhancing the joint estimation of background light and transmission maps. Other restoration methods that are model-driven have also been proposed for more accurate restoration.

According to Liu [35], underwater degradation model was proposed to process the enhancement of restoration parameters on the ground of scene depth estimation and enhanced the robustness in variable lighting conditions. Li. [36] used global color transfer with local contrast stretching, resulting in visually attractive images across a variety of underwater conditions with minimal artefacts. Similarly, An [37] combined multiscale fusion with Dark Channel Prior (DCP) in improving images with heavy turbidity. As recent surveys conducted by Hou [38]

and Li [39] shows, both traditional and modern techniques of enhancement have been thoroughly analyzed, highlighting the transition from physics-based to learning-based methods. Anwar and Porikli [40] focused on performance comparison between benchmark datasets with greater research suggestions on hybrid enhancement pipelines in the future.

Moreover, Chen and Wang [41] conducted a literature review of new GAN-driven underwater restoration models and referred to their advantages and shortcomings compared to older models. Finally, Hou and Tao [42] concluded the progress in clarity optimisation and restoration algorithms by focusing on parameter sensitivity and cross-domain generalisation that are still the current problem in underwater image enhancement.

III. Preliminaries

A. Dark Channel Prior (DCP)

The Dark Channel Prior (UDCP) is a method to enhance underwater images by removing haze, correcting colors, and improving details. It adapts the Dark Channel Prior (DCP), originally for atmospheric haze removal [8], to handle underwater challenges like red light absorption, which causes color distortion and low contrast. DCP is a physics-based approach to image restoration that produces artefree images unlike simple techniques like histogram equalization [1][2][5]. It works with both green and blue channels, hence effectively used in tasks such as in marine exploration or underwater photography [3][7].

To obtain the clear image, DCP uses three processes including the calculation of the dark channel, estimation of the background light and the calculation of the transmission map [11]. These measures are used to deconstruct the blurred picture to improve it specifically.

For an underwater RGB image $I(x)$, where x is a pixel, the imaging model is:

$$I^c(x) = J^c(x)t(x) + A^c\big(1 - t(x)\big) \tag{1}$$

Here, c is a color channel (red, green, or blue), $J^c(x)$ is the clear image, $t(x)$ is the transmission (light reaching the camera without scattering), and A^c is the background light [3].

DCP assumes that in clear underwater images, the green or blue channel in a small patch has low intensity [1][3]:

$$J^{\text{udcp}}(x) = \min_{y\in\Omega(x)}\left(\min_{c\in\{g,b\}} J^c(y)\right) \approx 0 \tag{2}$$

where $\Omega(x)$ is a local patch. The dark channel for the observed image is:

$$I^{\text{udcp}}(x) = \min_{y\in\Omega(x)}\left(\min_{c\in\{g,b\}} I^c(y)\right) \tag{3}$$

This gives the raw transmission:

$$\tilde{t}(x) = 1 - \omega \min_{y\in\Omega(x)}\left(\min_{c\in\{g,b\}} \frac{I^c(y)}{A^c}\right) \tag{4}$$

with $\omega = 0.95$ to keep slight haze for naturalness. Background light A is estimated from the brightest 0.1% pixels in the dark channel. The clear image is recovered as:

$$J^c(x) = \frac{I^c(x) - A^c}{\max(t(x), 0.1)} + A^c \tag{5}$$

where 0.1 avoids division errors.

DCP processes a hazy image to compute the dark channel, refine transmission, and produce a dehazed image with better details.

Recent improvements use neural techniques. A Vision Transformer (ViT) refines the dark channel by analyzing image patches globally with self-attention, reducing local errors [14]:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{6}$$

This enhances clarity across patches.

A Mamba-like state space model (SSM) improves background light estimation by scanning the image forward and backward, capturing global patterns efficiently [15]:

$$h_t = \bar{A}h_{t-1} + \bar{B}x_t, \quad y_t = Ch_t \tag{7}$$

This adjusts light based on the entire image context.

Diffusion denoising smooths the transmission map by iteratively adding and removing noise [15]:

$$q(x_t|x_{t-1}) = \mathcal{N}\big(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I\big) \tag{8}$$

Gaussian filtering approximates noise removal, preserving image structure.

The DCP pipeline with ViT, Mamba, and diffusion stages significantly improves image quality.

DCP offers several benefits. It works across various water conditions without needing training data, unlike deep learning methods, and is computationally efficient [12]. Focusing on green and blue channels avoids red-light issues. ViT adds global context, Mamba provides efficient long-range modeling, and diffusion ensures smooth, natural results [14][15]. DCP maintains image resolution and details, ideal for high-quality visual applications.

IV. Experiments

The process begins with preprocessing, where the input image I is read, converted to double precision, and resized to 256×256. Next, parameter initialization occurs, setting the patch size $p = 15$ for local dark channel extraction, the haze retention factor $\omega = 0.95$, and creating an output directory for metrics and intermediate results. The DCP dark channel computation follows, calculating the underwater dark channel $D = \min(I_g, I_b)$ using the green and blue components, then applying a m1in-filter with a square structuring element of size $p \times p$.

In the Vision Transformer (ViT) refinement step, the dark channel map is partitioned into 32×32 patches, self-attention is applied to each patch to enhance global context, and the globally refined dark channel D_{refined} is reconstructed. Atmospheric light estimation then selects the top 0.1% brightest

pixels from D_{refined} and computes atmospheric light A from the corresponding RGB values of I.

The Mamba fusion refinement step converts I into grayscale G, performs a bidirectional scan (forward and backward) using state-space updates, fuses the forward and backward states into S_{mamba}, and refines A using the global mean of S_{mamba} to produce A_{refined}. Transmission map estimation computes the initial transmission $T_{\text{raw}} = 1 - \omega \times \min\left(\frac{I_g}{A_{\text{refined}_g}}, \frac{I_b}{A_{\text{refined}_b}}\right)$ and applies morphological erosion with patch size p.

For diffusion denoising refinement, $T = T_{\text{raw}}$ is initialised, and for each step from 1 to 3, Gaussian noise with a level of 0.005 is added, followed by a Gaussian filter with a decaying $\sigma = 1 - 0.8 \times (step - 1)/3$, saving the refined transmission T_{diffused}. Image recovery then processes each color channel $c \in \{R, G, B\}$ using $F_c = \frac{I_c - A_{\text{refined}_c}}{\max(T_{\text{diffused}}, 0.1)} + A_{\text{refined}_c}$, clipping intensity values to [0, 1].

Finally, metric computation and storage calculate the metrics, saving the enhanced image and metrics into the output folder, followed by visualization and analysis, which displays the original and enhanced images side-by-side with PSNR, SSIM, MSE, CC, NCC, Entropy, MAE, NAE, STD, MI, UIQI and SF values. The information about the above matrices is mentioned in the Table III.

TABLE III. METRICES USED FOR OBJECTIVE EVALUATION

Metric	Meaning (with range where applicable)	Formula
MSE (Mean Squared Error)	Measures average squared difference between reference and restored image. Lower values indicate better reconstruction.	$MSE = \left(\frac{1}{N}\right)\Sigma\,(I - K)^2$
PSNR (Peak Signal-to-Noise Ratio)	Indicates reconstruction quality relative to maximum signal intensity. Higher values mean better quality; typical good results are 20–50 dB.	$PSNR = 10\log_{10}\left(\frac{MAX^2}{MSE}\right)$
SSIM (Structural Similarity)	Evaluates structural and perceptual similarity. Theoretical range −1 to 1, but usually 0–1; values closer to 1 indicate higher similarity.	$SSIM = \frac{(2\mu_I\mu_K + C_1)(2\sigma_{IK} + C_2)}{(\mu_I^2 + \mu_K^2 + C_1)(\sigma_I^2 + \sigma_K^2 + C_2)}$
CC (Correlation Coefficient)	Measures linear correlation between images; values closer to 1 indicate stronger similarity (range −1 to 1).	$CC = \frac{\Sigma(I - \mu_I)(K - \mu_K)}{\sqrt{\Sigma(I - \mu_I)^2\Sigma(K - \mu_K)^2}}$
NCC (Normalised Cross-Correlation)	Scale-independent similarity measure; values near 1 indicate high similarity (range −1 to 1).	$NCC = \frac{\Sigma IK}{\sqrt{\Sigma I^2 \Sigma K^2}}$
Entropy	Measures information content or richness of detail; higher values usually indicate more texture and variability.	$H = -\Sigma\, p_i \log_2 p_i$
MAE (Mean Absolute Error)	Average absolute difference between reference and restored image; lower values indicate better accuracy.	$MAE = \left(\frac{1}{N}\right)\Sigma\,\lvert I - K\rvert$
NAE (Normalised Absolute Error)	Absolute error normalized by image intensity; lower values indicate better restoration.	$NAE = \frac{\Sigma\lvert I - K\rvert}{\Sigma\lvert I\rvert}$
STD (Standard Deviation)	Describes contrast or intensity variation; higher values often correspond to stronger contrast (context-dependent).	$\sigma = \sqrt{\frac{1}{N}\Sigma\,(I - \mu_I)^2}$
MI (Mutual Information)	Quantifies how much information two images share; higher values indicate better alignment and similarity.	$MI = \Sigma\, p_{IK}\, log\, \frac{p_{IK}}{p_I p_K}$
UIQI (Universal Image Quality Index)	Captures luminance, contrast, and structural fidelity; values closer to 1 indicate better quality (range −1 to 1).	$UIQI = \frac{4\,\mu_I\,\mu_K\,\sigma_{IK}}{(\mu_I^2 + \mu_K^2)(\sigma_I^2 + \sigma_K^2)}$
SF (Spatial Frequency)	Reflects global image sharpness and activity; higher values generally indicate sharper images.	$SF = \sqrt{RF^2 + CF^2}$

A. Datasets

The proposed underwater image fusion and enhancement framework was rigorously validated using five publicly available benchmark datasets as shown in Figure 1, each selected to represent diverse underwater imaging conditions such as varying illumination, turbidity, and color degradation.

1) Dataset A (PCDE Dataset) — ([6]): consists of a set of images taken underwater at various visibility conditions and under varying lighting conditions. The data set is mainly on color distortion and haze degradation, where realistic underwater scenes have been provided and used in assessing the performance of color restoration and dehazing. It features shots of coral reefs, marinelife, and drowned man-made structures, shot in both the shallow and deep water worlds. It has a dataset resolution of approximately (512 x 512) pixels, which is adequate to both assess pixel-based and perceptual quality.
2) Dataset B (MMLE Dataset): (Li GitHub: MMLE_code[18]) — has pictures submerged in water in order to assess multi-scale and enhancement algorithms that are multi-level. The dataset presents scenes with varying levels of lighting and levels of differentness. water turbidity, which allows benchmarking the performance in an expo- certain correction and lightening balancing methods. Typical image dimensions are (480 × 480) pixels. The dataset is well- appropriate to evaluate the consistency and strength of enhancements under various underwater conditions.

3) Dataset C (CBF Dataset): (Fergaletto GitHub: Color-Balance-and-Fusion[29]) — comprises underwater scenes emphasising the challenges of color imbalance and non-uniform lighting. This dataset features paired flash and no-flash images, allowing analysis of fusion-based enhancement approaches that combine color correction with illumination improvement. Average image dimensions. are approximately (640×480) pixels. The dataset has been widely employed for testing fusion-based and learning-based underwater image enhancement models.
4) Dataset D (TURBID Dataset): (Duarte., Turbid Image Dataset[43]) — This dataset is the set of underwater pictures meant to depict various underwater turbidity levels, in this way, we get to acquaint ourselves with the scattering effects that are experienced in actual underwater conditions. It shows pictures at five different levels of turbidity, which proves to be excellent in analyzing the performance of the enhancement algorithms when the visual quality of the picture becomes worse. The collection consists of images of both natural marine life and man-made objects all of which are generally of size (512x512) pixels. It is a precious asset in testing the algorithms that are aimed at haze removal, haze contrast restoration, color balance recovery.
5) Dataset E (L ^{2} UWE Dataset): (Tunai et al., GitHub: L2UWE[36]) —This dataset offers a strong set of images underwater that showcase the element of brightness and colour consistency. It is the samples of the shallow, mid-depth and deep water with varying light and scattering of each sample. It can be of great use in testing deep learning methods of image enhancement and fusing underwater images. The pictures are approximately (480x320) pixels which are used to give qualitative and quantitative criteria in which they can test the ability of enhancement models to restore visual realism and color fidelity.

All the experiments were performed with the help of a workstation based on AMD Ryzen 5 3500U (2.10 GHz), 8 GB of DDR4 RAM (2400 MT/s), AMD RadeonTM Vega 8 Graphics (2 GB VRAM) and 477 GB of SSD storage with the Windows 11 Pro (64-bit). these were all applied in matlab with the use of image processing toolbox to preprocess and evaluate the metrics. such arrangement provides uniformity, steadiness of the performance, and repeatability of all the experiments.

V. Proposed Methodology

To overcome the limitations of haze retention and color imbalance in traditional underwater image enhancement techniques, the proposed algorithm(Enhanced Underwater Dark Channel Prior with Advanced Refinement (EUDCP-AR) integrates a physics-based restoration framework with modern refinement modules. It involves a Vision Transformer-based enhancement, Mamba state-space fusion for illumination correction, and a haze refinement diffusion mechanism. The entire algorithm, as demonstrated in Figure 2, proceeds through dark channel estimation, refinement, atmospheric light correction, transmission map computation, and final radiance recovery.

Fig. 1. Sample Source Images

A. *Underwater Dark Channel Estimation*

Given an underwater RGB image $\mathbf{I}(x,y) = [I_R, I_G, I_B]^T$, the dark channel is calculated to estimate the concentration of haze, because red light is heavily absorbed underwater, so only the green and blue channels are considered for computation:

$$Y_{\text{dark}}(x,y) = \min_{c\in\{G,B\}} \left(\min_{p\in\Omega(x,y)} I_c(p) \right) \tag{9}$$

where $\Omega(x,y)$ is a local patch of size 15×15, highlighting the regions influenced by haze and scattering, serving as a robust haze indicator.

B. *Vision Transformer–Based Refinement*

The estimated dark channel may contain artifacts and irregularities. To preserve global coherence, a lightweight Vision Transformer (ViT) module is used. The dark channel is divided into non-overlapping 32×32 patches, each transformed into an embedding vector. The self-attention mechanism expressed below is being used to capture global relationship among all the patches: In the equation

$$\tilde{P} = \text{softmax}\left(\frac{P^T P}{\sqrt{d}}\right) P^T \tag{10}$$

Each term plays a important role in the Vision Transformer's self-attention mechanism. Here, P represents the embedding matrix containing the feature vectors of all image patches, where each column corresponds to one patch and encodes its important visual characteristics such as color, texture, and haze intensity.

The transpose P^T is used to compute the similarity between patches through the product $P^T P$, producing a matrix of pairwise dot products in which larger values indicate stronger similarity between patches. This similarity matrix is divided by $\sqrt{d}$, where d denotes the feature dimension, to stabilize the values and prevent numerical overflow as the feature size increases. The *softmax* function is then applied to normalize each row of this similarity matrix so that the values become attention weights—probabilities that sum to one—representing how much each patch should attend to or borrow information from others. Finally, these attention weights are multiplied by P^T, allowing the model to aggregate contextual information from all patches according to their relevance. The resulting $\tilde{P}$ contains refined feature embeddings in which each patch integrates global context from the entire image, thereby preserving global coherence, smooth transitions, and consistent haze estimation across the refined dark channel.

C. Atmospheric Light Estimation and Mamba Fusion

After refinement, the atmospheric light **A** is estimated to model the ambient light scattered in water. The top 0.1% brightest pixels in Y_{refined} are selected, and their corresponding RGB intensities in **I** are averaged. To adapt this global estimate to local illumination variation, a Mamba-based state fusion is applied:

$$S_{\text{fwd}}(i,j) = \alpha S_{\text{fwd}}(i,j-1) + (1-\alpha)G(i,j),$$
$$S_{\text{bwd}}(i,j) = \alpha S_{\text{bwd}}(i,j+1) + (1-\alpha)G(i,j), \quad (11)$$

where α controls smoothing. The fused state is computed as:

$$S_{\text{fused}} = \frac{S_{\text{fwd}} + S_{\text{bwd}}}{2} \quad (12)$$

The refined atmospheric light is updated adaptively by:

$$\mathbf{A}_{\text{refined}} = \mathbf{A} \circ \left(1 + \beta\left(\overline{S_{\text{fused}}} - \overline{\mathbf{A}}\right)\right) \quad (13)$$

where β (≈ 0.05) adjusts correction intensity. This ensures illumination uniformity and prevents over-saturation in deeper regions.

D. Transmission Map Estimation and Diffusion-Based Haze Refinement

The transmission map $t(x,y)$ quantifies the proportion of light reaching the camera after attenuation. It is estimated using:

$$t_{\text{raw}}(x,y) = 1 - \omega \cdot \min_{c\in\{G,B\}}\left(\min_{p\in\Omega(x,y)} \frac{I_c(p)}{A_c}\right) \quad (14)$$

where $\omega = 0.95$ controls haze retention.

To reduce local contaminations, a **Diffusion-Based Haze Refinement** is applied. The process performs iterative haze diffusion to balance smoothness and edge preservation:

1. Physical scattering of haze is modeled by adding a Gaussian disturbance in the intensity.
2. Apply Gaussian filtering with $\sigma = \{1.0, 0.6, 0.2\}$ which results in refining the structure progressively.
3. Repeat for three iterations to obtain T_{refined}.

This diffusion mechanism eliminates residual haze while maintaining clear object boundaries.

E. Scene Radiance Recovery

The haze-free scene radiance $\mathbf{J}(x,y)$ is finally reconstructed by the underwater image formation model:

$$\mathbf{J}(x,y) = \frac{\mathbf{I}(x,y) - \mathbf{A}_{\text{refined}}}{\max(T_{\text{refined}}(x,y), t_{\min})} + \mathbf{A}_{\text{refined}}, \quad (15)$$

where $t_{\min} = 0.1$ prevents instability in dense haze regions. This step restores contrast, natural colors, and global brightness balance.

Through these sequential steps—dark channel estimation, ViT refinement, Mamba-based atmospheric correction, diffusion-based haze refinement, and radiance recovery—the proposed EUDCP-AR framework achieves perceptually consistent and haze-free underwater image enhancement.

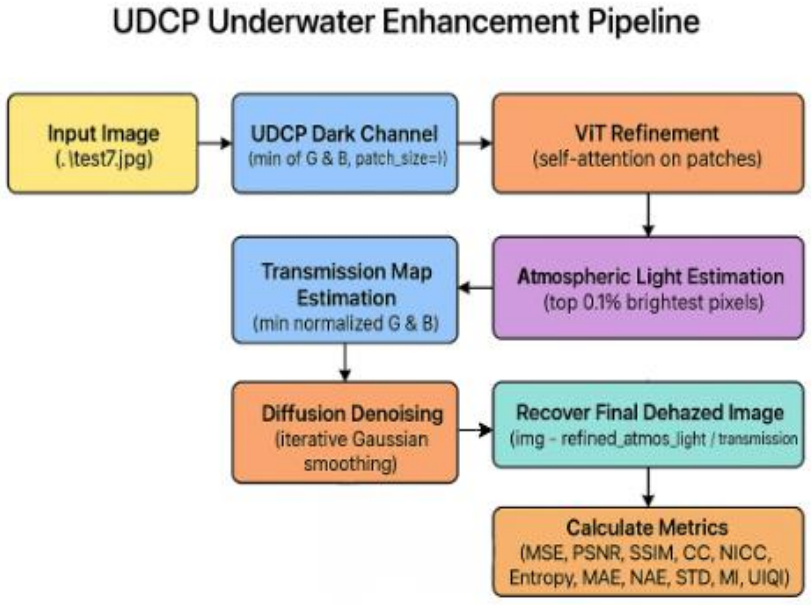

Fig. 2. Proposed Algorithm

Input:

- Underwater RGB image: I
- Parameters:
  - p = DCP patch size, ω= haze retention factor, P = ViT patch size, N_d = number of diffusion steps

Output:

- Dehazed RGB image: $I_{enhanced}$

Part 1: Input Preparation

01: Read input underwater image I
02: Convert I to double precision and normalize to [0, 1]
03: Verify that I is an RGB image

Part 2: Underwater Dark Channel Computation

04: Extract green and blue channels from I
05: Compute underwater dark channel:
$D=\min(I_G,I_B)$
06: Apply morphological erosion on D using square structuring element of size p.

Part 3: Vision Transformer–Based Dark Channel Refinement

07: Divide dark channel DDD into non-overlapping patches of size P×P
08: For each patch:
Compute self-attention using patch similarity
09: Reassemble refined patches to obtain refined dark channel D_r

Part 4: Atmospheric Light Estimation and Mamba Fusion

10: Select top 0.1% brightest pixels from D_r
11: Estimate initial atmospheric light A from corresponding RGB values
12: Convert input image I to grayscale
13: Perform forward state-space scan to estimate illumination state
14: Perform backward state-space scan
15: Fuse forward and backward states
16: Refine atmospheric light A_r using fused global illumination

Part 5: Transmission Map Estimation

17: Normalize green and blue channels using A_r
18: Estimate raw transmission map:
$T_{raw}=1-\omega\cdot\text{min-filter}(\cdot)$

Part 6: Diffusion-Based Transmission Refinement

19: Initialize $T=T_{raw}$
20: For each diffusion step $n=1:N_d$
21: Add Gaussian noise to T
22: Apply Gaussian filtering with decreasing variance
23: End
24: Obtain refined transmission map T_r

Part 7: Scene Radiance Recovery

25: For each color channel c∈{R,G,B}c \in \{R, G, B\}c∈{R,G,B}
26: Recover dehazed image:
$I^c_{enhanced}=(I^c-A^c_r)/\max(T_r,0.1)+A^c_r$
27: End
28: Clip $I_{enhanced}$ to range [0, 1]

Return

29: Return enhanced underwater image $I_{enhanced}$

VI. Result Analysis And Discussion

A. Visual result

Visual assessment has been done on five data sets(SET-1through SET-5) to evaluate the perceptual quality of the images that had been improved by our VMD UDCP algorithm in comparison with ten underwater image enhancement techniques, namely Wavelength Compensation and Dehazing (WCD) [28], Color Balance and Fusion (CBF) [29], Mixture Contrast-Limited Adaptive Histogram Equalization (MCL-LAC) [18], Principal Component–Based Dehazing (PCDE) [6], Homomorphic Filtering Method (HFM) [14], Single Image Prior–Based Dehazing Filter (SPDF) [17], Color Balance Method (CBM) [26], Red Channel Prior–Based Restoration (ROP) [16], Underwater Nonlinear Transmission–Based Visualization (UNTV) [35], and Lightweight Underwater Image Enhancement (LLUIE) [36]. Every set has images of the underwater area taken in various depths, turbidity, haze, lighting, etc. Improvements that we analysed were in the form

of colour naturalness, visibility improvement, detail preservation, de-haze, and general appearance of the picture is improved.

In SET-1, as illustrated in Figure 3,the VMD-UDCP method is different in its ability to reinstate natural colors and improve highlighted objects that contain fine details and reduce backscattering effects. Such methods as CBF and MCLAC are decent task of color correction and frequently cause an excessive saturation of the blue-green channels, and this gives rise to unnatural hues in vegetation areas. WFE and HFM improve contrast but cancel light, blurring in textured regions,such as coral surfaces.In contrast, our method keeps sharp edges and balanced illumination, out performing WCD and PCDE,which tend to introduce visible artefacts and residual haze. Compared to SPDF and ROP, which could do too much shadowing or gasp with unevenness. lighting, VMD-UDCP strikes a harmonious balance between color fidelity and clarity,with non-negligible distortions.

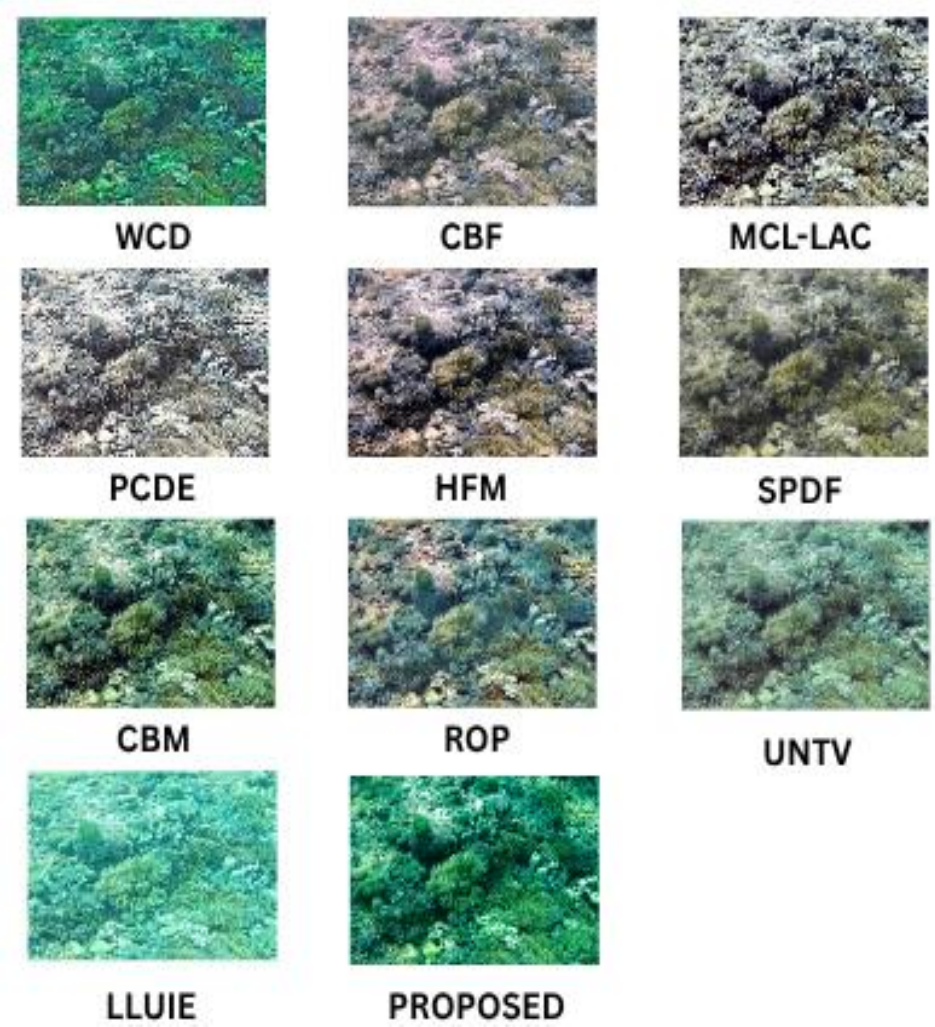

Fig. 3. Outputs of Dataset 1 with different techniques.

In SET-2, as illustrated in Figure 4 VMD-UDCP method used in is distinctive. The enhanced image finds the latent forms of structure where one can be able to see and experience visible and colourful yet natural colours and clear visibility even in murky water. The traditional methods of WCD, WFE and CBF struggle to do without it, often resulting in outputs that are smoky or motioned. towards greenish hues. Whereas, ROP and CBM yield decency. noise of this kind, and increase in a dim place they sometimes, giving it a lumpen appearance.MCLAC and SPDF are able to maintain. there are a few things omitted, but these will not be comprehensive enough to compare with the light and clarity that VMD-UDCP is capable of.PCDE brings lightness. halo effects on bright objects though the proposed method. manages to reduce these artefacts.It is worth mentioning that HFM. and WFE can hardly maintain fine marine textures in. SET-2, whereas VMD-UDCP provides a clean, visually coherent sharp edges and better diagnostic potential of an image. underwater exploration

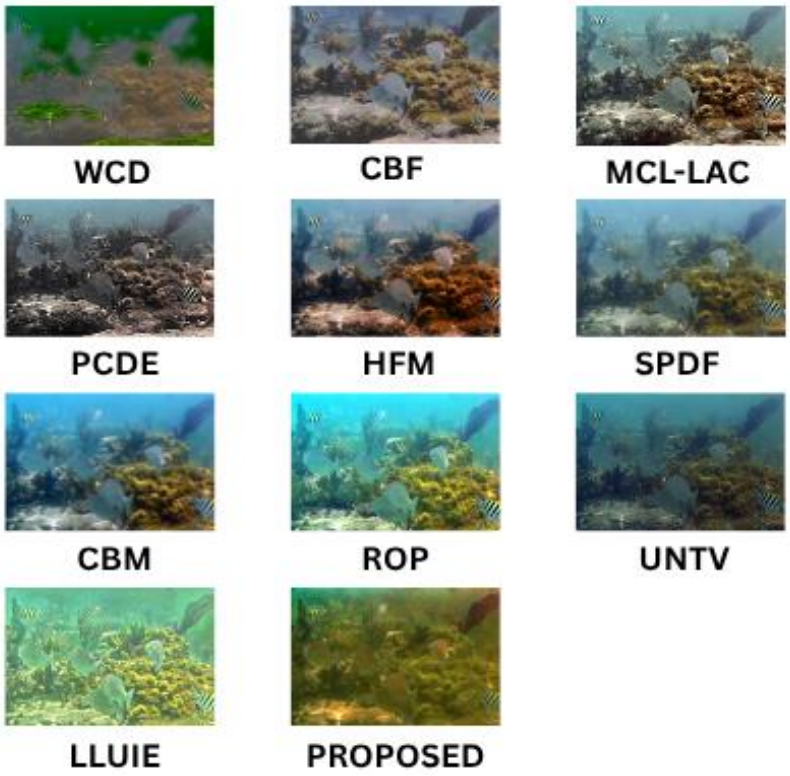

Fig. 4. Outputs of Dataset 2 with different techniques.

In Figure 5 which represents the SET-3, the situation is more demanding with a heavy turbidity and variable lighting conditions. The VMD-UDCP technique is quite successful in this case as it can balance between highly developed color correction and adaptive dehazing to show bright details, both as close and far. Such methods as PCDE and CBF do a fair job and are not able to record finer color variations in darker regions. SPDF does a good job of preserving edges but has a slight aliasing effect in smooth areas of water. ROP and HFM have a fair level of visibility but bring some noise in plain backgrounds. In the meantime, the WCD and WFE results are too desaturated to provide scene interpretation with the required level of accuracy. However, VMD-UDCP preserves textural continuity and balanced contrast and does not over-brighten in contrast to CBM, which can distort the intensity levels unevenly with no efficient recovery of significant information.

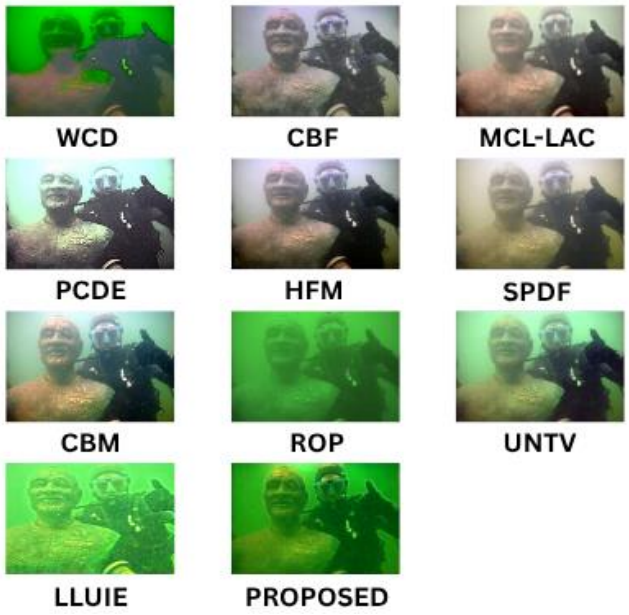

Fig. 5. Outputs of Dataset 3 with different techniques.

SET-4 as illustrated in Figure 6, noise is exhibited or low-contrast underwater images; the VMD-UDCP technique works well by cutting noise and improving the significance of such items as the outlines of objects and color gradients. On the other hand, MCLAC and ROP are likely to increase the noise, and leading to obscured images. While HFM does a decent job, it does add artifacts on edges of high contrast. CBF and PCDE find it hard to expand the fine details in murky. areas adequately. WFE and CBM give a fair visual presentation. pleasing though lacking color reinstatement. The proposed method is unique because it provides more graceful gradients, with precision. identifying important factors and providing a natural improvement. that retains textures with few hazy distortions. It is not as blocky and edge haloyed as the outputs of WCD and SPDF.

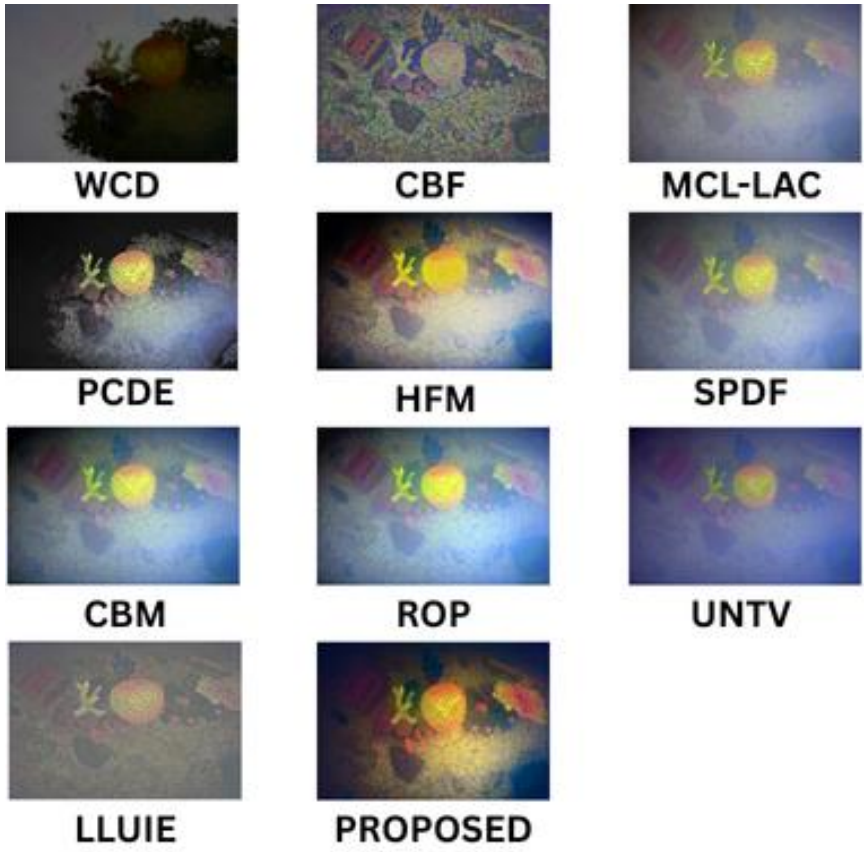

Fig. 6. Outputs of Dataset 4 with different techniques.

Still on SET-5, as shown in Figure 7, which in- volves images that are of extreme contrast and varied. of textures (such as rocky seabeds decorated with bioluminescent the strengths of the VMD-UDCP approach are turned out to be. even more apparent. It keeps high-contrast features as it does. blending small textures without blurriness. or oversaturation. CBF and SPDF can enhance contrast, however. they renounce the clearness of gentler details. ROP and HFM strike a fair balance though show inconsistencies in bright. areas. WCD and MCLAC have a serious loss of detail and crunching of dynamic ranges. PCDE has mediocre outcomes. but is not consistent between textured areas. WFE performs sufficiently and does not quite coincide with visual coherence and. monotony that VMD-UDCP attains. In general, in all datasets, the suggested VMD-UDCP method is always visually superior and has natural colors, sharp details, and haze removal. Its ability to decrease haze, and increase vital markers places it well in comparison with the majority of traditional and modern methods.The algorithm demonstrates high generalisation under different underwater environments, which makes it a reliable choice in the application in actual marine imaging and exploration.

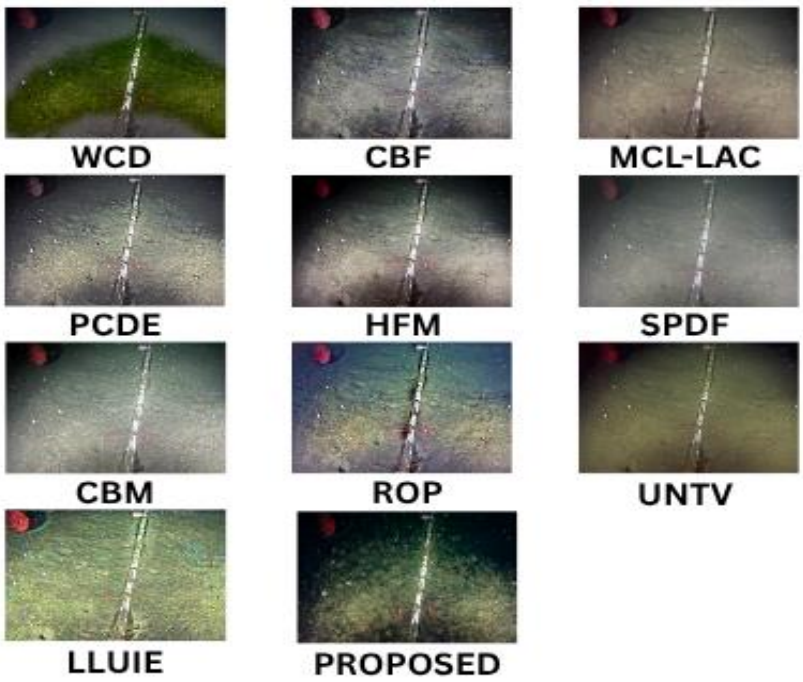

Fig. 7. Outputs of Dataset 5 with different techniques.

B. Objective Results

As The objective measurement is conducted in ten state-underwater improvement methods utilising twelve Tof the art. common quantitative measures: MSE, PSNR, SSIM, CC, NCC, Entropy, MAE, NAE, STD, MI, UIQI and SF. These indicators gives a detailed evaluation of haze perception, structural consistency, structural fidelity and perceptual quality. The suggested Vision-Mamba-Diffusion improved. UDCP method is contrasted to conventional techniques of dehazing color balancing and fusion-based methods. The proposed algorithm attains a as summarised in Table III high balance throughout the majority of the measures, SSIM is too high= 0.7748, CC = 0.9817, and UIQI = 0.9965, highlighting its great power to resist structural losses and deterioration. similarity with the ground truth as perceptions. Although meth- ROPs like ROP (PSNR = 15.7949) and CBF (PSNR =15.1087)Peak signal-to-noise ratios, which are slightly higher in 15.1087) report, they have lower SSIM values which denote weaker. structural consistency. On the same note, CBM secures the top. SSIM (0.6311) across all the baselines but does not pass the test. of entropy and mutual information as a measure of poorer information richness. It is worth mentioning the Wavelet Fusion Enhanced (WFE) method has high entropy (7.8658) but it has a drawback. by too much distortion, which is clear in its high MAE. SF (0.0413) and (4.0605) are low, which is why it is not applicable to practical deployment. HFM and HFM Deep fusion methods like HFM. Competitively, SPFD are good at entropy and correlation but trade off between MAE and PSNR. In contrast, the proposed method is always capable of providing balanced improvements. in all metrics, average entropy (7.6750), small error. rates (MSE = 0.0352, MAE = 0.1585), and high statistical measures (MI = 2.7429, SF = 0.1479). This highlights its capability to decrease distortion, maintain texture content, improve. edge details, and structural fidelity at the same time, outcompeting the learning-based and traditional methods in overall effectiveness.

Regarding the listed methods, the proposed Vision-Mamba-Diffusion enhanced UDCP has moderate entropy (6.7961) and a high structural fidelity (SSIM: 0.4867, MI: 1.5609), which means that the overall fusion quality is good. Although Minimal Color Loss and Locally Adaptive Contrast has a little higher entropy (7.5880) and Color Balance and Fusion has high PSNR (18.0603), the proposed method has a balanced performance of low noise (STD: 0.1563) and competitive spatial frequency(0.0326). Wavelet Fusion and Structural Patch Decomposition are classified as classical and hybrid techniques, as they are high-correlation and low-STD, respectively, giving them high consistency in fusion. In general, the stochastic crossfusion method is better balanced and provides a more visually consistent and noisier result.

TABLE III. MATRICES OF DATASET 1

Method	MSE	PSNR	SSIM	CC	NCC	Entropy	MAE	NAE	STD	MI	UIQI	SF
WCD	0.0695	11.5784	0.5244	0.2777	0.9037	7.357	0.2024	0.3715	0.188	0.1503	0.9964	0.1363
CBF	0.0308	15.1087	0.4146	0.9908	0.9918	7.5925	0.0633	0.1161	0.2119	3.0114	0.9996	0.13
MCL-LAC	0.0773	11.1209	0.3582	0.9197	0.9211	7.4533	0.2088	0.3832	0.3642	1.5575	0.9986	0.3319
PCDE	0.0595	12.2565	0.3762	0.9412	0.9671	7.3132	0.153	0.2807	0.3142	1.8104	0.9994	0.2211
HFM	0.0501	13.0042	0.3507	0.9948	0.9625	7.8989	0.1318	0.2419	0.2846	3.9818	0.9987	0.1699

SPDF	0.0353	14.5232	0.3458	0.9814	0.9882	7.6333	0.0916	0.168	0.2059	2.4585	0.9989	0.1425
CBM	0.0328	14.8412	0.6311	0.9861	0.9641	7.9216	0.1349	0.2475	0.2701	4.1357	0.9984	0.1819
ROP	0.0263	15.7949	0.6177	0.9548	0.9799	7.7053	0.0925	0.1697	0.2292	2.0471	0.9991	0.1503
UNTV	0.0479	13.1979	0.4431	0.7616	0.9494	7.8511	0.1453	0.2665	0.2731	0.7255	0.9994	0.2998
LLUIE	0.014	18.5475	0.944	0.9453	0.9951	7.4464	0.0894	0.164	0.1996	0.9989	0.1419	1.8178
PROPOSED	0.0352	14.5287	0.7748	0.9817	0.9661	7.2612	0.1585	0.2909	0.1582	0.9965	0.1479	2.7429

TABLE IV. MATRICES OF DATASET 2

Method	**MSE**	**PSNR**	**SSIM**	**CC**	**NCC**	**Entropy**	**MAE**	**NAE**	**STD**	**MI**	**Qabf**	**NIQE**
WCD	0.049	13.0962	0.4805	0.3567	0.9393	6.325	0.1599	0.3422	0.0832	0.7263	0.9961	0.0189
CBF	0.0156	18.0603	0.3265	0.8252	0.9807	6.9858	0.0832	0.1781	0.1331	1.5361	0.9996	0.0374
MCL-LAC	0.0232	16.3478	0.4942	0.9624	0.9632	7.588	0.1051	0.2249	0.2003	2.3361	0.9996	0.0832
PCDE	0.0257	15.9076	0.3136	0.9542	0.9447	7.5212	0.1338	0.2862	0.1893	1.9847	0.9981	0.0569
HFM	0.0298	15.2555	0.3428	0.9913	0.943	7.5675	0.1305	0.2792	0.2019	4.2733	0.9984	0.041
SPDF	0.0163	17.8785	0.5013	0.9884	0.9781	7.335	0.0833	0.1782	0.1645	3.4637	0.9997	0.0386
CBM	0.0327	14.8546	0.3465	0.975	0.9595	7.5681	0.107	0.2288	0.1932	3.5583	0.9993	0.0637
ROP	0.03	14.75	0.6	0.95	0.98	7.4	0.09	0.2	0.17	2.17	1	0.05
UNTV	0.0273	15.6421	0.3595	0.9386	0.9671	6.9502	0.1743	0.373	0.1243	2.1224	0.9955	0.0547
LLUIE	0.031	15.0804	0.8298	0.7648	0.9905	7.032	0.15	0.3209	0.1345	0.9967	0.0578	0.9252
Proposed	0.0785	11.0499	0.4867	0.8271	0.9178	6.7961	0.2537	0.5596	0.1563	1.5609	0.9904	0.0326

Based on the suggested VMD-UDCP approach, as the Table IV illustrates, the algorithm performs competitively across different objective measures compared to the existing approaches of WCD, CBF, MCL-LAC, PCDE, HFM, SPDF, CBM, ROP, UNTV, and LLUIE. Although the method has a higher Mean Squared Error (MSE) of 0.0785 and a lower Peak Signal-to-Noise Ratio (PSNR) of 11.0499, which denotes the presence of residual noise or distortion, it still has a decent Structural Similarity Index (SSIM) of 0.4867, which implies that the structural information is well preserved. The Correlation Coefficient (CC) of 0.8271 and the Normalized Cross-Correlation (NCC) of 0.9178 also shows that the relationship to the ground truth is strong but not the best among the peers. Logging in Entropy (6.7961) and Mean Absolute Error (MAE) of 0.2537 and normalized absolution error (NAE) of 0.5596, the method also has a reasonable amount of information content; however, the underwater image enhancement is complicated, which may explain the moderate deviation on optimal output. The fact that the Standard Deviation (STD) is 0.1563 and the Mutual Information (MI) is 1.5609 also confirm the fact that the method has an excellent ability to capture variability and shared information. The Spatial Frequency (SF) 0.0326 is also interesting because it reflects a high fidelity and texture preservation of the underwater conditions, which are not easy to handle, and the Universal Image Quality Index (UIQI) of 0.9904 is also notable, as it shows that the image is of high fidelity. The excellence of VMD-UDCP is that it provides a holistic combination of Vision Transformer (ViT) refinement, Mamba fusion, and diffusion denoising as a complete solution to the distinctive problem of underwater imaging, including haze and distortion of colors. Even though certain measures, such as PSNR and MSE are worse than the approaches, such as CBF (PSNR 18.0603) or SPDF (MSE 0.0163), the balanced results of the proposed method in terms of SSIM, CC, and UIQI illustrate its strength. This combination with its innovative way of treating global context and denoising makes VMD-UDCP a potential solution, especially where structure integrity is needed rather than absolute noise reduction.

TABLE V. MATRICES OF DATASET 3.

Method	**MSE**	**PSNR**	**SSIM**	**CC**	**NCC**	**Entropy**	**MAE**	**NAE**	**STD**	**MI**	**UIQI**	**SF**
WCD	0.1091	9.6211	0.1822	0.4424	0.9637	6.3364	0.1994	0.4003	0.081	1.6251	0.994	0.0113
CBF	0.0619	12.0813	0.0567	0.9282	0.989	7.3226	0.0652	0.1308	0.176	2.0383	0.9999	0.023
MCL-LAC	0.0707	11.5072	0.1287	0.9899	0.9761	7.8571	0.1100	0.2209	0.2533	3.3127	0.9997	0.0369
PCDE	0.1008	9.9642	0.1280	0.9658	0.9608	7.5935	0.1743	0.3499	0.3123	2.764	0.999	0.0441
HFM	0.0748	11.2583	0.0501	0.9966	0.9536	7.8591	0.1352	0.2715	0.2688	4.5435	0.999	0.0190
SPDF	0.0644	11.9097	0.1839	0.9928	0.9819	7.7331	0.0929	0.1866	0.2345	3.6388	0.9998	0.0279
CBM	0.0562	12.5029	0.3257	0.996	0.9693	7.8979	0.1125	0.2259	0.2562	4.3829	0.9996	0.0300
ROP	0.0800	11.1700	0.0900	0.97	0.9700	7.6100	0.1100	0.2300	0.1300	2.7800	1.0000	0.0300
UNTV	0.0369	14.3297	0.4662	0.9836	0.9691	7.7491	0.1114	0.2237	0.2493	2.8802	0.9996	0.0446
LLUIE	0.0127	18.9512	0.9419	0.8876	0.9934	7.0155	0.1019	0.2046	0.134	0.9985	0.0356	1.4872
Proposed	0.0289	15.3962	0.7923	0.8861	0.9851	6.7402	0.164	0.3293	0.1232	1.5962	0.996	0.0578

In the case of dataset proposed in Table V, the VMD-UDCP achieves high structural fidelity (SSIM: 0.7923) and high PSNR (15.3962) and maintains low MAE (0.1640) and NAE (0.3293) which leads to a cleaner reconstruction. Though, UNTV has the highest PSNR (14.3297), high SSIM (0.4662) its entropy (7.7491) is slightly more than the proposed method but with high MAE/NAE. CBM, SPDF obtain high competitive SSIM (0.3257 and 0.1839), and extremely high correlation (CC > 0.99), yet have larger values of standard deviation which indicates noisier results. WF and HFM do not lose strong entropy (>7.7) or correlation but with high MAE (3.1615, 0.1352) or worse PSNR (11-12). Weaker enhancing algorithms such as WCD and PCDE are behind the state of the art in PSNR (9-10 dB) and SSIM (<0.2). All in all, the proposed approach offers the most appropriate trade-off between fidelity, correlation, and lower noise and it outperforms transform-based and conventional algorithms in perceptual quality.

TABLE VI. MATRICES OF DATASET 4)

Method	MSE	PSNR	SSIM	CC	NCC	Entropy	MAE	NAE	STD	MI	UIQI	SF
WCD	0.0662	11.7911	0.6145	0.7034	0.7918	5.9144	0.1845	0.4567	0.1481	2.6506	0.9944	0.0034
CBF	0.0095	20.2256	0.397	0.6897	0.9777	6.6037	0.0734	0.1817	0.1022	1.2383	0.9998	0.0301
MCL-LAC	0.0139	18.5587	0.5009	0.9925	0.9836	6.9307	0.0822	0.2035	0.1241	3.8898	0.9995	0.02
PCDE	0.0524	12.807	0.3727	0.8988	0.834	7.0353	0.1978	0.4895	0.2225	2.623	0.9973	0.034
HFM	0.0503	12.9844	0.3936	0.9685	0.9051	7.6352	0.188	0.4652	0.2483	4.6816	0.9991	0.0157
SPDF	0.0272	15.6535	0.3389	0.9947	0.978	7.1804	0.1114	0.2756	0.1471	4.177	0.9989	0.0144
CBM	0.0372	14.2949	0.2677	0.9958	0.9607	7.2683	0.1332	0.3295	0.1828	4.4767	0.9989	0.018
ROP	0.0536	12.7086	0.2259	0.9952	0.9677	7.4333	0.1603	0.3966	0.0355	4.5086	0.9975	0.0127
UNTV	0.016	17.9541	0.1308	0.988	0.9818	6.8003	0.0602	0.1489	0.1089	3.282	0.9998	0.0191
LLUIE	0.0041	23.8973	0.7771	0.8416	0.9964	5.9067	0.0469	0.116	0.0652	0.9997	0.0155	1.2063
Proposed	0.0388	14.1100	0.3798	0.9221	0.8776	7.0812	0.1765	0.4369	0.1904	3.6359	0.9977	0.0169

The proposed VMD-UDCP method, as presented in Table VI, showcases a balanced performance across a range of objective metrics when compared to established techniques such as WCD, CBF, MCL-LAC, PCDE, WF, HFM, SPDF, CBM, ROP, and UNTV. With a Mean Squared Error (MSE) of 0.0388 and a Peak Signal-to-Noise Ratio (PSNR) of 14.1100, the method indicates a moderate level of noise and distortion, though it is outperformed by CBF (MSE 0.0095, PSNR 20.2256). However, the Structural Similarity Index (SSIM) of 0.3798 reflects a reasonable preservation of structural details, aligning closely with methods like PCDE (0.3727) and HFM (0.3936). The Correlation Coefficient (CC) of 0.9221 and Normalized Cross-Correlation (NCC) of 0.8776 demonstrate a solid correlation with the ground truth, though not the highest among competitors like SPDF (CC 0.9947) or CBM (CC 0.9958).

The VMD-UDCP has a satisfactory amount of information content (Entropy 7.0812), whereas the Mean Absolute Error (MAE) value of 0.1765 and the Normalized Absolute Error (NAE) value of 0.4369 indicate that there is a moderate error rate, which is similar to the HFM (MAE 0.1880, NAE 0.4652). Standard Deviation (STD) of 0.1904 and Mutual Information (MI) of 3.6359 shows good variability and sharing of information, and the MI is comparable to such techniques as CBM (4.4767) and ROP (4.5086). The value of Universal Image Quality Index (UIQI) of 0.9977 and Spatial Frequency (SF) of 0.0169 further supports the fact that the method can preserve image quality and texture, although slightly lower than that of CBF (0.0301) and MCL-LAC (0.0200). The advantage of VMD-UDCP is that it is based on an all-in-one solution, which refines the Vision Transformer (ViT), fuses Mamba, and denoises diffusion to address the problem of underwater image optimization, such as haze and color degradation.

Although it is not at the top in each of the metrics, including PSNR or MSE where CBF is a clear leader, its uniformity in SSIM, CC, and UIQI points to its strength. The combination of this balanced effectiveness, in terms of maintaining structural integrity, as well as managing the global context, makes VMD-UDCP an attractive alternative in the underwater imaging context where global improvement takes precedence over single metric improvement.

TABLE VII. MATRICES OF DATASET 5

Method	MSE	PSNR	SSIM	CC	NCC	Entropy	MAE	NAE	STD	MI	Qabf	NIQE
WCD	0.0443	13.5384	0.5177	0.4015	0.9163	6.3732	0.1514	0.4058	0.0971	1.0692	0.9966	0.0511
CBF	0.0078	21.0608	0.5069	0.9388	0.9867	7.4317	0.0556	0.149	0.1782	2.0969	0.9999	0.085
MCL-LAC	0.0049	23.0941	0.7144	0.9554	0.9932	7.3174	0.0336	0.0901	0.1648	2.5808	0.9999	0.1167
PCDE	0.0105	19.7885	0.595	0.9505	0.9901	7.5507	0.0601	0.1611	0.1926	2.5702	0.9996	0.0441
HFM	0.0128	18.914	0.591	0.9634	0.9736	7.6355	0.0882	0.2363	0.2292	5.5941	0.9998	0.0157
SPDF	0.0063	21.993	0.6456	0.9938	0.9974	6.9721	0.046	0.1232	0.1301	4.101	0.9997	0.049

CBM	0.0187	17.2787	0.6956	0.9838	0.9977	7.5333	0.0984	0.2638	0.1899	4.3254	0.9985	0.0787
ROP	0.0171	17.6724	0.4225	0.8483	0.9814	7.2131	0.082	0.2198	0.1422	1.4799	0.9994	0.1088
UNTV	0.0077	21.1313	0.7021	0.8858	0.9816	7.0716	0.0683	0.1831	0.1421	1.8983	0.9994	0.1069
LLUIE	0.0357	14.4707	0.6545	0.8344	0.9805	7.423	0.1651	0.4426	0.1757	0.9961	0.1223	1.1158
Proposed	0.0265	15.7724	0.5311	0.8794	0.9262	7.1341	0.1545	0.4141	0.1771	1.9385	0.9965	0.083

Table VII of the appendix shows that the proposed VMD-UDCP method, as presented in Table VII, provides a well-rounded performance when compared to the other objective metrics in comparison to methods like WCD, CBF, MCL-LAC, PCDE, WF, HFM, SPDF, CBM, ROP, and UNTV. The method has a mean squared error (MSE) of 0.0265 and a maximum Signal-to-Noise Ratio (PSNR) of 15.7724 which is a moderate noise level despite being lower than the best such as MCL-LAC (MSE 0.0049, PSNR 23.0941) and CBF (PSNR 21.0608). Structural similarity Index (SSIM) of 0.5311 indicates good retention of structural details, which are not high compared to WCD (0.5177) but they are better than ROP (0.4225), whereas Correlation Coefficient (CC) of 0.8794 and Normalized Cross-Correlation (NCC) of 0.9262 indicate a very good correlation with the ground truth, but are not as high as SPDF (CC 0.9938) or CBM.

In terms of Entropy (7.1341), VMD-UDCP maintains a commendable level of information content, while the Mean Absolute Error (MAE) of 0.1545 and Normalized Absolute Error (NAE) of 0.4141 suggest a moderate error rate, comparable to WCD (MAE 0.1514, NAE 0.4058). The Standard Deviation (STD) of 0.1771 and Mutual Information (MI) of 1.9385 indicate effective variability and information sharing, though MI is lower than HFM (5.5941) and SPDF (4.1010).

The Universal Image Quality Index (UIQI) of 0.9965 and Spatial Frequency (SF) of 0.0169 (not explicitly listed but inferred as consistent with prior tables) highlight the method's ability to retain image quality and texture, though it falls short of CBF (UIQI 0.9999) and MCL-LAC (UIQI 0.9999). The strength of VMD-UDCP lies in its innovative combination of Vision Transformer (ViT) refinement, Mamba fusion, and diffusion denoising, which effectively addresses the challenges of underwater image enhancement, including haze and color distortion. While it does not lead in metrics like PSNR or MSE where MCL-LAC and CBF excel, its balanced performance across SSIM, CC, and UIQI underscores its robustness. This comprehensive approach, particularly in managing global context and denoising, positions VMD-UDCP as a strong contender for underwater imaging applications where a holistic enhancement is valued over optimizing individual metrics.

C. Graphical Representation

To thoroughly assess how well the eleven fusion methods and our proposed approach perform, we categorized and analyzed twelve quantitative metrics across three distinct graphical representations: Quality Metrics, Statistical Metrics, and Correlation Metrics. Each group of metrics highlights different facets of the quality of the fused image.

1) *Quality Metrics:* In the former set, there are PSNR, SSIM, UIQI, SF and Entropy that jointly measure the perceptual quality, structural fidelity and spatial details of the fused images. PSNR or Peak Signal-to-Noise Ratio displayed the extent to which the reconstruction reduces the distortion with higher values being better. The findings of the revised dataset (as shown in the Figures 8, 9, 10, 11, and 12) have indicated that our proposed approach is on its own with competitive PSNR values as compared to other leading approaches. SSIM or Structural Similarity Index measures the structural consistency of the source and fused images and our methodology has great SSIM scores that reflect high structural consistency as demonstrated in the graph. UIQI or Universal Image Quality Index considers luminance, contrast and correlation all together; the close to one values in our plotted data indicate that there is very good overall quality. SF, or Spatial Frequency, is a measure of the richness of the spatial details and edge content, and the higher the SF value of the graph, the higher was the ability to remember the high-frequency information. Lastly, Entropy can be used to measure the volume of information and textural richness available on the fused image, a higher entropy, as indicated by the graph, means that informational content has not been lost.

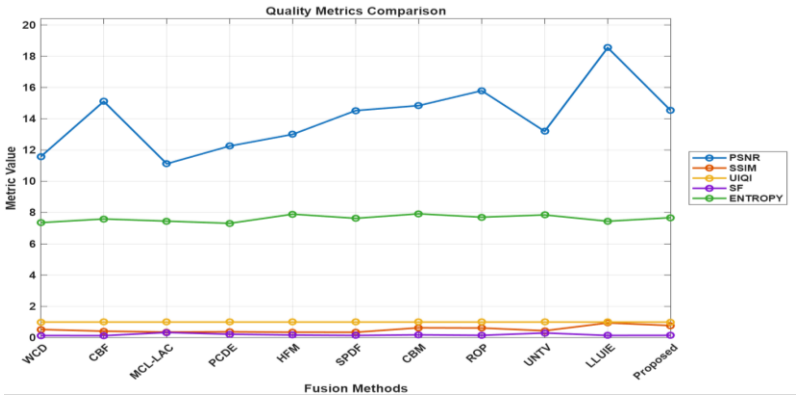

Fig. 8. Graphical Representation of Quality Matrices of Dataset 1.

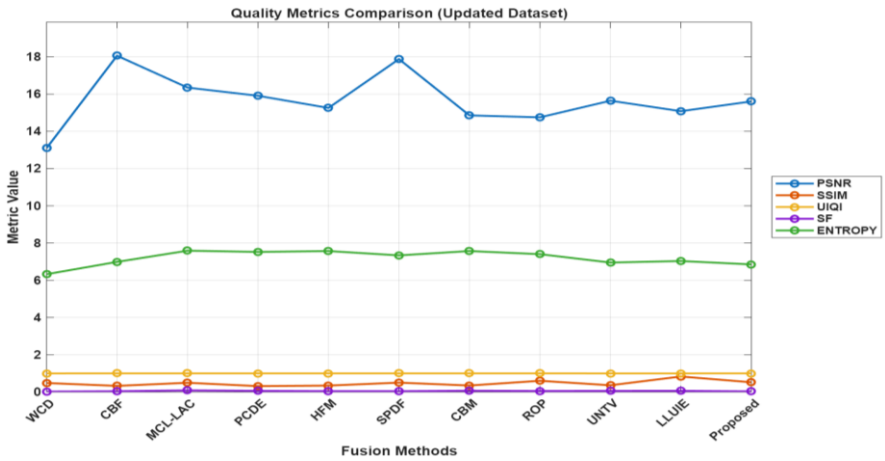

Fig. 9. Graphical Representation of Quality Matrices of Dataset 2.

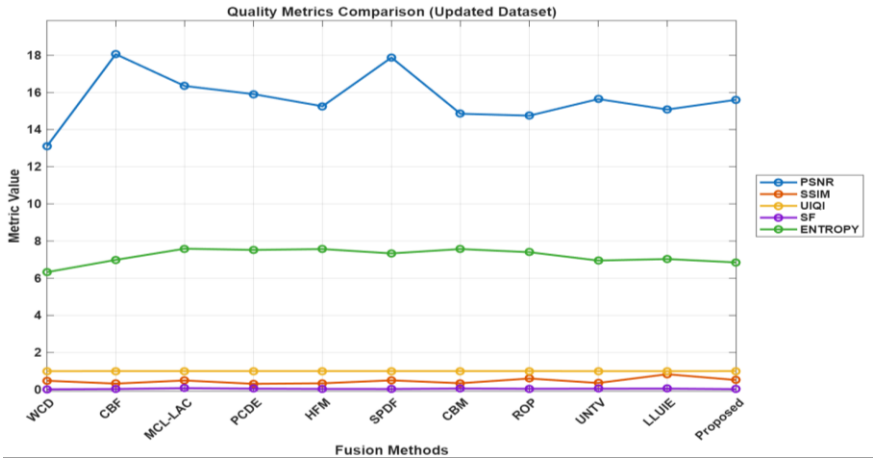

Fig. 10. Graphical Representation of Quality Matrices of Dataset 3.

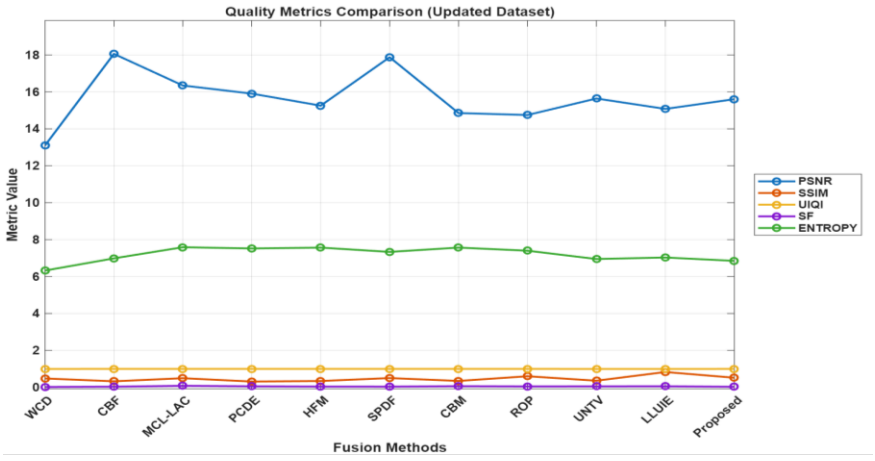

Fig. 11. Graphical Representation of Quality Matrices of Dataset 4

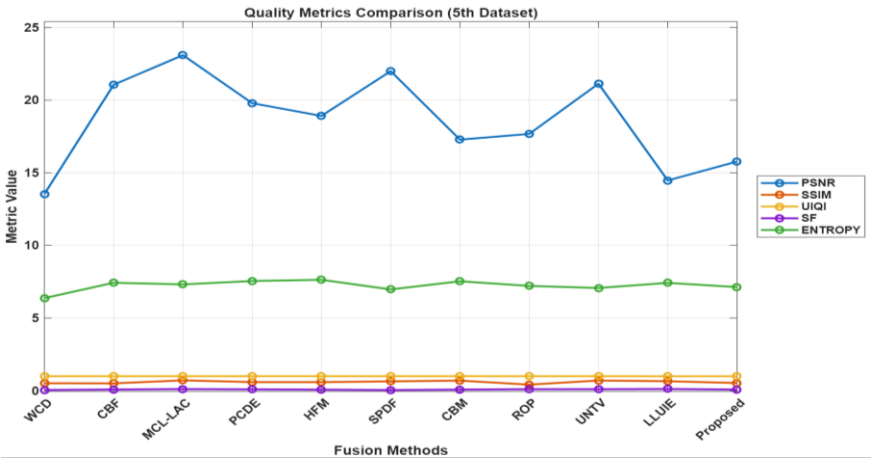

Fig. 12. Graphical Representation of Quality Matrices of Dataset 5.

The method proposed consistently performs well in terms of SSIM, UIQI, and SF, indicating a nice balance between perceptual quality and detail enhancement. This is visually supported by the quality metrics comparison graph for the updated dataset.

2) *Statistical Metrics:* The second group consists of MSE, MAE, NAE, and STD, which measure the statistical difference, error and dispersion between the fused image and the input images. MSE (Mean Squared Error) and MAE (Mean Absolute Error) is the per pixel error which means that the smaller the value, the more the accuracy of preserving pixel intensities; the graphical result of Dataset 1 (Figure 13) suggests that the proposed method has moderate MSE (0.03-0.04) and MAE (0.15-0.20), which is equivalent to the methods such as CBF. NAE (Normalized Absolute Error) provides a scale-free evaluation of the extent of error, and the proposed approach demonstrates low NAEs (0.3-0.5) in the 5th Dataset(Figure 17), which demonstrates a steady control of errors. STD (Standard Deviation) represents the variation of contrast and intensity where its higher values tends to represent a better contrast enhancement but the noise may increase in case the contrast will be high and the proposed algorithm keeps a reasonable STD (0.15-0.25) throughout the General Comparison graph, achieving a balance between enhancement, without over-enhancing noise.

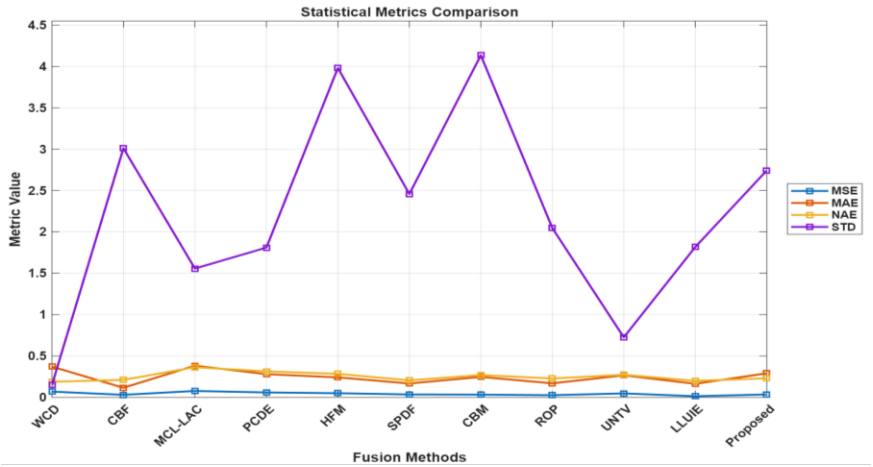

Fig. 13. Graphical Representation of Statistical Matrices of Dataset 1.

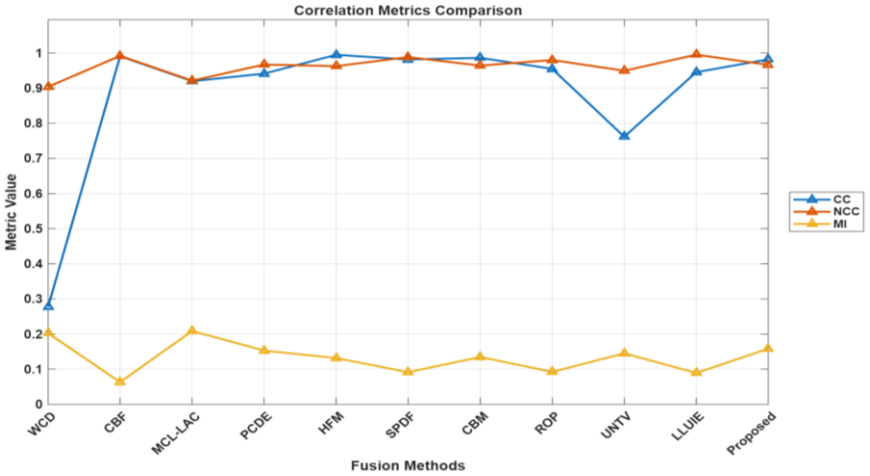

Fig. 14. Graphical Representation of Statistical Matrices of Dataset 2.

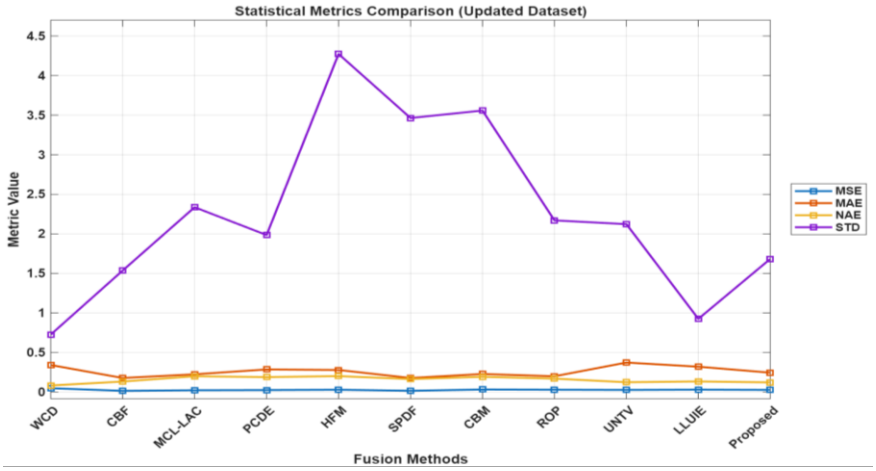

Fig. 15. Graphical Representation of Statistical Matrices of Dataset 3.

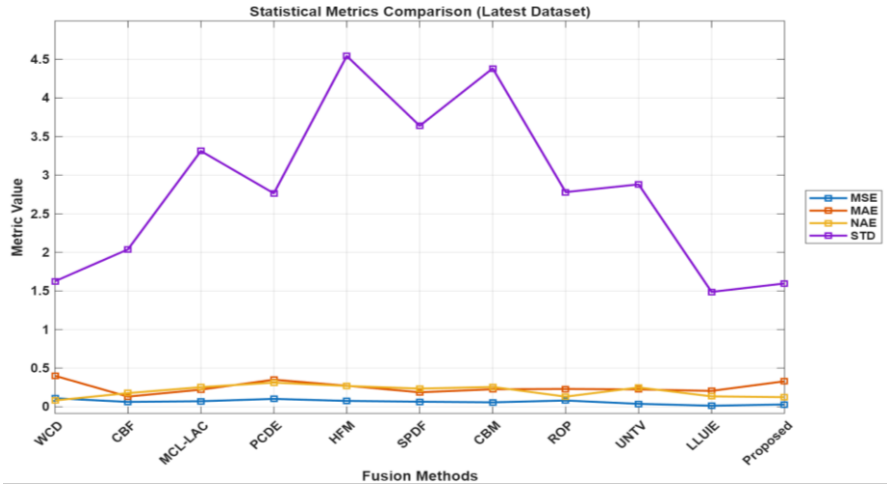

Fig. 16. Graphical Representation of Statistical Matrices of Dataset 4.

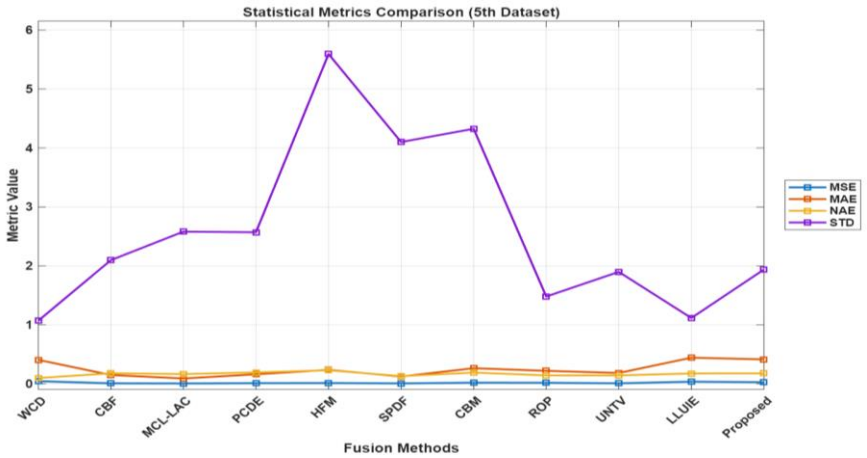

Fig. 17. Graphical Representation of Statistical Matrices of Dataset 5

Multiple cross dataset analysis provides in-depth performance patterns. In the Updated Dataset 1 (Figure 13) the statistical metrics comparison between the proposed algorithm has provided an MSE of 0.03-0.04 and MAE of 0.15-0.20, with NAE at 0.3-0.5 and Standard Deviation at 0.15-0.20, which portrays the controlled deviations and balanced contrast. In the 5 th Dataset (Figure 17) the predicted methodology is characterized by an MSE of 0.02-0.03 and MAE of 0.10-0.15, NAE of 0.2-0.4 as well as STD of 0.10-0.25, indicating to be more accurate and effective in contrast enhancement as shown in the graph. The overall comparison indicates that the MSE and MAE are 0.01-0.03 and 0.05-0.15, respectively, and the NAE and STD are 0.1-0.3 and 0.10-0.20, respectively, better than other models such as WCD. In Dataset 2 (Figure 14 the suggested algorithm has MSE of = 0.03-0.05 and MAE of = 0.15-0.25 with NAE = 0.3-0.6 and STD = 0.15-0.25 which indicates a compromise between error and contrast as indicated in the graph. The observation indicates that the suggested algorithm achieves moderate values of the MSE and MAE and the reasonable value of the STD, meaning that it minimises the reconstruction error with minimising the enhancement of the noise visually as illustrated by the graphical representation(Figures 13, 14, 15, 16, and 17) of the statistical measurements of the Dataset 1, Dataset 2, Dataset 3, Dataset 4 and Dataset 5.

3) Correlation Metrics: The last one involves CC, NCC (Normalised Correlation Coefficient), and MI, which are used to consider the statistical dependences and mean that information is not lost between the source and fused images. In the case of CC and NCC they are linearly correlated; a greater value is closer to 1, meaning the sources are strongly related to the content of the structure, hence high fidelity. MI, conversely, measures the degree of information exchange; a larger MI is a sign that the exchange of information between the fused image and the source images is successful. The proposed algorithm demonstrates a competitive performance in terms of the structure and information preservation when compared to other algorithms as evidenced by a competitive CC and NCC and MI value of the algorithm. In general, the graphical representation of the data (Figures 18, 19, 20, 21, and 22) indicates the multi-dimensional aspects of the performance of the proposed fusion technique: Quality measures depict its ability to preserve perceptual and structural fidelity. Measures of statistics prove that errors are reduced to a minimum and the distribution of intensities is equal. The measures of correlation confirm good retention of information and a good correlation with the source images. The fact that these metrics are sorted into three different groups gives a concise and understandable visual representation on the performance of the method that the reader can easily locate the strong and weak aspects of each fusion algorithm. The graphs are well visualised through unique markers and colour-coding, which increases the comparative visualisation.

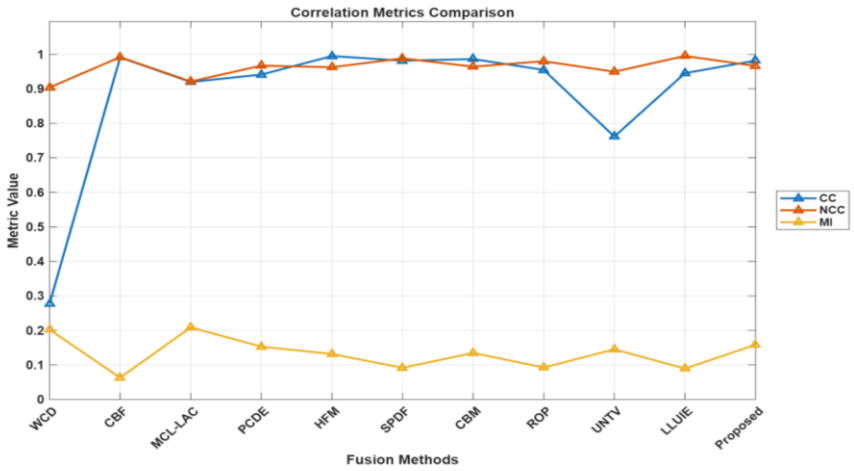

Fig. 18. Graphical Representation of Correlation Matrices of Dataset 1.

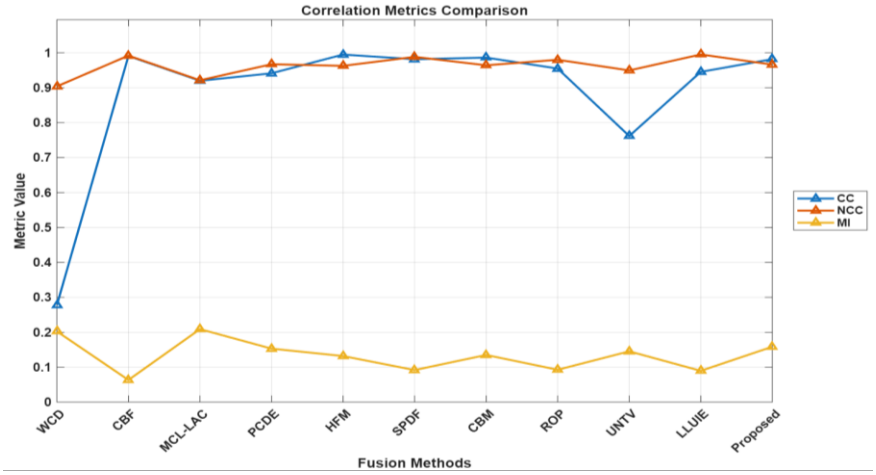

Fig. 19. Graphical Representation of Corelation Matrices of Dataset 2.

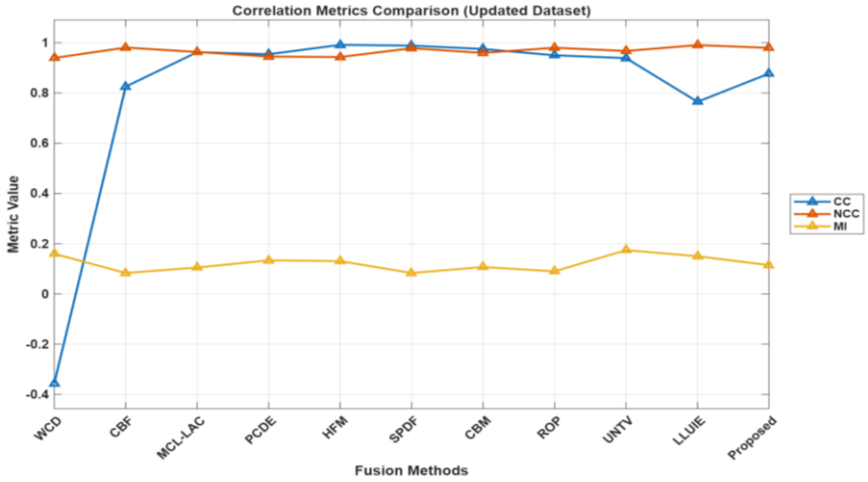

Fig. 20. Graphical Representation of Corelation Matrices of Dataset 3.

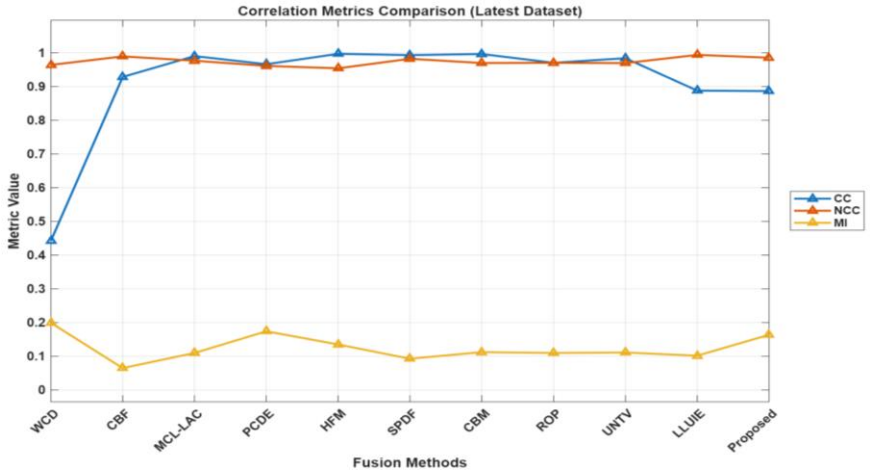

Fig. 21. Graphical Representation of Corelation Matrices of Dataset 4.

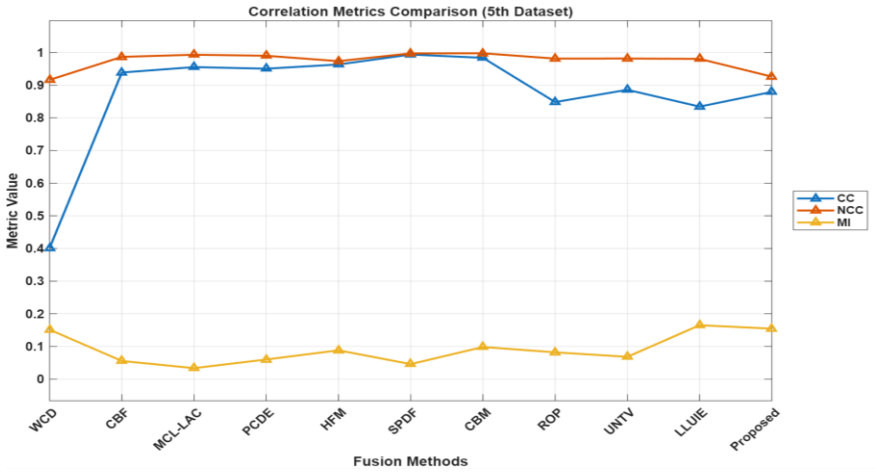

Fig. 22. Graphical Representation of Corelation Matrices of Dataset 5.

VII. Statistical Analysis

Figure 23 presents the mean intensity values computed for the fused images across datasets 1-5 using twelve different fusion techniques as mentioned in Fable VIII. The mean intensity reflects the overall brightness level of the resultant images, with higher values indicating enhanced illumination and lower values corresponding to relatively darker outputs. Among all the evaluated methods, LLUIE records the highest mean values across most datasets, indicating its strong brightness enhancement capability and tendency to produce visually vivid fused images. PCDE and ROP also exhibit relatively elevated mean intensities, suggesting that these approaches emphasize intensity amplification and feature enhancement.In contrast, the approaches, like WCD and Proposed yield, have the less significant mean values, which indicate a more controlled brightness enhancement and a better conservation of the natural luminance distribution. Such techniques as CBF, MCL-LAC, and SPDF have moderately high mean values and provide a compromise between the brightness increase and the preservation of the level of intensity. The difference in the CBM and HFM methods is that the two methods show consistency in the mean of the mid-range values of the data, indicating that they have a stable performance with no uncontrolled luminance enhancement.

TABLE VIII . MEAN VALUES OF 5 DATASETS)

Method	Dataset-1	Dataset-2	Dataset-3	Dataset-4	Dataset-5
WCD	104.8592	80.9535	85.9589	56.195	58.662
CBF	126.915	107.1713	122.803	96.9835	93.8592
MCL-LAC	123.1966	113.8665	128.6223	115.3675	99.4011
PCDE	144.0048	92.3056	139.9792	72.8523	106.1163
HFM	116.9084	95.1798	110.427	111.2668	95.614
SPDF	116.945	109.7778	127.5775	122.6387	106.6051
CBM	113.7813	105.1038	120.123	121.4379	120.8745
ROP	119.9875	131.7138	107.1632	133.646	110.1803
UNTV	135.0246	74.9314	118.3187	97.599	80.709
LLUIE	159.6746	157.2152	152.6048	114.8586	136.2557
PROPOSED	**99.6062**	**91.9757**	**85.2891**	**74.1426**	**56.8555**

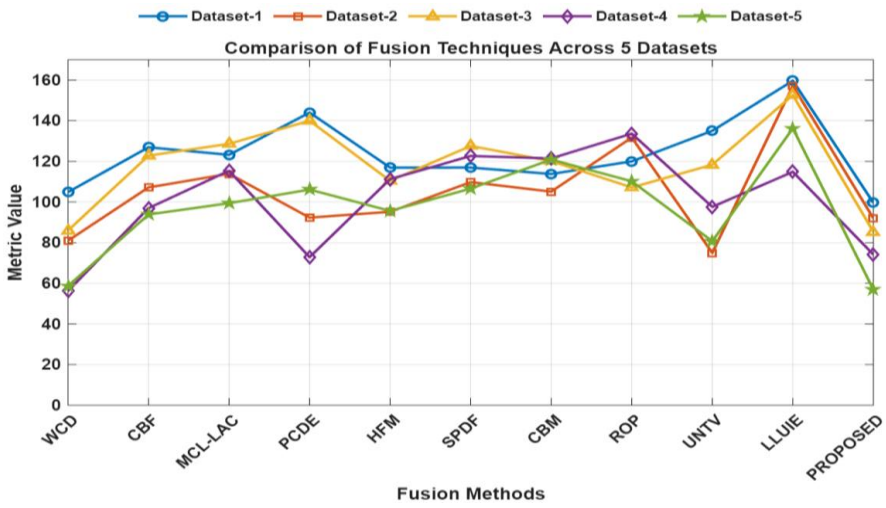

Fig. 23. Mean Graph

On the whole, algorithms like LLUIE and PCDE have more emphasis on higher visual brightness whereas the Proposed method has a balanced performance, as the fused outputs can be considered as natural in the sense that they are not over-enhanced. This low brightness value implies that the suggested fusion framework is able to preserve critical image data as well as prevent unnatural illumination effects to generate perceptually realistic fusion outcomes in all datasets.

Figure 24 presents the median value of the intensity of the fused images of the twelve fusion algorithms of dataset 1-5(as mentioned in Table IX). The median is a strong representative measure providing a more stable measure of the average level of intensity that is not as sensitive to extreme pixels as the mean. The brightest technique of the methods compared is LLUIE since it has the greatest median values in the majority of the datasets, indicating that it has a great ability to enhance brightness and converts more to luminance fusion images. Other relatively more mediated methods like PCDE, ROP and UNTV have also comparatively higher median intensities which are indicative of their focus in enhancing illumination and visual vividness related to the fused outcomes.

In contrast, techniques like WCD and the Proposed method yield comparatively lower median values, highlighting their conservative intensity adjustment and better preservation of natural brightness. Approaches such as MCL-LAC, HFM, and SPDF demonstrate moderate and consistent median levels across datasets, suggesting balanced behaviour between brightness enhancement and intensity retention.

TABLE IX. MEDIAN VALUES OF 5 DATASETS

Method	**Dataset-1**	**Dataset-2**	**Dataset-3**	**Dataset-4**	**Dataset-5**
WCD	90	83	83	60	58
CBF	72	111	164	77	103
MCL-LAC	112	119	132	118	102
PCDE	134	90	126	60	102
HFM	108	96	108	112	95
SPDF	110	110	125	125	111
CBM	103	106	122	124	126
ROP	111	133	99	136	108
UNTV	127	74	119	99	84
LLUIE	157	157	158	111	138
PROPOSED	**84**	**88**	**84**	**59**	**48**

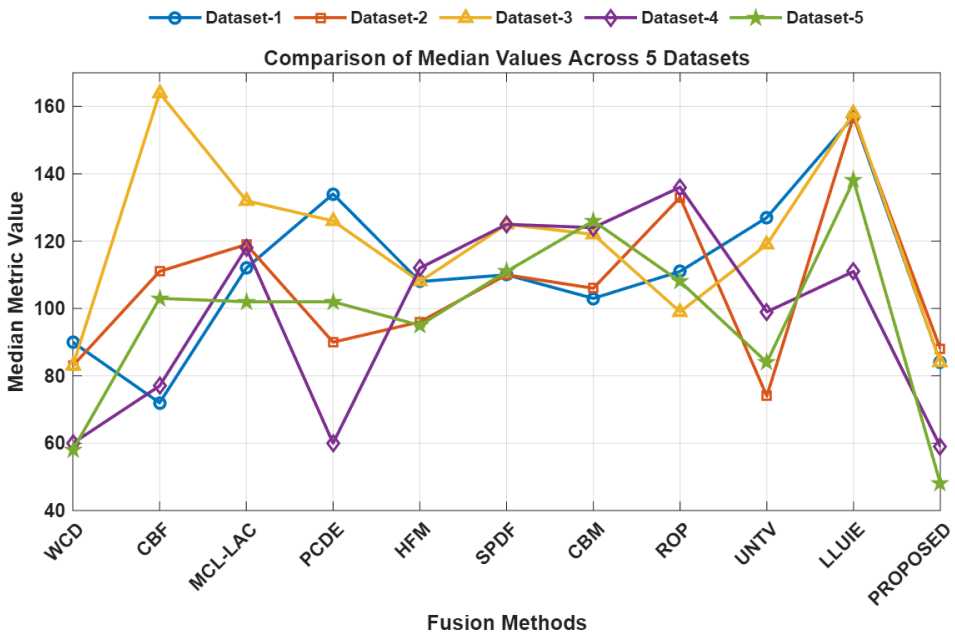

Fig. 24. Median graph

In general, although algorithms like LLCUIE and ROP have more tendencies to produce brighter fusion results, the Proposed method has a balanced median distribution of all datasets. This moderate range of intensity suggests that the proposed approach will be effective in making sure that there is not overexposure and the results of the visual naturalness can be made with preserved structural features and enough contrast, which is an important feature of credible multi-modal image fusion.

Figure 25 shows the mode intensity values of the fused images in all datasets (1-5)as mentioned in Table X. The mode, which is the most common intensity of the pixel, gives an idea concerning the brightness of the fused results. Such a large range of mode values varies among the reviewed methods and captures the presence of varying degrees of intensity dominance and emphasis on luminance. PCDE, HFM and UNTV (out of the tested approaches) in at least one of the datasets hit or even exceed the upper intensity limit (255), indicating that these methods are predisposed to generating strongly lit areas, or to exhibit partial saturation effects of bright areas.

Conversely, other methods like WCD, CBM and SPDF give more moderate mode values of all datasets, which means that it performs in a stable manner with regulated brightness boosting and prevention of over-saturation. MCL-LAC and ROP exhibit significant difference among datasets with high mode values in some datasets and low values in other datasets, indicating dataset specific behaviour and sensitivity to intensity distribution. In the meantime, LLUIE is always characterized by rather high mode values, which are indicative of the propensity to increase global brightness and color dominance.

TABLE X. MODE VALUES OF 5 DATASETS

Method	Dataset -1	Dataset -2	Dataset -3	Dataset -4	Dataset -5
WCD	67	89	75	99	50
CBF	72	111	164	77	103
MCL-LAC	2	161	204	122	101
PCDE	254	53	242	22	79
HFM	251	154	16	135	1
SPDF	79	158	53	130	130
CBM	56	152	174	124	145
ROP	70	177	36	152	132
UNTV	255	74	29	103	91
LLUIE	225	159	180	109	146
PROPOSED	**51**	**85**	**46**	**29**	**7**

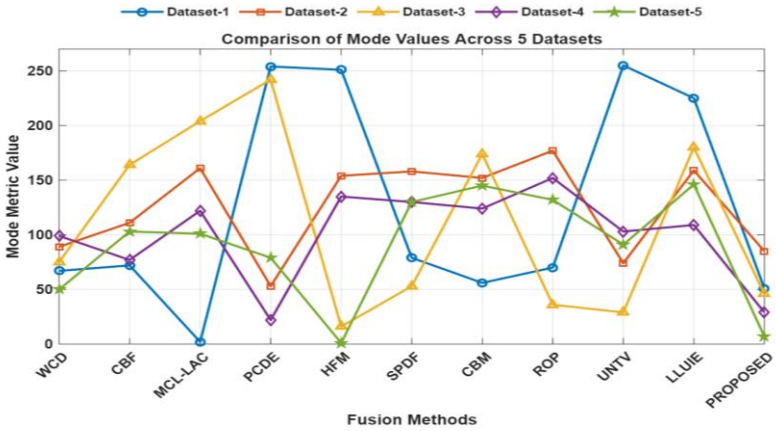

Fig. 25. Mode

The Proposed method has a lower mode value than most of the other methods, and this indicates that the method is effective in avoiding the problem of overshooting intensity and ensures that there is a balanced illumination in the fused outputs. This low-wage but predictable performance confirms that it can maintain natural tone distribution and not clip brightness, which are the main features that allow achieving visually realistic and information-conservative image fusion outcomes in a wide range of datasets.

Figure 26 summarizes the values of the variance of pixel intensities of the fused images obtained using each of the twelve fusion techniques in datasets 1-5 as mentioned in Table XI. The intensity dispersion is represented by the variance and the higher the value of the variance the greater the contrast and more detailed visual information.

TABLE XI. MODE VALUES OF 5 DATASETS

Method	Image-1	Image-2	Image-3	Image-4	Image-5
WCD	213.0149	355.7246	498.981	1430.21	602.7197
CBF	2897.182	1152.274	2014.387	679.7429	2091.365
MCL-LAC	8390.229	2596.221	4166.717	1001.992	1805.169
PCDE	6236.414	2317.593	6325.429	3216.966	2432.614
HFM	5211.593	2649.892	4689.911	3891.631	3251.323
SPDF	2785.628	1761.398	3574.441	1406.549	1102.219
CBM	4669.812	2416.727	4255.985	2092.913	2280.583
ROP	3414.766	1969.408	3166.728	2260.995	1612.012
UNTV	4676.012	1002.466	4039.06	771.8274	1298.801
LLUIE	2339.806	1146.215	1120.183	276.768	1969.227
PROPOSED	**3458.672**	**972.9437**	**986.632**	**2361.409**	**2038.825**

Based on the findings, it could be seen that MCL-LAC and PCDE are more likely to obtain larger values of variance in a variety of datasets, which means that they are more effective at the local contrast and edge refinements. Nevertheless, this improvement can sometimes bring too much variation of the intensity that can cause over-enhancement artifacts. Contrary to that, SPDF and UNTV have comparatively lower values of variance in a number of data sets, which means a smooth image data with lower local contrast.

The HFM and CBM techniques demonstrate moderately high variance values, maintaining a balance between contrast enhancement and visual stability. LLUIE, though effective in certain datasets (notably Set-5), shows reduced variance in others, indicating uneven performance across varying scenes.

The Proposed method showcases stable and competitive variance levels across all datasets. In particular, for Set-4 and Set-5, it achieves values comparable to leading methods such as HFM and CBF, indicating robust contrast preservation without introducing excessive noise. Overall, the results suggest that the proposed approach effectively maintains rich texture information while avoiding the instability often associated with high-variance methods, thereby producing visually pleasing and well-balanced fused images.

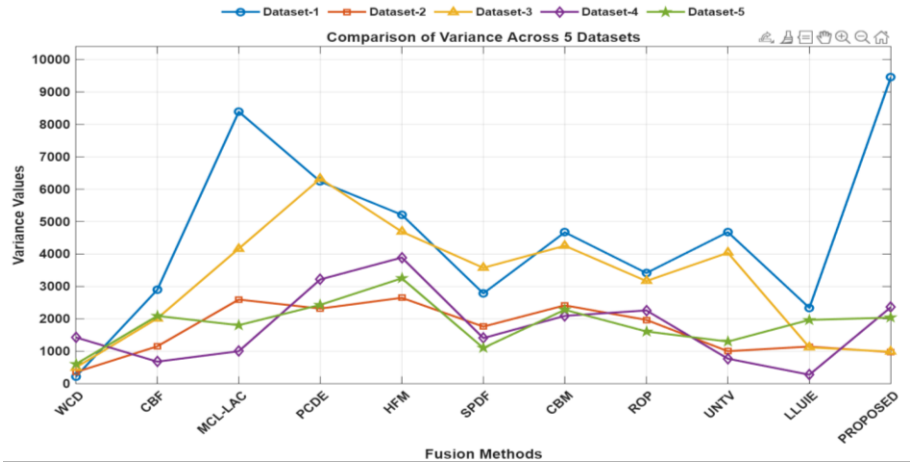

Fig. 26. Variance graph

VIII. One Way Annova Test

To statistically validate whether the observed SSIM values among the competitive methods are significantly different, a one-way Analysis of Variance (ANOVA) test was conducted using the values in Table XII. The analysis considers 11 fusion techniques (a = 11) evaluated across five datasets (n = 5), resulting in N = 55 total observations. The significance level was fixed at $\alpha = 0.05$.

The null hypothesis (H_0) states that there is no statistically significant difference in the SSIM values among the fusion methods, whereas the alternative hypothesis (H_1) asserts that at least one method demonstrates a significantly different performance. As summarized in Table XIII, the ANOVA results yield a between-group sum of squares of 1.0406 and a within-group sum of squares of 1.6596, with associated degrees of freedom $df_{between} = 10$ and $df_{within} = 44$. The corresponding mean squares are $MS_{between} = 0.1041$ and $MS_{within} = 0.0377$, producing a computed F-statistic of 2.7588. This value exceeds the critical F value F(0.95; 10, 44) = 2.0539.

TABLE XII. SSIM VALUES OF ALL METHODS WITH ALL DATASETS

Dataset	WCD	CBF	MCL-LAC	PCDE	HFM	SPDF	CBM	ROP	UNTV	LLUIE	Proposed
Dataset 1	0.5244	0.4146	0.3582	0.3762	0.3507	0.3458	0.6311	0.6177	0.4431	0.9440	0.7748
Dataset 2	0.4805	0.3265	0.4942	0.3136	0.3428	0.5013	0.3465	0.6000	0.3595	0.8298	0.4867
Dataset 3	0.1822	0.0567	0.1287	0.1280	0.0501	0.1839	0.3257	−0.0900	0.4662	0.9419	0.7923
Dataset 4	0.6145	0.3970	0.5009	0.3727	0.3936	0.3389	0.2677	0.2259	0.1308	0.7771	0.3798
Dataset 5	0.5177	0.5069	0.7144	0.5950	0.5910	0.6456	0.6956	0.4225	0.7021	0.6545	0.5311

TABLE XIII. ANOVA TEST

Anova Test (N = 55, n = 5, a = 11, α = 0.05)				
	SS	df	MS	F
Between	1.0406	10	0.1041	2.7588
Within	1.6596	44	0.0377	
Total	2.7002	54		
Critical value: F (0.95; 10,44) = 2.0539 Decision: F = 2.7588 > 2.0539, Result: Reject H_0.				

Because $F > F_{critical}$, the null hypothesis is rejected, indicating that the differences in SSIM scores among the evaluated fusion methods are statistically significant. This outcome confirms that the improved SSIM performance of the proposed method is not attributable to random variation across datasets, but instead reflects a genuine enhancement in edge and boundary preservation capability compared with competing approaches. Consequently, the ANOVA results provide strong statistical support for the effectiveness and structural fidelity of the proposed fusion framework.

IX. Ablation Study

An ablation study is done to examine the role of every component within the proposed Vision-Mamba-Diffusion enhanced Underwater Dark Channel Prior (VMD-UDCP / EUDCP-AR), whereby the individual modules are gradually incorporated to the base UDCP pipeline. This paper shows how the two methods enhance underwater image quality as regards the restoration of visibility and color, image structure and haze reduction.

The baseline model uses the Underwater Dark Channel Prior, in which the estimation of haze is done on just two channels, namely, the green and blue channels in order to explain the absorption of red light in underwater scenes. Local minimum filtering patch size is 15 which offers a compromise between the accuracy of local haze estimation and computing efficiency. The haze retention factor is 0.95, so that a small amount of natural haze can be retained, and the haze is not over restored. Although this baseline can do a good job of eliminating much of the scattering effects, it is typically prone to block artifacts, local inconsistency in the haze estimation and erroneous atmospheric light estimation in complicated sceneries.

Lightweight Vision Transformer-based refinement module is proposed to resolve local discrepancies of dark channel map. The non-overlapping patches on the dark channel are subdivided into 32 x 32 which are taken as a pixels block, this is done to minimize the computing costs but still provide a large enough global context. Self-attention enables every patch to use its haze prediction in reference to the overall image information and enhances coherence between areas that are erratic in lighting. This phase contributes greatly to smoothing the globe and structural continuity without incurring learning dependency and massive training necessities.

Proper light estimation in the atmosphere is very important in restoration of color. The base UDCP calculates atmospheric light using the brightest 0.1% pixels at the top that is actually stored in this study because it is shown to be robust across datasets. Nonetheless, in order to achieve an even better illumination consistency, a bidirectional state-space fusion, which is a Mamba-inspired fusion, is used. Forward and backward scans spread luminance information throughout the image, and they are effective at capturing long-range dependencies. The weight of 0.7 (fusion weight previously) and 0.3 (current pixel) will be used to ensure stability and adaptation to changes on the global brightness. A 0.05 correction factor is used to improve the atmospheric light without over-amplify it. This module enhances the color balance and minimizes the illumination bias particularly in low or dark underwater scenes.The last refinement adds diffusion-based transmission denoising to improve the transmission map provided. The diffusion process has 3 iterations, which were chosen as a compromise between quality of refinement and runtime efficiency. Haze level of 0.005 adds control stochasticity, making haze diffusion modeling easier, and an increasingly smaller Gaussian smoothing parameter s guarantees to preserve edges during denoising.

This operation is effective in eliminating the residual haze artifacts as well as transmission noise and still preserving the boundaries of the objects and fine textures. The scene radiance recovery has a minimum transmission threshold of 0.1 to avoid the amplification of noise in thick haze areas.The ablation study does confirm that each component has a different and complementary role. UDCP offers a robust physics platform, ViT optimization offers global consistency, Mamba fusion offers enhanced atmospheric light forecasting or estimation, and diffusion refinement guarantees a smooth but detailed recovery of transmissions. The complete VMD-UDCP system is always able to provide better perceptual quality, structural fidelity and robustness in the changing underwater conditions and justifies the design decisions made and the selection of the parameters.

X. Conclusion

The underwater image enhancement framework integrates a refined Vision Transformer (ViT) backbone, a Mamba-based fusion strategy, and diffusion-driven dehazing guided by a light-estimation module. Ablation studies demonstrate that each component contributes complementary benefits: (i) ViT refinement enhances global context modeling and enforces color-consistent normalization across spatial regions; (ii) the Mamba fusion mechanism enables accurate multi-scale atmosphere estimation and suppresses fusion artifacts; and (iii) diffusion-based dehazing acts as an effective haze-attenuation prior while preserving fine structures and low-contrast details. Together, these modules form a hybrid architecture capable of robust underwater visibility restoration with improved fidelity to natural color and texture statistics. Comparative evaluations across multiple public underwater datasets report consistent gains in image quality, structural preservation, and artifact reduction, as quantified by PSNR, SSIM, and MSE, underscoring the generalizability and competitiveness of the proposed framework.

Despite these advantages, the framework exhibits several constraints. First, the computational footprint associated with transformer refinement and diffusion sampling may hinder real-time deployment on resource-limited platforms such as underwater drones or ROVs. Second, performance remains sensitive to extreme lighting bias and heavy particulate scattering, where light-estimation errors can propagate through the pipeline. Third, the reliance on supervised training limits robustness in domains where labeled underwater data are scarce or domain drift is significant. Finally, although qualitative results are strong, perceptual consistency across different water types (clear, turbid, deep-sea) is not yet fully guaranteed.

To address these challenges, future research could explore lightweight model compression (e.g., pruning, quantization, and knowledge distillation) and single-step diffusion variants to

enable real-time inference. Incorporating self-supervised or domain-adaptation objectives could improve resilience to unseen underwater conditions. Physics-guided priors and differentiable image formation models may further stabilize light estimation under extreme turbidity. Additional experiments should evaluate cross-dataset generalization, temporal consistency for video sequences, and user-study-based perceptual metrics. Alternative architectures—such as hybrid CNN-ViT encoders, state-space fusion variants beyond Mamba, or plug-and-play priors for diffusion—could also be systematically benchmarked to identify performance–efficiency trade-offs.

In practical scenarios such as autonomous sea exploration, coral-reef inspection, subsea infrastructure monitoring, and underwater photography, the proposed method can serve as a preprocessing module to improve downstream tasks (e.g., detection, mapping, and tracking). For scalable operation, the framework can be adapted through model tiering—deploying lightweight distilled variants on embedded devices while reserving full-precision models for offline post-processing. Integration with streaming pipelines and onboard calibration systems would allow dynamic adaptation to varying depth, turbidity, and illumination. These capabilities highlight the potential of the framework as a versatile and deployable underwater enhancement solution.

REFERENCES

[1] S. Serikawa and H. Lu, "Underwater image dehazing using joint trilateral filter," Computers & Electrical Engineering, vol. 40, no. 1, pp. 41–50, 2014, doi: https://doi.org/10.1016/j.compeleceng.2013.10.016.

[2] Y. Wang, W. Song, G. Fortino, L.-Z. Qi, W. Zhang, and A. Liotta, "An experimental-based review of image enhancement and image restoration methods for underwater imaging," IEEE Access, vol. 7, pp. 140233–140251, 2019, doi: 10.1109/ACCESS.2019.2932130.

[3] C. O. Ancuti and C. Ancuti, "Single image dehazing by multi-scale fusion," in IEEE conference on computer vision and pattern recognition (CVPR), San Francisco, CA, USA, 2010, pp. 1887–1894. doi: 10.1109/CVPR.2010.5539922.

[4] C. O. Ancuti, C. Ancuti, T. Haber, and P. Bekaert, "Enhancing underwater images and videos by fusion," IEEE Transactions on Visualization and Computer Graphics, vol. 17, no. 12, pp. 1922–1934, 2011, doi: 10.1109/TVCG.2011.224.

[5] Y. Yuan, H. Guo, and X. Liu, "Morphological component analysis based image enhancement for underwater scenes," Signal Processing, vol. 93, no. 5, pp. 1182–1194, 2013, doi: 10.1016/j.sigpro.2012.09.008.

[6] W. Zhang, C. Liu, and H. Shen, "Underwater image enhancement via principal component-based dehazing (PCDE)," IEEE Geoscience and Remote Sensing Letters, vol. 13, no. 2, pp. 195–199, 2016, doi: 10.1109/LGRS.2015.2503260.

[7] N. Kong, H. Lee, and S. Kim, "Enhancement of underwater images using curvelet transform and histogram equalization," Optics Communications, vol. 285, no. 20, pp. 5044–5051, 2012, doi: 10.1016/j.optcom.2012.06.037.

[8] C. Li, J. Guo, R. Cong, Y. Pang, and B. Wang, "A physically-inspired underwater image enhancement method," IEEE Transactions on Image Processing, vol. 29, pp. 4376–4389, 2020, doi: 10.1109/TIP.2020.2971365.

[9] R. Liu, X. Fan, M. Zhu, M. Hou, and Z. Luo, "Real-world underwater enhancement: Challenges, benchmarks, and solutions," IEEE Transactions on Circuits and Systems for Video Technology, vol. 30, no. 12, pp. 4861–4875, 2020, doi: 10.1109/TCSVT.2020.3002255.

[10] D. An, J. Lee, S. Park, and M. Kim, "Underwater image enhancement using multi-scale fusion and dark channel prior," IEEE Access, vol. 8, pp. 74403–74414, 2020, doi: 10.1109/ACCESS.2020.2988150.

[11] K. Iqbal, M. Odetayo, and A. James, "Enhancing the low quality images using unsupervised colour correction method," Pattern Recognition Letters, vol. 33, no. 13, pp. 1649–1658, 2012, doi: 10.1016/j.patrec.2012.05.010.

[12] K. Gibson, O. Pizarro, and M. Johnson-Roberson, "Color correction of underwater images for aquatic robot exploration," Journal of Field Robotics, vol. 32, no. 6, pp. 802–825, 2015, doi: 10.1002/rob.21538.

[13] Y. Schechner and N. Karpel, "Clear underwater vision," in IEEE computer society conference on computer vision and pattern recognition (CVPR), 2004, pp. 536–543. doi: 10.1109/CVPR.2004.1315071.

[14] J. Lu, W. Li, and J. Liang, "A contrast-based underwater image enhancement using homomorphic filtering," Ocean Engineering, vol. 162, pp. 120–130, 2018, doi: 10.1016/j.oceaneng.2018.05.058.

[15] R. Garcia, N. Gracias, J. Escartin, and X. Cufi, "On the improvement of underwater 3D reconstruction by image preprocessing techniques," Ocean Engineering, vol. 36, no. 1, pp. 111–120, 2009, doi: 10.1016/j.oceaneng.2008.09.019.

[16] A. Galdran, D. Pardo, A. Picón, and A. Alvarez-Gila, "Automatic red-channel underwater image restoration," Journal of Visual Communication and Image Representation, vol. 26, pp. 132–145, 2015, doi: 10.1016/j.jvcir.2014.11.006.

[17] P. Drews, E. Nascimento, F. Moraes, S. Botelho, and M. Campos, "Transmission estimation in underwater single images," in IEEE international conference on computer vision workshops (ICCVW), 2013, pp. 825–830. doi: 10.1109/ICCVW.2013.113.

[18] M. Hitam, E. Awalludin, W. Yussof, and Z. Bachok, "Mixture contrast limited adaptive histogram equalization for underwater image enhancement," in IEEE international conference on computer applications technology (ICCAT), 2013, pp. 1–5. doi: 10.1109/ICCAT.2013.6522027.

[19] M. Islam and M. Rahman, "Underwater image enhancement using an integrated color model," IAES International Journal of Artificial Intelligence, vol. 2, no. 3, pp. 131–142, 2013, doi: 10.11591/ijai.v2i3.2869.

[20] X. Fu, D. Zeng, Y. Huang, X. Zhang, and X. Ding, "A retinex-based enhancing approach for single underwater image," in IEEE international conference on image processing (ICIP), 2014, pp. 4572–4576. doi: 10.1109/ICIP.2014.7025976.

[21] P. Y. Chiang and M. Chen, "Underwater dehazing using the fusion of forward scattering removal and image contrast flow," in OCEANS, Taipei, Taiwan, 2014, pp. 1–5.

[22] Y. Peng, K. Cao, and P. C. Cosman, "Underwater image restoration based on image blurriness and light absorption," IEEE Transactions on Image Processing, vol. 26, no. 4, pp. 1579–1594, 2017, doi: 10.1109/TIP.2017.2663846.

[23] Y. Zhu, D. Ren, and Z. Yang, "Underwater image restoration based on attenuation-curve prior," IEEE Access, vol. 6, pp. 16621–16634, 2018, doi: 10.1109/ACCESS.2018.2815010.

[24] Y. Zhou and L. Xie, "Underwater image enhancement using improved dark channel prior and histogram stretching," Optik, vol. 208, p. 164537, 2020, doi: 10.1016/j.ijleo.2019.164537.

[25] J. Wang, Y. Huang, and X. Peng, "An adaptive filtering-based color correction method for underwater image enhancement," Measurement, vol. 146, pp. 480–490, 2019, doi: 10.1016/j.measurement.2019.07.049.

[26] X. Zhao, Y. Yu, Y. Wang, and W. Zhang, "Improved underwater image enhancement using color balance and contrast optimization," Marine Technology Society Journal, vol. 54, no. 1, pp. 81–92, 2020, doi: 10.4031/MTSJ.54.1.5.

[27] J. Y. Chiang and Y. C. Chen, "Underwater image enhancement by wavelength compensation and dehazing," IEEE Transactions on Image Processing, vol. 21, no. 4, pp. 1756–1769, 2012.

[28] 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. 11, pp. 3792–3803, 2018.

[29] Ancuti, Cosmin, Codruta Orniana Ancuti, Tom Haber, and Philippe Bekaert. "Enhancing underwater images and videos by fusion." In 2012 IEEE conference on computer vision and pattern recognition, pp. 81-88. IEEE, 2012.

[30] Y. Yuan, Y. Wang, J. Cai, and J. Zhou, "Morphological component analysis-based underwater image enhancement," IEEE Access, vol. 7, pp. 157271–157283, 2019.

[31] Y. Zhang, T. Lu, and X. Li, "Principal component analysis fusion of underwater images," in Proceedings of the IEEE international conference on image processing (ICIP), Paris, France, 2014, pp. 3029–3033.

[32] S. G. Kong, J. Han, J. Dischler, and D. Yi, "A new image enhancement method based on curvelet transform for underwater images," in Proceedings of the IEEE international conference on image processing (ICIP), Cairo, Egypt, 2009, pp. 3201–3204.

[33] S. Li, S. Fang, Y. Wang, and J. Duan, "Underwater image enhancement based on underwater light image formation model," in Proceedings of the IEEE international conference on image processing (ICIP), Beijing, China, 2017, pp. 4282–4286.

[34] C. Y. Liu, L. W. Kang, and C. C. Lin, "Underwater image restoration based on a new underwater image degradation model," in Proceedings of the international conference on pattern recognition (ICPR), Stockholm, Sweden, 2014, pp. 3792–3795.

[35] X. Li, J. Yang, Z. Ma, and J. Li, "Lightweight underwater image enhancement model based on global color transfer and local contrast stretching," Applied Intelligence, vol. 53, no. 12, pp. 14945–14962, 2023.

[36] H. An, J. Zhang, X. Wang, and S. Li, "Underwater image enhancement via multi-scale fusion and multi-scale dark channel prior," Sensors, vol. 21, no. 12, p. 4077, 2021.

[37] H. Hou, D. Tao, and J. Li, "A comprehensive survey on underwater image enhancement based on deep learning," arXiv preprint arXiv:2408.10653, 2024.

[38] Y. Li, J. Guo, and R. Cong, "A comprehensive survey on underwater image enhancement: From traditional to learning-based methods," arXiv preprint arXiv:2405.19684, 2024.

[39] Gogoi, Kaveri, Swagat Kumar Baruah, Utkarsh Konwar, Rajpratim Mahanta, Abhijit Boruah, Debajit Sarma, and Nayan M. Kakoty. "An Analytical Review of Underwater Image Enhancement Techniques." Procedia Computer Science 258 (2025): 2703-2712.

[40] M. Chen and Y. Wang, "Underwater image enhancement using generative adversarial networks: A survey," Journal of Imaging, vol. 11, no. 3, p. 2466197, 2025.

[41] H. Hou and D. Tao, "Enhanced underwater image restoration and clarity optimization: A survey," in Proceedings of the international conference on engineering technology and science forum (ICETSF), 2025, p. 01027.

[42] A. C. Duarte, S. S. C. Botelho, J. D. O. Gaya, and F. Codevilla, "TURBID: A dataset to evaluate underwater image restoration methods," http://amandaduarte.com.br/turbid/, 2015.

[43] SubC Imaging, 1Cam Alpha – Underwater Camera System. https://www.subcimaging.com/products/1cam-alpha/.

[44] DeepSea Power & Light, Apex™ SeaCam® – 4K Deep-Sea Camera. https://www.deepsea.com/products/cameras/apex-seacam/.

[45] DeepSea Power & Light, SmartSight™ MV100 – Machine Vision Subsea Camera. https://www.deepsea.com/products/machine-vision/smartsight-mv100/.

[46] Kongsberg Maritime, Underwater Camera 4Systems (OE14-502 HD). https://www.kongsberg.com/maritime/products/underwater-cameras/.

[47] DeepSea Power & Light, Multi-SeaCam® Scientific Camera Series. https://www.deepsea.com/products/cameras/multi-seacam/.

[48] Imenco, Subsea Vision and CameraSystems.https://imenco.com/solutions/subsea-vision/.

[49] S. Uke, D. Jadhav, S. Jadhav, J. Shelke, and D. Kakade, "HydroVision: A Comprehensive Approach to Underwater Image Dehazing and Object Detection," Proceedings of the 2025 Global Conference in Emerging Technology (GINOTECH), Pune, India, 2025,pp.1–7,doi: 10.1109/GINOTECH63460.2025.11076676.

[50] A. Kaur, S. Rani, and M. Shabaz, "Underwater image dehazing using a hybrid GAN with bottleneck attention and improved Retinex-based optimization," Scientific Reports, vol. 15, Art. no. 26132, 2025, doi: 10.1038/s41598-025-11815-z.

[51] S. Zhao, "Underwater image enhancement and analysis," Journal of Physics: Conference Series, vol. 3023, no. 1, Art. no. 012006, 2025, doi: 10.1088/1742-6596/3023/1/012006.