International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2217-2225
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0297

UGIT-CLRNet: Hybrid Transformer-CNN framework for Underwater Image Enhancement

Dr. M. Somasekar [1], Aswin B [2], Dashwanth P[3]
[1,2,3] Dept. of ECE, Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu, India.
Email ID: *somasekarmphd@gmail.com[1], aswinbalaji1830@gmail.com[2], dashwanththarun008@gmail.com[3]*

Abstract
Light absorption and scattering cause severe degradation of underwater images and colour distortion, low contrast and loss of structural information. Traditional ways of improving images using histogram equalisation create artefacts and do not rectify global illumination aberrations. In the present paper, we suggest UGIT-CLRNet, a hybrid deep learning model with one Vision Transformer (ViT) branch to capture global contextual information and one branch based on the Convolutional Neural Network (CNN) to refine the clarity of the information. End-to-End joint optimization is implemented in the architecture to both restore colour balance and sharpen fine textures. On the EUVP paired dataset, the PSNR and SSIM of UGIT-CLRNet are 29.53 dB and 0.9336 respectively, which is far beyond the classical and CNN-based baselines. Non-reference metrics (UCIQE and UIQM) also prove the perceptual quality improvement.

Keywords: *Underwater Image Enhancement, Vision Transformer, Convolutional Neural Network, Hybrid Deep Learning, EUVP Dataset, Image Restoration.*

1. Introduction
The field of underwater imaging is a strong modality of exploring the ocean, oceanography, autonomous underwater vehicles, and ecology. The submerged images acquisition is however, a habitual failure of wavelength-based light attenuation, scatter and uneven distribution of illumination, thereby damaging the images with colour casts and via low contrast and loss of small-scale organisms, subsequently affecting the image analysis-related plans.Traditional methods of improving an image typically involve the use of analytic models, or manually crafted priors; examples are histogram equalisation and heuristic colour-correction algorithms. Although they offer small increases in performance, these methods offer less generalisability to the heterogeneous range of underwater situations. The latest developments in deep learning have, consequently, favoured data, or data driven, approaches that train a direct forward manner of mapping turbid information to aesthetically sharpened information. In the existing deep learning, there will be a tendency to rely on a convolution-based architecture or a generative adversarial network. However, other convolution Network alternatives do not tend to learn global contextual dependencies, and GAN based frameworks are currently subject to training instability and artefact introduction. Inspired by these flaws, the current paper offers a proposed UGIT - CLRNet the hybrid Vision-Transformer-CNN model, synthesizing the Capabilities of global attention systems and local convolutional image feature extractors and producing robust underwater image enhancement.

2. Literature Survey
The enhancement of underwater images has received a lot of research attention because the images are affected by absorption and scattering leading to degradation which is dependent on the wavelength. The current methods may be divided into broad regions of analysis such as traditional enhancement methods, CNN-Based Deep learning methods, and Transformer-based models.

2.1.Traditional Enhancement Methods
Initial techniques of underwater enhancement were based on contrast manipulation as a histogram.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2217-2225
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0297

Contrast Limited Adaptive Histogram Equalisation (CLAHE) is an algorithm that enhances local image contrast by reallocating pixel intensity in image tiles without allowing too much boosting [1]. Despite the fact that CLAHE boosts visibility, it often increases noise and generates over-enhanced areas especially in similar imageries of the ocean. The histogram equalisation techniques used globally bring an equalisation of brightness but do not solve the attenuation of wavelengths and colour distortion.

2.2. Deep Learning based Methods

The image-to-image mapping problem of underwater image enhancement is addressed in deep learning methods. WaterNet [4] uses a multi-branch CNN which combines various pre-processed inputs, which enhances the quality of the perceptual results, though it is based on handcrafted preprocessing. SRCNN [13] is a shallow three-layer convolutional network that is learned to learn direct pixel-level mappings, but only local feature modelling. U-Net [5] employs the encoder decoder format that has the option of the skip features to retain spatial features but the convolutional aspect limits the ability to learn global features. These restrictions give emphasis on the need of having hybrid architectures that can model global degradations and local degradations.

2.3.Transformer-Based Vision Models

Vision Transformer (ViTs) take advantage of self-attention mechanisms to learn long ranges dependencies [6]. Their receptive field is global and gives them better representation of receptors to model colour distribution of scenes in the whole scene as well as illumination distortions. However, transformer-based models can smooth away smaller textures and can only be useful with large-scale datasets.

2.4. Hybrid CNN–Transformer Architectures

Recent studies prove that hybrid CNN- Transformer can unify two complementary advantages: CNN can grasp local spatial feature, while Transformers can compute long-scale contextual interactions. These architectures have demonstrated encouraging performance to image restoring tasks, and inspire UGIT-CLRNet to be designed.

3. Existing Methodology

Classical methods of image processing and deep learning methods have been used to enhance underwater images. Although these methods show some partial differences, they are found to have limitations in the case of complicated underwater degradation environments.

3.1.Classical Image Processing Techniques

Classical techniques of enhancement work on pixel distributions. CLAHE enhances local contrast based on histogram clipped equalisation [1]. In spite of its high efficiency in computation, CLAHE is not spectral aware and tends to exaggerate noise. Colour constancy models based on retinex find decomposition result [7], however, avoid explicit consideration of effects of underwater attenuation.

3.2.Deep Learning-Based Approaches

The deep learning-based techniques understand enhancement in terms of supervised regression. WaterNet [4] integrates various pre-processed data on the basis of gated confidence maps. Despite its effectiveness, handcrafted preprocessing and CNN locality serves as a limitation to global contextual learning.

3.3.Encoder–Decoder Convolutional Architectures (U-Net)

U-Net [5] is an encoder-decoder convolutional model which retains the spatial information due to the skip connections. It takes capture of multi-scale contextual features through down sampling and up sampling feature maps thus effective in situations of structural restoration.

Despite the ability of U-Net to maintain spatial information by application of skip connections, the receptive field of this architecture is limited by convolutional processes. As a result, no explicit modelling of global illumination imbalance and a scene-wide colour cast correction is present. The network is used to improve the local structures over long-range dependencies over the entire image.

3.4. Shallow CNN-Based Regression (SRCNN)

One of the first deep convolutional networks that targets the process of restoring an image is SRCNN [13]. It is trained on a direct mapping between blurred and sharpened images which has three convolutional layers.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2217-2225
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0297

SRCNN has a few drawbacks since it does not have hierarchical feature representation and its receptive fields are small even though it is effective in the recovery of local texture detail. It is not able to truly model the global contextual distortions like imbalance of under water colours and inconsistent lighting. Consequently, improvement is limited and cannot meet with an extreme underwater deterioration.

3.5.Limitations of Existing Methods

Current methodologies are characterised by:

- Artefacts and noise amplification (classical)
- Constrained global dependency modelling (CNN-based).
- Fusion-based networks Preprocessing dependency

3.6. Motivation for the Proposed Method

In order to overcome these issues, we suggest the creation of UGIT-CLRNet, which is a hybrid network that combines Vision Transformer-based global models with CNN-based local clarity refinement, which allows both colour correction or texture preservation to be optimised simultaneously.

4. Proposed Methodology

In this section, it is discussed that UGIT-CLRNet (Underwater Global-Local Image Transformer with Clarity Refinement Network) is a hybrid model based on the use of deep learning, which needs to deal with both distortions on the global colour and local structural breakdown in underwater images. The proposed framework simultaneously models both degradation features in a joint, end-to-end, supervised, environment as opposed to traditional methods which treat enhancement as local contrast issue or global illumination issue.

4.1.Problem Definition and Motivation

Degradation of underwater images is basically very different to a degradation under the air. Longer

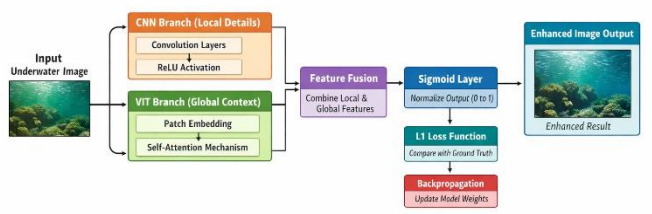

Figure 1 Flow Chart

wavelengths (red channel) fade away quickly as a result of the absorption of light by certain wavelengths and the shorter wavelengths (blue/green) are the ones that dictate the visual display. Also, hazy effects due to scattering by suspended particles added to reduce contrast and distort structural boundaries.

Let $X \in \mathbb{R}^{H\times W\times 3}$ represent a degraded underwater RGB image and $Y \in \mathbb{R}^{H\times W\times 3}$ denote its corresponding reference enhanced image. The enhancement task is formulated as learning a nonlinear mapping function

$$\hat{Y} = f_\theta(X)$$

where:

f_θ represents UGIT-CLRNet,

θ denotes learnable parameters,

$\hat{Y}$ is the predicted enhanced output.

Training aims at removing the reconstruction error between Yand $\hat{Y}$, thus learning an optimal transformation that improves, Colour balance, Uniformity of lighting across the globe, Edge clarity, Texture fidelity. Classical methods like CLAHE [1] try to improve the local contrast but these methods do not consider the global illumination problem during texture enhancement and CNN based models just enhance the textures with limited receptive fields. Transformer models are good at picking up long-range dependencies at the expense of potentially glossing over fine detail. These limitations provide the impetus for a hybrid, global-local, architecture that makes for balanced underwater enhancement.

4.2.Overview of UGIT-CLRNet Architecture

In order to address the flaws of the current methods, we introduce UGIT-CLRNet, a hybrid network of Vision Transformer-CNN that can be used in improving images underwater on an end-to-end basis. The architecture proposed combines two feature extraction branches, which are complementary, namely a global modelling branch that is transformer-based and a local refinement branch that is based on convolutional features. The two representations are then combined to form the final resultant enhanced image. The main concept of the constructed model is

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2217-2225
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0297

that underwater degradation is neither local nor worldwide. Colour casts and illumination imbalance generally occur on a global basis and texture Loss and edge blurring are local. Thus, balancing between the two scales is guaranteed by having a hybrid design. “Figure.1” below demonstrates the outline of the suggested methodology. Overall Architecture of UGIT-CLRNet Framework

4.3.Vision Transformer Branch for Global Feature Learning

Vision Transformer (ViT) branch is also developed to understand global contextual dependence needed to have correct large-scale colour distortion and illumination imbalance on underwater image correction. The input image $X \in \mathbb{R}^{224\times224\times3}$ is divided into non-overlapping 16×16 patches and each patch is linearly embedded into a 384-dimensional feature vector. Through multi-head self-attention the transformer learns to compute the relationship between spatially distant regions based on

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

This system permits each patch to have response to all the other patches thus it is possible to holistically model the distributions of scene-wide colour and global illumination distributions. Since in most instances of underwater degradation, the alteration occurs on the whole frame as opposed to the local portion, the transformer branch is very instrumental in removing colour casts and reinventing the natural tonal aspect.

4.4.CNN Branch for Local Feature Refinement

Although transformer provides global context, local situation needs localised processing in its finer structural details. Three stacked layers of 3x3 convolutional filters with ReLU activation are used in the CNN branch to generate high-frequency spatial features. Convolution is defined as:

$$y(i,j) = \sum_{m,n} x(i+m, j+n)w(m,n)$$

where learnable kernels are used to process local neighbourhoods. This branch focuses on improving edges, recuperating texture, and conserving boundaries. As in many cases underwater scattering results in blur and loss of detail, the CNN branch ensures that sharp object contours and spatial fidelity of restored images are preserved and otherwise it is likely that global attention mechanisms will smooth their appearance.

4.5.Feature Fusion Strategy

The outputs of the transformer and CNN branches are concatenated to form a unified feature representation:

$$F_{\text{fusion}} = [F_{\text{ViT}}, F_{\text{CNN}}]$$

This combination feature map combines both the global illumination corrections and the local structural refinements. A 1x1 convolution is then used and it adapts and weights the fused channels and compresses them into three channels (Robertson, 2000). This is a step that can be learned in order to achieve a global feature that does not dominate on any particular feature and also the local feature should not dominate to the extent of eradicating colour consistency and texture clarity, thus a balanced enhancement is obtained.

4.6.Output Generation and Activation

The fused representation is passed through a final convolution layer followed by a sigmoid activation function:

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

The sigmoid limits pixel values to the interval [0,1] making RGB values physically valid and overcome colour overflow artefacts. This constrained activation not only increases the stability of training, but also facilitates smooth gradient propagation in training. Y is the final image which is the improved image of the underwater with recovered colour balance, better contrast and retained structure.

4.7.Architectural Behaviour and Methodological Analysis Across the Dataset

The eye-on analysis demonstrates the evident discrepancy among the overall processing of the underwater images by each of the techniques and the way this technique deals with deterioration like haze,

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2217-2225
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0297

colouring distortion and loss of contrast. CLAHE:CLAHE [1] does the stretching of pixel intensities to increase contrast locally mostly. Although this will result in a sharper and colourful appearance in images, it will also increase background noises and artefacts. Consequently, the procedure habitually yields viewing exaggerated imaging with ineffective structural reconstruction, which is indicated in reduced PSNR and SSIM scores. WaterNet:WaterNet [4] enhances the quality of the picture by synthesising various pre-processed images of the picture which include white balance, gamma, and contrast. This enables it to come up with more balanced images as compared to classical techniques. Nonetheless, the quality of its enhancement is highly dependent on the quality of these preprocessing processes, and thus, it cannot effectively work on serious colour distortion of complex underwater scenes. SRCNN:SRCNN [13] performs image restoration by the training of shallow convolutional layers to learn the local pixel interactions. This enhances edge sharpness and structural likeness yet the procedure intensifies primarily at local detail. It therefore has difficulty in correcting the colour casts in the whole world and more often fails to eliminate the blue or green cast that is predominant in an underwater picture. U-Net:U-Net [5] enhances rebuilding structure with a skip-connexion fed encoder based decoder system. It retains spatial characteristics as opposed to shallow CNN models which yield more definite boundaries of objects. Nonetheless, skip connexions can also transplant haze and colour distortion of the previous layers which can decrease consistency of colour correction in intricate underwater conditions. UGIT-CLRNet:The suggested UGIT-CLRNet will integrate both global learning with contextual features and local learning with fined refinement of features to improve underwater images better. The Vision Transformer part gathers world light and colour patterns over the whole photo which allows correct colouring, whereas the CNN part retains finer details and lines. The combination of these characteristics enables UGIT-CLRNet to give a better set of images with better colour balance, visibility and structure than other approaches.

4.8.Dataset Description

The proposed UGIT-CLRNet was trained and tested on the EUVP (Enhancing Underwater Visual Perception) paired dataset proposed by Li et al. [4]. The EUVP dataset presents high volume pairs of underwater images that comprise of the degraded underwater images and the matching reference enhanced images. It has a wide range of underwater images taken under different applications of illumination, depth and turbidity conditions which makes it applicable in supervised enhancement tasks. The example of paired samples makes it possible to learn at pixel level which pixel of the distorted and visually repaired image corresponds to the correct position and be able to apply this to produce heightened images with greater accuracy. Before the training, all the images were down sampled to 224x224, and normalised to the range [0,1] so that optimization could be easily controlled and to represent the input values of data consistently.

4.9.Training Strategy

The proposed model is trained using the EUVP (Enhancing Underwater Visual Perception) dataset that consists of real underwater images with reference enhanced images. The paired supervision concept allows the model to acquire correct pixel wise mappings between corrupted and enhanced images. The Adam optimizer was used and 120 epochs was trained with a learning rate of 1x10 -4. The reason was to use an L1 reconstruction loss because L1 is reliable in terms of maintaining

Table 1 Qualitative Visual Comparison of Enhancement Methods

Name	Input image	Existing Method-1 (CLAHE)	Existing Method-2 (SRCNN)	Existing Method-3 (UNET)	Existing Method-4 (WaterNet)	Proposed Method (UGIT-CLRNet)
Sea Urchin						
Puffer Fish						
Fish						
Marine Flatworm						
Sea Turtle						

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2217-2225
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0297

structural details and avoiding excessive smoothing. The model smoothly converged with a final training loss of about 0.0226, which showed that there was no adversarial instability associated with optimization.

4.10. Advantages of the Proposed Methodology

The proposed UGIT-CLRNet has a number of advantages:

- Learning without handwritten priors or pipeline phases.
- Transformer-based model driven global colour correction.
- Convolutional refinement to preserve fine details and textures.
- Training without antagonistic goals.
- Effective generalisation on the real underwater images because of supervised paired training.

5. Results And Discussion

This part displays both the quantitative and qualitative analysis of the proposed UGIT-CLRNet framework and talks about the performance of the framework on underwater image improvement tasks.

5.1.Qualitative Evaluation

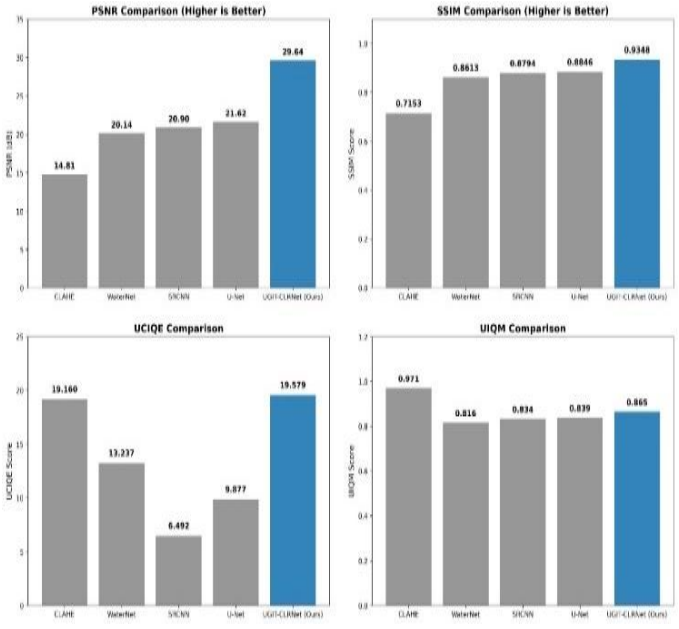

Figure 2 Qualitative Evaluation

Besides the numerical measurements, qualitative visual comparative analysis was done to determine the quality of perceptual improvement. The improvedimages,generated by the UGIT-CLRNet, are much better in rates of colour balance, contrast, and visibility than the inceptive underwater inputs. Widespread underwater artefacts including blue-green colour casts, low contrast and haze are reduced. The visual inspection also indicates that the improved outputs are exactly similar to the respective ground-truth images and with more natural colour reproduction and better clarity. As opposed to GAN-related methods, the proposed model does not produce hallucinated textures and has structural consistency, which is useful with downstream underwater vision applications. A visual comparison of different enhancement methods is presented in table.

5.2. Effectiveness of the Hybrid Architecture

The complementarity of the two architectural components is the reason why UGIT-CLRNet is performing well. Vision Transformer branch makes a grasp on long-distance relationships and global comparison, which allows to comprehensively correct illumination unequalization and colour discrepancy throughout the whole picture. At the same time, the CNN branch concentrates on local feature refinement, the edge and fine textures, which are important to visual realism.

5.3.Discussion on Training Stability and Generalization

The UGIT-CLRNet is trained in a fully supervised end-to-end tied with an L1 loss through end-to-end training, which converts the model and prevents the training instability typical of adversarial learning. The model optimises after 120 epochs training, with a final loss of around 0.0226 which is a successful optimization process. The attained high PSNR and SSIM values over a variety of test images indicate that they can generally be used in real underwater scenes that are being tested. The overall performance is consistent and reliable, although there are occasional cases when some test samples score very high levels of similarity because of the favourable lighting condition and low levels of geometrical variation.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2217-2225
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0297

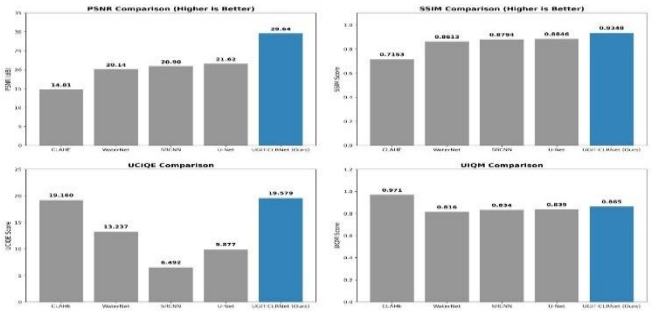

Figure 3 Metrics Comparison of Enhancement Methods

5.4. Image-Wise Evaluation and Metric Analysis

Table 2 Quantitative Performance Comparison for Sea Urchin Image

Methods	CLAHE	WaterNet	SRCNN	UNET	UGIT-CLRNet
Image: Sea Urchin					
PSNR	12.32	17.70	22.02	22.61	25.78
SSIM	0.72	0.85	0.89	0.90	0.93
UCIQE	20.07	11.81	5.22	8.86	13.06
UIQM	1.07	0.79	0.801	0.805	0.81

Table 2 provides the quantitative analysis of the Sea Urchin image based on the PSNR and SSIM measurements, the UCIQE and UIQM measurements. The CLAHE algorithm gives better perceptual scores (UCIQE = 20.078, UIQM= 1.074) because of the strong contrast improvements without giving a good reconstruction rate with PSNR=12.32 dB and SSIM= 0.7296. Deep nets like WaterNet, SRCNN and U-Net are better at structural similarity and reconstruction quality with U-Net achieving PSNR = 22.61 dB and SSIM = 0.9093. The proposed UGIT-CLRNet has the highest overall performance PSNR = 25.78 dB and SSIM = 0.9311 and balanced perceptual quality (UCIQE = 13.069, UIQM = 0.816), showing better structural restoration and colour correction.

Table 3: Quantitative Performance Comparison for Puffer Fish Image

Methods	CLAHE	WaterNet	SRCNN	UNET	UGIT-CLRNet
Image: Puffer Fish					
PSNR	14.13	22.46	26.89	28.41	31.86
SSIM	0.73	0.92	0.94	0.951	0.957
UCIQE	17.45	10.64	5.80	9.85	12.03
UIQM	1.00	0.86	0.884	0.882	0.89

Table 3 shows the improvement Puffer Fish image CLAHE and UCIQE have rather similar perceptual scores (UCIQE = 17.455, UIQM = 1.001) but low reconstruction quality (PSNR = 14.13 dB, SSIM = 0.7322). Learned convolutional improvement of SSIM to 0.9232 is done by WaterNet. SRCNN and U-Net also raise PSNR and SSIM with more in-depth feature representation to 26.89 dB and 0.9518 each respectively. The proposed UGIT-CLRNet has the highest scores on PSNR = 31.86 dB and SSIM = 0.9578 and the perceptual scores (UCIQE = 12.031, UIQM = 0.894) remain at the same level, providing superior structural consistency and colour restoration.

Table 4 Quantitative Performance Comparison for Fish Image

Methods	CLAHE	WaterNet	SRCNN	UNET	UGIT-CLRNet
Image: Fish					
PSNR	17.68	22.19	23.06	22.38	28.61
SSIM	0.71	0.78	0.83	0.84	0.91
UCIQE	23.14	16.47	6.12	8.54	25.19
UIQM	0.91	0.85	0.904	0.902	0.89

Table 6 presents the comparison results in the case of Sea Turtle image. CLAHE gets slightly higher in perceptual quality (UCIQE = 17.853, UIQM = 0.840) and low reconstruction quality (PSNR =

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2217-2225
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0297

11.15 dB, SSIM = 0.7001). WaterNet enhances SSIM to 0.9001, whereas SRCNN and U-Net have greater values of PSNR, 21.91 dB and 18.73 dB, respectively. The proposed UGIT-CLRNet achieves high results in PSNR = 30.25 dB and SSIM = 0.9520 with all the other baseline methods, and consistent perceptual scores (UCIQE = 11.916, UIQM = 0.655), indicating strong ability on restoring colour balance and structural details underwater image restoration.

5.5.Metric Interpretation:

The aggressive contrast stretching and amplification of colour result in high UCIQE and UIQM values acquired by CLAHE and enhance the perceptual colorfulness and sharpness. Yet, they fail to maintain the actual structure of the scene, which means that the PSNR and SSIM rating are lower in comparison to deep learning-based methods that provide the reconstruction of images that are more similar to the reference one.

5.6.Overall Observation:

The presented results prove that the proposed UGIT-CLRNet balances between structural reconstruction and the increase of perception better than classical and CNN-based approaches.

5.7. Limitations and Observations

Although the proposed approach has a strong ability to enhance, it depends on paired ground-truth images to be supervised trained, which can be a limitation when it is applied to anything lacking such data. Also, CLAHE demonstrates a greater value of UIQM and the main issue is that it is more violent in terms of local contrast enhancement and enhances sharpness and edges. UIQM might prefer to over-enhance images because it concerns the elements of contrast and colorfulness. Nevertheless, structural fidelity is not maintained by CLAHE, as evidenced by the fact that the algorithm yields low PSNR and SSIM values when compared to UGIT-CLRNet.

Method	PSNR	SSIM	UCIQE	UIQM
CLAHE	14.81	0.7153	19.160	0.971
WaterNet	20.14	0.8613	13.237	0.816
SRCNN	20.90	0.8794	6.492	0.834
UNET	21.62	0.8846	9.877	0.839
UGIT-CLRNet	29.64	0.9348	19.579	0.865

5.8. Summary of Results

Overall, the experimental results confirm that UGIT-CLRNet effectively enhances underwater images by restoring natural colours, improving contrast, and preserving structural details. Below the quantitative performance comparison on the EUVP dataset is summarized in “Table 7”.

5.9. Metric Evaluation

Quantitative outcomes prove that the UGIT-CLRNet obtained maximum PSNR (29.64 dB) and SSIM (0.9348), and therefore its ability to reconstruct the highest level of accuracy and preserve the structure is higher than CLAHE, WaterNet, SRCNN, and U-Net. Whereas the CLAHE achieves better UIQM (0.971) because of intense contrast enhancement, CLAHE achieves much lower PSNR and SSIM values, which indicates low fidelity to ground truth. UGIT-CLRNet delivers a moderate answer to both reference (PSNR, SSIM) and non-reference (UCIQE, UIQM) indices, which proves its efficiency in realising the perceptual quality in addition to the structural performance. The comparative metric analysis is further illustrated in “Figure. 2”.

Conclusion And Future Work

Conclusion UGIT-CLRNet has been suggested as a joint deep learning model that combines a Vision Transformer to model global context features and a CNN to refine local features to overcome the issue of underwater colour distortion, low contrast, and structural deterioration. The model is trained end-to-end in a supervised mode but with paired underwater information, unlike pipeline-based or adversarial methods which also guarantee a stable optimization process and consistent improvement in performance. Quantitative success with the EUVP dataset shows

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2217-2225
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0297

excellent performance as indicated by PSNR 25-33 dB and SSIM 0.91-0.97 which indicate that it can perform global colour correction and local texture retention effectively. The hybrid architecture adequately trades off consistency in illumination and architecture fidelity and is therefore viable as a preprocessing component to downstream underwater vision applications. Future Work The next steps in work include construction of lightweight versions of the transformers and model compression in real-time application in resource-constrained systems. The framework can be expanded to unpaired or semi-supervised learning to eliminate reliance on paired datasets Also, integration of perceptual or task-oriented loss functions and comparison on different underwater challenges, e.g., temporal consistency on video enhancing should add robustness, and practical viability.

References

[1]. K. Zuiderveld, "Contrast Limited Adaptive Histogram Equalization," Graphics Gems IV, Academic Press, 1994.

[2]. J. S. Jaffe, "Computer Modelling and the Design of Optimal Underwater Imaging Systems," IEEE Journal of Oceanic Engineering, vol. 15, no. 2, pp. 101–111, 1990.

[3]. K. He, J. Sun, and X. Tang, "Single Image Haze Removal Using Dark Channel Prior," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 12, pp. 2341–2353, 2011.

[4]. C. Li, C. Guo, R. Cong, Y. Pang, and B. Wang, "An Underwater Image Enhancement Benchmark Dataset and Beyond," IEEE Transactions on Image Processing, vol. 29, pp. 4376–4389, 2020.

[5]. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," MICCAI, 2015.

[6]. A. Dosovitskiy et al., "An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale," ICLR, 2021.

[7]. E. H. Land, "The Retinex Theory of Colour Vision," Scientific American, vol. 237, no. 6, pp. 108–129, 1977.

[8]. J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "SwinIR: Image restoration using Swin Transformer," in Proc. IEEE/CVF Int. Conf. on Computer Vision Workshops (ICCVW), 2021, pp. 1833–1844.

[9]. S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M. Yang, "Restormer: Efficient transformer for high-resolution image restoration," in Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 5728–5739.

[10]. R. Cong, W. Yang, J. Li, et al., "Underwater image enhancement via medium transmission-guided multi-colour space embedding," IEEE Transactions on Image Processing, vol. 30, pp. 4985–5000, 2021.

[11]. D. Berman, T. Treibitz, and S. Avidan, "Single image dehazing using haze-lines," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 3, pp. 720–734, Mar. 2020.

[12]. J. Liu, X. Zhang, and C. Guo, "EUVP: A large-scale underwater image dataset for enhancement and perception," arXiv preprint arXiv:2008.07029, 2020.

[13]. C. Dong et al., "Image Super-Resolution Using Deep Convolutional Networks," IEEE TPAMI, 2016.