Multimedia Tools and Applications (2024) 83:75603–75625
https://doi.org/10.1007/s11042-024-18550-z

Underwater image enhancement using lightweight vision transformer

Muneeba Daud[1] · **Hammad Afzal**[1] · **Khawir Mahmood**[1]

Received: 23 February 2023 / Revised: 20 September 2023 / Accepted: 22 January 2024 /
Published online: 19 February 2024

Abstract
Deep learning-based models have recently shown a strong potential in Underwater Image Enhancement (UIE) that are satisfying and have the right colors and details, but these methods significantly increase the parameters and complexity of the image processing models and therefore cannot be deployed directly to the edge devices. Vision Transformers (ViT) based architectures have recently produced amazing results in many vision tasks such as image classification, super-resolution, and image restoration. In this study, we introduced a lightweight Context-Aware Vision Transformer (CAViT), based on the Mean Head tokenization strategy and uses a self-attention mechanism in a single branch module that is effective at simulating long-distance dependencies and global features. To further improve the image quality we proposed an efficient variant of our model which derived results by applying White Balancing and Gamma Correction methods. We evaluated our model on two standard datasets, i.e., Large-Scale Underwater Image (LSUI) and Underwater Image Enhancement Benchmark Dataset (UIEB), which subsequently contributed towards more generalized results. Overall findings indicate that our real-time UIE model outperforms other Deep Learning based models by reducing the model complexity and improving the image quality (i.e., 0.6 dB PSNR improvement while using only 0.3% parameters and 0.4% float operations).

Keywords Tokenization · Feature extraction · Image enhancement · Vision transformers

1 Introduction

Scattering and absorption make it difficult for light to travel through water, thus underwater object detection is difficult beyond 20-meter depth, whereas in muddy water the visibility

✉ Hammad Afzal
hammad.afzal@mcs.edu.pk

Muneeba Daud
mdaud.msse-27mcs@nust.student.edu.pk

Khawir Mahmood
khawir@mcs.edu.pk

[1] National University of Sciences and Technology, Islamabad, Pakistan

drops at only a 5-meter depth. In water, the light beam is reflected and deflected by dissolved salts, and organic and inorganic substances [1]. There are three essential parts of the light beam that the camera detects in underwater imaging systems. The direct portion of the light is not scattered and is directly reflected from an object. The second potion of light that is reflected after striking the target item at some angle is said to have undergone forward scattering [43]. Typically, this kind of scattering causes an image to appear blurry. The third component is backward scattering, which occurs when a light beam strikes an imaging system without first returning from the object. It only serves to further reduce an image's contrast [2]. All of the above phenomenons are explained in Fig. 1.

Images can be degraded because of a variety of factors and correcting just one of these factors might affect the other. As a result, when using image enhancement algorithms, many important factors must be considered for the Image Quality Assessment (IQA) like sharpness, dynamic range, and distortion [42]. The Human Visual System (HVS) processes visual data and connects the eyes to the brain. Evaluating underwater image quality uses subjective and objective methods. Subjective assessment by humans is impractical for underwater photos. Objective evaluation employs metrics like MSE, PSNR [4], SSIM [5], MAD [6] and no-reference UIQM [7], which aligns with the HVS. Deep learning based techniques have advanced significantly in the domain of image enhancement in recent years. These techniques, however, rely on intricate network structures and use an excessive amount of processing power. For instance, the generalizability of existing approaches is likely to be constrained by their tendency to bias towards a narrow range of brightness values and scenarios [3].

Some transformer-based architectural designs are utilized to tackle this problem. Utilization of backbone models that have been pre-trained on extensive datasets, such as ImageNet, is a common practice in achieving superior performance in high-level computer vision tasks like object detection and semantic segmentation. In contrast, algorithms designed for low-level vision tasks, such as image denoising, super-resolution, and deraining, are trained directly on task-specific data [10].

Our research is based on the idea that if a tiny dataset is sufficiently innovative, pre-trained models won't be able to aid with training on that domain and the model won't be suitable for that dataset. Although transfer learning is a powerful technique, it necessitates a pre-trained model for each dataset, increasing training time and complexity, requiring an additional forward pass [11]. Learning-based techniques have advanced significantly in the field of photo improvement in recent years. However, the implementation of the enhancement approaches on lightweight devices becomes significantly more challenging because they depend on complicated network architectures and use a lot of Computational resources. Therefore, our

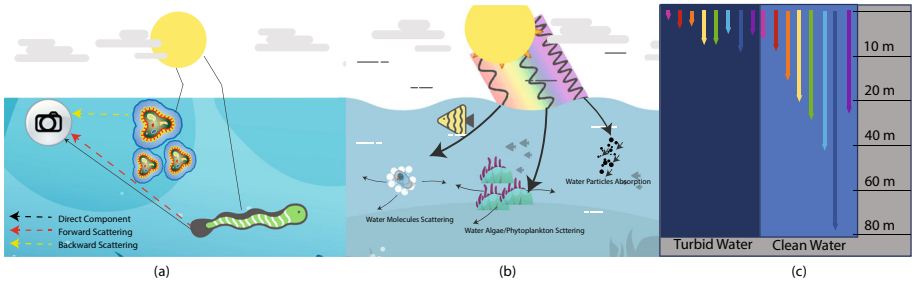

Fig. 1 (a) Components of camera light (b) Underwater light scattering and absorption (c) Attenuation of light of different wavelengths in clean and turbid water

proposed network aims to generate high-quality underwater images by utilizing vision transformers which are less complex and require less time, while also considering the inconsistent attenuation characteristics of different color channels and regions in underwater images.

To the finest of our knowledge, the methodology and framework presented in this work have not previously been used in the process of improving the quality of underwater images. The primary highlights of this work are:

- We created a new Context-Aware Vision Transformer (CAViT) that is simpler in design and uses fewer parameters and can perform the UIE task in real-time.
- We evaluated the proposed method using common datasets, to improve the network's generalization.
- We performed an ablation study to determine the optimal settings for our model. This was done by applying White Balancing and Gamma Correction methods.

The rest of the research article has been organized as follows: Section 2 gives an overview of the related work in the domain of Underwater Image Enchantment (UIE). Section 3 discusses the architecture of the proposed methodology. Section 4 describes the datasets used in this study. Section 5 is the results and ablation study of the best baseline model. Section 6 summarizes the research work and presents the limitations.

2 Literature review

The UIE task can be approached with various techniques, which can be grouped into two main categories i.e. Traditional UIE techniques, Machine & Deep Learning based methods. Further elaborations regarding this division is given below:

- The traditional UIE method includes two types of algorithms i.e. Physical models and non-physical models. For physical model-based UIE it is difficult to evaluate numerous parameters at once, and model hypotheses are not always reasonable in the complex and dynamic undersea environment. Also, the technology is expensive to adopt and is dependent on the development of an imaging system. Non-physical-based models have a limited number of direct applications for improving underwater image quality. The issue with the enhanced images is that either they have too little contrast or too much exposure [2].
- The lack of a large dataset with various underwater settings and high-fidelity reference photos is a problem for the current Machine Learning based underwater image enhancement (UIE) algorithms. Additionally, the uneven attenuation in various color channels and space regions is not fully considered for enhanced images. Deep learning approaches immediately learn the translation relationship between the source input images and the clean underwater image without being constrained by model assumptions or previous conditions.

The categories and their related model are further explained in the following sections.

2.1 Underwater image enhancement analysis using traditional techniques

The physical model-based restoration techniques use the reverse process of the imaging paradigm to obtain a clear image. A similar popular recovery model is the Jaffe-McGlamery underwater imaging model [16]. Using the Jaffe-McGlamery underwater imaging model, the

light obtained by the camera E_t was divided into three parts: the light reflected straight from an object, E_d the forward scattered portion, which is small-angle light reflected from a target, E_f and the backscattered light which is non-target reflected light E_b.

$$E_t = E_d + E_f + E_b \tag{1}$$

Meng et al. [17] exploited The Dark Channel Prior (DCP) based recovery, which depends on the sharpening method's maximum a posteriori probability (MAP) which improves visibility, lessens fuzziness, and enhances foreground retention textures but too many new parameters are added. In spite of the significant scattering effect of murky water, integral imaging technology has a tremendous impact since it can combine signals from several images. Cho [18] Single-photon imaging with a threshold was suggested by Li et al. [19]. By using this technique, photos captured in a high-loss underwater environment are reconstructed. The Peak Signal to Noise Ratio (PSNR) is theoretically improved by applying photon-limited computational techniques. The underwater dark channel prior (UDCP) technique, which only takes into account the blue and green channels, was proposed by Drews et al. [20] and achieved a more precise transmittance map compared with the DCP algorithm, increasing the restoration impact. Its robustness and dependability, however, fall short of the assumptions' restrictions.

There are a number of techniques proposed using Non-Physical Model Enhancement methods, including histogram-based, retinex-based, and visual fusion-based algorithms. Zhuang et al. [21] created a Bayesian and Retinex framework that enhances a single underwater image using multi-order gradient priors. This algorithm is effective for color correction, preserves the naturalness of the image, and improves the visibility of structures and details. However, the algorithm takes a long time to optimize. Song et al. [22] proposed a method that uses global stretching and multiscale fusing of dual models to eliminate unwanted color variations. The technique employs white-balancing and uses contrast and spatial signals in combination with a saliency weight coefficient method. However, there are some limitations to this approach, particularly with regard to the depth of the color model. Li et al. [23] suggested an approach for underwater hybrid systems that stretches the histogram and uses an improved white balance method. The method improves contrast and saturation, eliminates scattering-related blur, enhances color adjustment, reduces haze, and increases the clarity of details by creating a variable brightness and saturation enhancement model. However, this approach requires obtaining multiple fusion images and fusion weights. The process of image enhancement, utilizing the histogram equalization (HE) algorithm, involves the transformation of the image histogram from a narrow unimodal distribution to a more balanced distribution. Subsequently, the adaptive histogram equalization (AHE) method was formulated with the aim of enhancing the local contrast of the image. The algorithm known as Contrast Limited Adaptive Histogram Equalization (CLAHE), enhances the computational efficiency [24]. Iqbal et al. [25] introduced an unsupervised color correction method (UCM) for underwater images that relies on color correction and selective histogram stretching and can successfully reduce blue deviation while enhancing the low-component red channel and brightness. Huang et al. [26] suggest a straightforward yet efficient technique for improving shallow-water images called relative global histogram stretching (RGHS), which is based on appropriate parameter acquisition. The two components of the suggested method are color correction and contrast adjustment.

2.2 Underwater image enhancement analysis using machine learning & deep learning methods

Li et al. [27] developed an underwater image enhancement network called Ucolor, which uses medium transmission-guided multi-color space embedding. The network has a multi-color space encoder and a medium transmission-guided decoder. However, it does not produce visually satisfactory results when processing an underwater image with limited lighting. In order to process underwater images, Wang et al. [28] presented a parallel convolutional neural network with two parallel branches, a transmission estimation network, and a global ambient light estimate network. To avoid the halo effect and maintain edge properties, the network uses multiscale estimation and cross-layer connectivity. The contrast improvement, however, is not strong enough. WATER-NET, a gated fusion network, was suggested by Li et al. [29]. The underwater image is enhanced using white balance, histogram equalization, and gamma correction algorithms, and the final image is produced by integrating the confidence graphs of various enhancement techniques the reference model performs well in terms of generalization and has an opportunity for improvement.

For the purpose of enhancing underwater images, Guo et al. [30] presented the UWGAN, a new multiscale dense generated adversarial network that incorporates residual multiscale dense blocks into the generator. The discriminant uses the spectral normalization calculation method to stabilize discriminant training. Uplavikar et al. [31] proposed an algorithm for enhancing underwater images using domain adversarial learning. This algorithm improves the learning data for underwater target detection algorithms. An end-to-end dual generative adversarial network (DuGAN) for improving underwater images is proposed by Zhang et al. [32]. In which two discriminators are utilized to complete adversarial training. However, this solution relied on a user-guided way to gather reference photos, making it challenging to train with fresh images. Islam et al. [33] proposed FUnIE-GAN, an underwater image enhancement algorithm that is computationally efficient and able to run in real time. The algorithm uses a simpler generator model resulting in fast inference. The discriminator part of the GAN model is based on patch-level information rather than global recognition. However, the model is trained on a specific dataset, and its performance on other datasets or in other underwater environments may not be as good.

Transformer is a Seq2Seq framework that replaces conventional recurrent neural network used in Natural Language Processing (NLP) nearly entirely by introducing a self-attention strategy and using position embedding to account for the position information. Computer vision is undergoing a revolution because of the introduction of a new architecture called Vision Transformers (ViT). Vaswani et al. [8] The adoption of transformers in computer vision tasks, inspired by the significant achievements in natural language processing, has resulted in the emergence of Vision Transformers (ViTs). In recent years, vision transformers have gained significant prominence, as demonstrated by Dosovitskiy et al. [9].

Attention has intrinsic complexity of $O(N^2)$, which means that to evaluate the complexity of every pixel in relation to every other pixel in any low-resolution images like 256x256 pixels, the number of calculations would be enormous. Therefore, to make this strategy effective it is suggested in the paper "A picture is worth 16x16 words" [9] that the image would be divided into patches and are projected linearly to produce vectors, which are then combined with knowledge about the patch's location within the picture and fed into a traditional Transformer Encoder. The ViT model consists of multiple components, namely Image Tokenization, Positional Embedding, Classification Token, the Transformer Encoder, and a Classification Head. The insertion of information regarding the patch's original position

inside the image is essential since, despite being crucial to completely comprehending the image's content, this knowledge would be lost during the linear projection. The result relating to this patch being the one that is taken into account and fed into a Multi-Layer Perceptron (MLP). An additional vector is inserted that is unrelated of the picture being analyzed and is utilized to get global data on the entire image. This procedure is described visually in Fig. 2.

Efficient vision transformers There are several ways to interpret a model's efficiency. It could be referring to the model's memory footprint, which is significant when the memory of the accelerators on which the model is running is constrained. In addition to training and inference costs, efficiency may also refer to computing costs, such as the number of FLOPs. In particular, models frequently have to work within a severely constrained computational budget for on-device applications. Khan et al. [12]

Data-Efficient Image Transformers (DeiT) were proposed by Touvron et al. [13] in an effort to lessen reliance on data. Without extensive pretraining on dataset like ImageNet-21k. Additionally, DeiT variations were developed even further thanks to their inventive knowledge transfer method, particularly when a convolutional model was used as the instructor.

Token-to-token ViT (T2T- ViT), which employs a window and attention-based tokenization technique, was proposed by Yuan et al. [14]. Their tokenizer creates three sets of feature maps using three sets of kernel weights, extracting patches from the input feature map in a manner akin to convolution. In addition, a T2T tokenizer is more sophisticated and has more parameters than a convolutional one.

Convolutional vision Transformer (CvT) [15] presents convolutional transformer encoder layers that use convolutions rather than linear projections for the QKV in self-attention. In their tokenization process, they also include convolutions, and they report results that are competitive with those of other vision transformers on ImageNet-1k. All of these papers present findings after initial training on ImageNet (or bigger) data sets.

Image Processing Transformer (IPT) is the name of a pre-trained model Chen et al. [34] proposed based on the Transformer architecture. It is capable of restoring images in a variety of ways, including super-resolution, denoising, and deraining. IPT has a shared encoder-

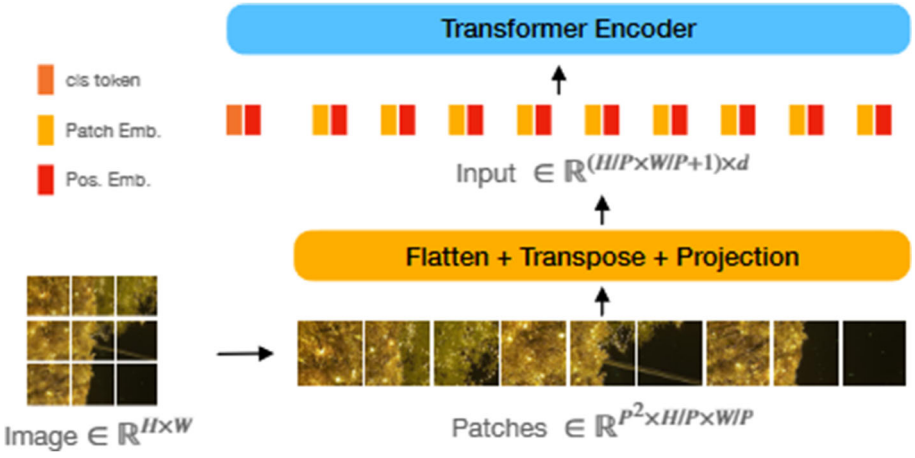

Fig. 2 Vanilla transformer architecture where image to sequence translation is shown where $X \epsilon R^{L \times d}$ denote a sequence of vectors $(x1, x2, \cdots, xL)$, where d is the embedding dimension of each vector

decoder Transformer body as well as many heads and tails that can each do a distinct task independently. Peng et al. [35] created a channel-wise multi-scale feature fusion transformer (CMSFFT) and a spatial-wise global feature modeling transformer (SGFMT) based on the attention mechanism, which they then integrated into the U-shape Transformer. They also created a multi-color space loss function that includes RGB, LAB, and LCH. An innovative underwater image improvement technique called UDAformer, developed by Shen et al. [36] is based on the Dual Attention Transformer Block (DATB), which also includes the Channels Self-Attention Transformer (CSAT) as well as Shifted Window Pixel Self-Attention Transformer (SW-PSAT). Huang et al. [37] proposed an Adaptive Group Attention (AGA) method, which is used within the Swin Transformer (ST) module. This method dynamically selects channels that are visually similar based on dependencies, reducing the need for additional attention parameters. Sun et al. [38] presented a network that includes the SwinMT module, which has two components: a unit for extracting low-frequency features and another for recovering high-frequency features to produce a high-quality image.

3 Proposed method

We create the CAViT model, which is focused on image improvement tasks and can operate without stacking convolution to extract structural data more effectively. Similar to word embedding in NLP, patches of the photograph are tokenized and turned into token embedding in our model. CAViT actively understands token-wise dependencies for input images rather than directly computing pixel-wise connections. CAViT enhances the image with excellent efficiency. Along with being highly effective, CAViT can intuitively learn the semantic information and hence produce results that are more semantically meaningful than CNNs. Nevertheless, obtaining comparable performance with CNN often requires a large amount of training data or extra supervision else cannot perform as expected due to the lack of inductive biases.

3.1 Overall architecture

An overview of the proposed lightweight Context-Aware Vision Transformer (CAViT) model is represented in Fig. 3. We introduced a novel UIE method, to start, given an unprocessed underwater image $I \epsilon R^{(H*W*C_I)}$. We first divide the image into patches of the form $I_p \epsilon R^{((H/P)(W/P)*C_I)}$, where P is the patch's size. The flattened image patches are then considered as a series of tokens $I_T \epsilon R^{(L*C_T)}$, wherein C_I & C_T are the input channel and transformer dimensions, respectively. And $L = (H/P)(W/P)$. The Context-Aware Vision Transformer module will then receive the created tokens I_T as inputs and produce an output structural map as $S_I \epsilon R^{(L*L*C_T)}$ using bilinear interpolation. Table 1 Presents the detailed Pseudocode of the proposed algorithm.

3.1.1 Tokenization strategy

In practice, the input features of the original image are designed to have a big dimension $t_i \epsilon R^{(P^2*C_I)}, where I = 1, 2, 3, \cdots, N$ necessitating high training parameters (e.g., 33 M parameters in [34]). An alternate approach is to derive input features from a sequence using CNN's feature maps. Therefore, the Mean Head method, where Adaptive Average Pooling immediately reduces the spatial size followed by Linear Head Embedding, is used in this

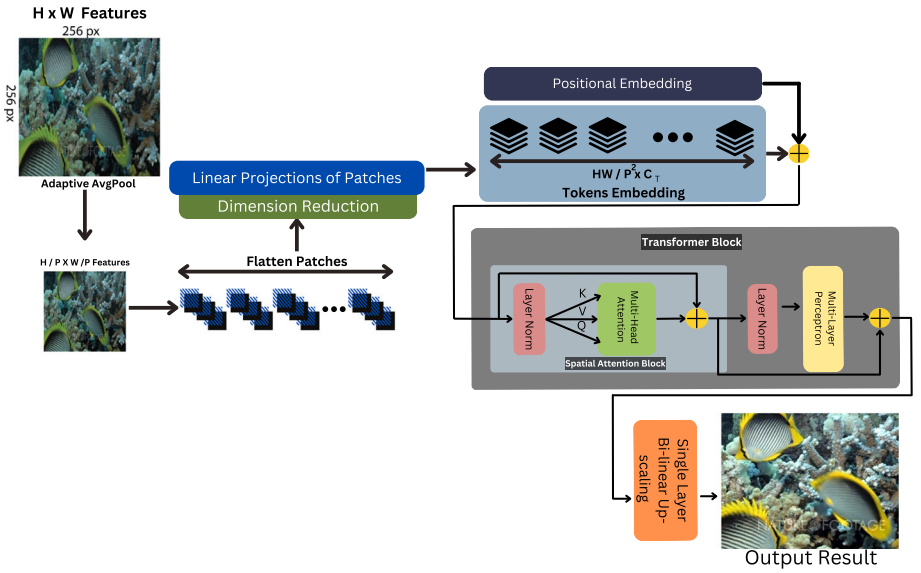

Fig. 3 Overall Architecture: We used a single branch transformer design with one encoder block, that explicitly combines the global and local context extraction modules to lessen model complexity. The conventional Transformer design used in this study was inspired by [15]

study to lower memory usage. This was inspired by the squeeze-and-excitement block [18] as shown in Fig. 4.

Input spatial resolution of the image is shown by the $H * W$ features. Similar to [14], Adaptive Average Pooling is applied to the image, using a 7 x 7 convolution with stride 4 and 16 output channels and then we feed the output features to the Linear Head Embedding. Linear Head Embedding is used in order to split input features into patches directly, which is then followed by linear projections. Behind every Patch, there is a cascading dimension reduction procedure. By using Mean Head, we decreased the tokenization complexity as much as possible. Empirically demonstrating the advantages of such a spatial dimension reduction for a transformer architecture. In essence, the structure maintains the same number of spatial tokens across all layers of the network and include a spatial reduction layer across every patch. Although the self-attention operation is not constrained by spatial distance, the spatial size of the feature has an impact on the size of the spatial area participating in attention. Utilizing CAViT's spatial reduction layers further increases the architecture's capacity.

The patches are projected linearly to produce vectors, which are then combined with knowledge about the patch's location within the image and then fed into a traditional Transformer Encoder. In order to keep position information we add a 1D learning positional embedding $p \epsilon R^{(C_T/2)}$ to Transformer inputs, as shown in Fig. 5.

3.1.2 Attention mechanism

We used a single branch with a local-global spatial attention module of the Transformer to process token sequences used in the area of computer vision. Which includes a Multi-Layer Perception (MLP) with a skip connection and a Multi-head Self-Attention (MSA) module. We choose GELU as the non-linearity function and LayerNorm (LN) as the normalization

Table 1 The description of the proposed algorithm

Pseudocode description of the proposed algorithm
Algorithm: Context-Aware Vision Transformer (CAViT) Training
Input:
- Parameters and hyperparameters (e.g., depth, heads, dropout, etc.)
- Training data (underwater images and targets)
- Patch size P
- Transformer dimensions C_I and C_T
Initialization: Define optimizer and loss function, enhance-net neural network with CAViT modules
Main Training Loop:
for epoch in range(number-of-epochs):
for batch in training-data:
Input-Images = preprocess(batch.underwater-images)
patches = divide-image-into-patches(input-images, P)
tokens = flatten-image-patches-to-tokens(patches)
reduced-tokens = apply-mean-head(tokens, C_T)
output-structural-map = context-aware-vision-transformer (reduced-tokens, C_T)
L = (H // P) * (W // P)
structural-map-with-interpolation = bilinear-interpolation (output-structural-map, L)
enhanced-images = enhance-net(structural-map-with-interpolation)
loss = compute-loss(enhanced-images, batch.targets)
End of Training

function. In summary, the transformer module can be defined as:

$$A_0 = I_T + p \tag{2}$$

$$\tilde{A_N} = MSA(LN(A_{N-1})) + A_{N-1} \tag{3}$$

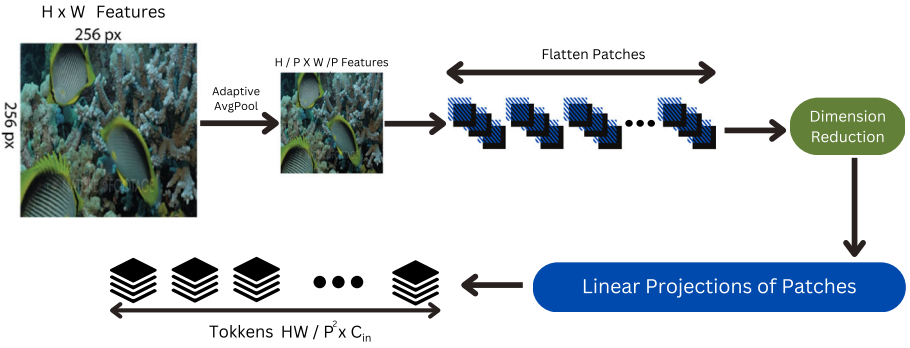

Fig. 4 Mean head / squeeze-and-excitation tokenization strategy

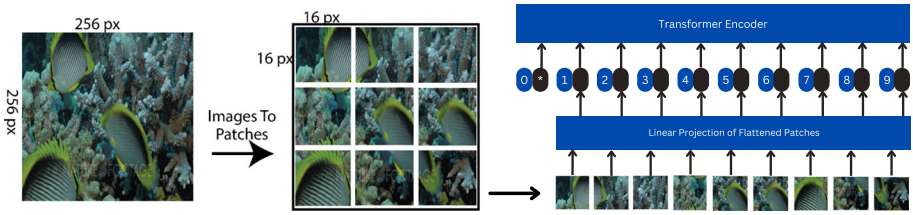

Fig. 5 Dividing an image into patches and then feeding them into the transformer encoder

$$A_N = MLP(LN(\tilde{A_N})) + \tilde{A_N} \tag{4}$$

where N represents the Transformer's depth (number of basic transformer blocks i.e. 1 for our model). In a typical Transform block, a linear layer creates the projections from the input features, Query (Q), Key (K), & Value (V), but only accomplishes global spatial interactions. It seems sense to substitute a convolution with a kernel size of 3 x 3 for the linear layer in order to employ more local information, as this simultaneously reinforces the channel and spatial augmentation. In order to cover neighboring tokens for the convolutions 2-D Block convolutions are utilized to analyze the rearranged picture tokens as opposed to 1-D convolutions, which are used for processing sentences in natural language processing (NLP).

3.1.3 Defining the loss function

To objectively evaluate the performance of the model, we use a combination of Gradient Loss and Cosine Similarity Loss as a loss function. Gradient Loss not only collects low-frequency information, such as the L1 loss but also, by adding a second-order constraint acquires high-frequency information. L1 loss is employed by minimizing the MAE (Mean Absolute Error) between generated and ground truth patches during network training. Let G and $\tilde{G}$ denote the gradient map of X and $\tilde{X}$, where $\tilde{X}$ and X denote the restored and the real

Fig. 6 Li [29] dataset encompass a wide variety of underwater settings, with different aspects of quality degradation, and includes a diverse range of image information. It also include high-quality reference images that correlate to the underwater images

image, respectively. $Q(r)$ and $Q(g)$ are the distribution of $\tilde{G}$ and G respectively. Gradient Loss is stated as:

$$L_{gd} = E_{\tilde{G}\sim Q(r),6G\sim Q}\|\tilde{G} - G\|_1 \tag{5}$$

The Cosine Similarity Loss is stated as:

$$L_{coss} = 1 - \frac{A * B}{\|A\| * \|B\|} \tag{6}$$

where A and B are the predicted and ground truth vectors, respectively, and ||A|| and ||B|| are their L2 norms. The linear summation of the two loss functions results in the total loss function, which is given as:

$$L_{sum} = L_{coss} + L_{gd} \tag{7}$$

4 Underwater image datasets

There are two datasets utilized in this study:

- The UIEB dataset: Li et al. [29] introduced The underwater Image Enhancement Benchmark (UIEB) dataset consisting of 950 real-world underwater photos. It is further subdivided into 890 pairs of raw underwater photographs and associated high-quality reference images, and the remaining 60 difficult images without reference (this set of images is denoted as Test-U60). Out of 890 images 300 images are used as training dataset denoted as Train-U and 90 images are used as testing dataset denoted as Test-U90. The dataset was annotated from different Internet sites, relevant papers, and video footage.

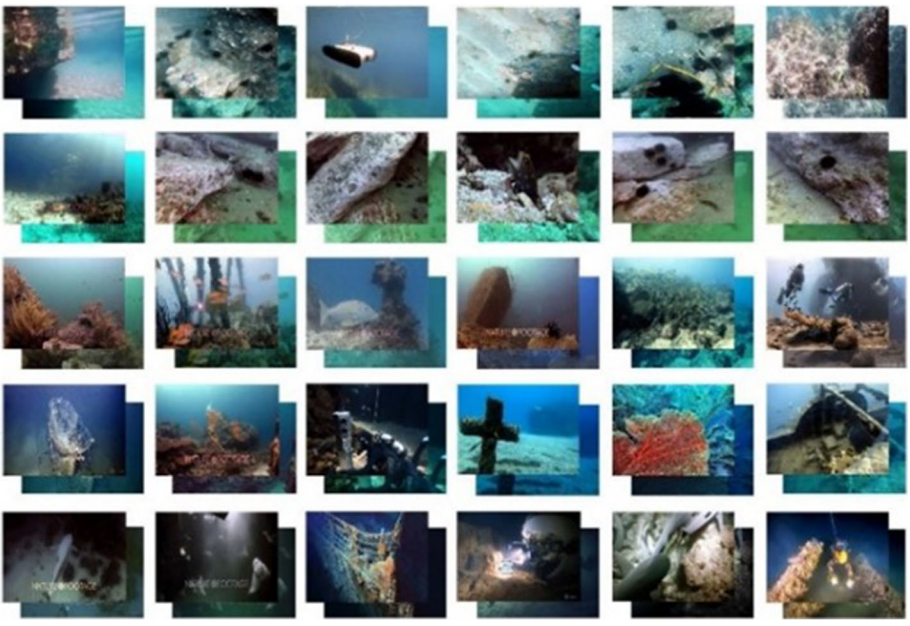

Fig. 7 Peng et al. [12] created an LSUI (Large-Scale Underwater Image) dataset that surpasses current underwater datasets in terms of coverage of underwater scenes and visual quality of reference photos

It contains a variety of underwater scenes and aquatic animals. To create high-quality reference images as shown in Fig. 6, 12 image enhancement techniques were applied to the training dataset. Volunteers choose the final, high-quality reference photos.

- The LSUI Dataset: Peng et al. [35] created the dataset that contains 5004 pairs of natural underwater photographs. Compared to the current underwater datasets, it has better reference photographs with more varied underwater habitats as shown in Fig. 7 The images in Large-Scale Underwater Image (LSUI) dataset have richer features and diverse underwater settings (lighting conditions, water kinds, and targeted categories). The LUSI dataset containing 1500 images denoted as Train- L is being utilized as the training images for the proposed transformer models. And the rest 70 images demoted as Test-L70 are utilized as the testing dataset for the proposed dataset.

5 Results and discussion

On both UIEB dataset and LUSI dataset, we carried out extensive trials. The Mean Square Error (MSE), Peak Signal to Noise Ratio (PSNR), Most Apparent Distortion (MAD) and Structural SIMilarity index (SSIM) are commonly employed as full-reference evaluation metrics for measuring the similarity between an image and a reference image. A higher PSNR, SSMI and MAD value indicates a closer match in terms of image content, while a lower MSE score signifies a greater similarity in terms of structure and texture. The evaluation of underwater image quality is commonly conducted using non-reference metric such as the and the Underwater Image Quality Measure (UIQM). These metrics assess various aspects of image quality, including color density, saturation, sharpness, and contrast. A higher score in UIQM indicates better human visual perception of the underwater image.

5.1 Underwater images quality assessment techniques

For underwater images to be more visually appealing, evaluating their quality is essential. Subjective and objective approaches can be used to classify quality evaluation techniques. In a subjective evaluation, human observers give ratings based on their viewpoint, such as mean opinion scores (MOS). Due to its labor-intensive nature and lack of automation, especially in the perspective of contemporary deep learning techniques, this method is however impracticable for underwater photos. On the other hand, objective evaluation employs computer algorithms for instantaneous scoring. Full reference (comparing photos), reduced reference (with only part of the picture), or no reference (independent evaluation) are the possible options. Human Visual System (HVS) refers to the intricate sensory system in humans that is in charge of processing visual data and enabling us to experience and understand the environment around us. The complex neurological connections between the eyes and the brain are included in the HVS. The evaluation of full reference picture quality involves comparing a reference image with a distorted image and quantifying the discrepancies in order to derive a numerical score. Traditional evaluation indicators commonly used in full-reference and semi-reference evaluations include Mean square error (MSE), peak signal-to-noise ratio (PSNR) [4], structural similarity index (SSIM) [5] and most apparent distortion (MAD) [6]. The initial approaches to full-reference quality assessment mostly relied on the analysis of distortion energy. However, MAD [6] incorporated and analyzed luminance and contrast masking to assess the perceived distortion based on high-quality images. The underwater image colorfulness measure (UICM), the underwater image sharpness measure (UISM), and

the underwater image contrast measure (UIConM) are the three underwater image attribute measures that make up the UIQM (underwater image quality measure) [7], which is a no-reference quality metric. Each offered attribute measure is motivated by the characteristics of human visual systems (HVSs), and each attribute is chosen for analyzing one component of underwater picture degradation. The experimental findings show that the methods accurately assess the quality of underwater images in line with human perceptions.

5.2 Objective image quality evaluation metrics

Non-reference evaluation and full-reference evaluation are the two basic categories of objective assessment methods. In this work, we will evaluate our model on both categories.

5.2.1 Full-reference image quality evaluation

We carried out a full-reference assessment using Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) measures, which reflect the similarity of the enhanced images with the reference images (provided in the dataset). Higher values of the PSNR and SSIM indicate the better visual quality of the images. The equation below displays the Mean Square Error.

$$MSE = \frac{1}{M * N} \Sigma_{(M,N)} [l_{1(m,n)} - l_{2(m,n)}]^2 \tag{8}$$

$l_{1(m,n)} - l_{2(m,n)}$ stands for the original and enhanced images, respectively. A peak signal-to-noise ratio is given by the following equation.

$$PSNR = 2 \log_{10} \frac{L - 1}{RMSE} \tag{9}$$

Root Mean Squared Error is referred to as RMSE. Three crucial elements are extracted from an image via the Structural Similarity Index (SSIM) metric; Structure, Contrast, and Luminance. These three elements serve as the foundation for the comparison of the two photos. And finally, the SSIM score is given by,

$$SSIM(x, , y) = [l(x, y)]^{\alpha} . [c(x, y)]^{\beta} . [s(x, y)]^{\gamma} \tag{10}$$

where α, β, γ denote the relative importance of each of the three components.

The Most Apparent Distortion (MAD) is a perceptual image quality metric that quantifies the perceived quality difference between a reference image and a distorted image. It's defined as the average of the absolute pixel-wise differences between the luminance values (brightness values) of corresponding pixels in the reference and distorted images. Mathematically, MAD can be expressed as follows: Let $R(x, y)$ be the luminance value of the pixel at coordinates (x, y) in the reference image, and let $D(x, y)$ be the luminance value of the pixel at the same coordinates in the distorted image. The MAD between the reference image and the distorted image is calculated as,

$$MAD = \frac{1}{N} \Sigma_{x=1}^{W} \Sigma_{x=1}^{W} \| R(x, y) - D(x, y) \| \tag{11}$$

5.2.2 No-reference image quality evaluation

UIQM focuses on Underwater Image Color Measure (UICM), Underwater Image Sharpness Measure (UISM), and Underwater Image Contrast Measure (UIConM). Better visual perception is indicated by a higher UIQM score. The equation for the UIQM, which combines the measures of color, sharpness, and contrast, is provided by:

$$UIQM = \alpha.UICM + \beta.UISM + \gamma.UIConM \tag{12}$$

The weighted coefficients, $\alpha = 0.0282$, $\beta = 0.2953$, and $\gamma = 3.5753$ are used to balance the values of the three measures.

- Underwater Image Colorfulness Measure (UICM): Both Red-Green (RG) and Yellow-Blue (YB) color components are evaluated by the UICM in order to gauge the effectiveness of color-correcting algorithms.
- Underwater Image Sharpness Measure (UISM): First, each color component is used to create edge maps, which are then used to measure the sharpness. Then, to determine the grayscale edge maps, the resulting edge maps are multiplied by the original color component.
- Underwater Image Contrast Measure (UIConM): Contrast is an underwater visual performance factor. Backscattering is typically to blame for the contrast reduction in underwater photos.

5.3 Experimental analysis of proposed transformer based model

On both UIEB dataset and LUSI dataset, we carried out extensive trials. The all obtained statistical results will be discussed in the following sections.

5.3.1 Preprocessing of images for CAViT

Both the training and the test images are downsized to 1200 x 900 pixels depending on their longest side [40]. In order to achieve comparability, the dataset was split into 80% training examples and 20% test data. Additionally, random cropping, resizing, flipping, and rotating are used to enhance the training data.

5.3.2 Implementation details

Pytorch-based implementation is carried out on NVIDIA-SMI 460.32.03 GPU and CUDA 11.2. The Adam optimizer is used for processing with a preset learning rate of $1e^{-4}$. The default setting was set to use Transformer depth 1 as concluded with multiple experiments that increasing the depth of the transformers did not improve any artifacts of the suggested pipeline. Each image will be divided into 32 by 32 tokens. The skip connection ratio is set to 0.1 and the scale factor in Q and K vector is 8, and the number of heads of the Transformer is set to 8. CAViT and CAViT_G denote the model with and without gamma correction, respectively (explained in Section 5.3.3). The rest of the details are shown in Table 2.

5.3.3 Ablation study

The White Balancing and Gamma Correction branch tries to improve the appearance of underwater images by eliminating undesired color casts brought on by various illuminants. It

Table 2 Defining the training hyper-parameters (number of epochs, batch size and tokenization features) for LUSI (training dataset Train-L) and UIEB (training dataset Train-U) for the proposed CAViT and Gamma correction based $CAViT_G$ Models

	Tokenization	Branch	Epochs	Batch	Features
CAViT, CAViT_G	Adaptive Average Pooling (Mean Head Strategy)	Single branch model	30	8	24

is employed before passing the image to the transformer module. Because underwater images suffer noticeably when water depths are greater than 30 feet, the goal of the White Balancing and Gamma Correction branch is to improve the overall contrast and brighten up dark areas of underwater image images.

In Table 3 we reported the Full-Reference Test on Both Dataset (LUSI & UIEB). Using the Test-L70 for the LUSI dataset & Test-U90 for the UIEB dataset. We evaluate both the models i.e simple as well as the gamma correction based model with evaluation metrics; PSNR and SSIM. It can be concurred that the gamma based model gives more PNSR and SSIM values but their running time is a little more then the base model.

The enhancement outcomes of our proposed methods for both CAViT and CAViT_G for the LUSI dataset are visually compared in Fig. 8, using the images from the train dataset of LUSI Train-L. The results given by (c) are most similar to the reference image, which has less color artifacts and high-fidelity object areas, which also supports the hypothesis given in ablation study as well as the evaluation results.

In Table 4 we reported the Non-Reference Test on Both Dataset (LUSI & UIEB). Using the Test-L70 for the LUSI dataset & Test-U90 for the UIEB dataset. We evaluate both the models i.e simple as well as the gamma correction based model with evaluation metrics; UICM, UISM, UIConM and UIQM. It can be endorsed that the gamma based model gives more overall UIQM value UICM, UISM, UIConM being its components.

The enhancement outcomes of our proposed methods for both CAViT and CAViT_G for the UIEB dataset are visually compared in Fig. 9, using the images from the train dataset of UIEB Train-U. The results given by (c) are most similar to the reference image, which has less color artifacts and high-fidelity object areas, which also supports the hypothesis given in ablation study as well as the evaluation results.

It can be concluded from Table 5 that CAViT_G has a lower latency than CAViT on the LUSI dataset, indicating that it is faster in making inferences. On the UIEB dataset, CAViT

Table 3 Full-reference test on both dataset (LUSI & UIEB) using test-l70 & test-U90, respectively

Model	Test set	Training time (s)↓	PSNR (dB)↑	SSIM↑
CAViT	Test-L70	**2273.31**	24.80	0.93
CAViT_G	Test-L70	10253.25	**25.76**	**0.95**
CAViT	Test-U90	**1057.23**	21.37	0.89
CAViT_G	Test-U90	4174.406	**23.54**	**0.96**

Presenting evaluation metrics (training time, number of parameters, PSNR and SSIM) for both models CAViT and CAViT_G
The top results are bold

Fig. 8 Enhancement results of CAViT and CAViT_G trained on LUSI. (a): Input images. (b): Enhanced results using the model trained on Train-L. (c): Enhanced results using the model trained with the Gamma Correction component on Train-L. (d): Reference images (recognized as ground truth (GT))

has a slightly lower latency than CAViT_G, but the difference is minimal. The choice between CAViT and CAViT_G depends on specific priorities. If low latency and good image quality are crucial, CAViT_G might be preferred. However, CAViT still has competitive performance.

Figure 10 shows the visual results from the inference results form the LUSI dataset for both the models i.e. CAViT and CAViT_G. It may be argued that LSUI contains reference photos with richer underwater settings and higher visual quality than existing underwater image datasets, which enhanced the tested network's capacity for improvement and generalization.

Figure 11 shows the visual results from the inference results form the UIEB dataset for both the models i.e. CAViT and CAViT_G. UIEB have more intensities of different color gradations and better visibility and brightness.

Table 4 Non-reference test on both dataset (LUSI & UIEB) using test-l70 & test-U90, respectively

Model	Test Set	UICM↑	UISM ↑	UIConM ↑	UIQM↑
CAViT	Test-L70	5.19	5.59	**0.19**	2.49
CAViT_G	Test-L70	**5.59**	**5.79**	0.19	**2.69**
CAViT	Test-U90	**8.25**	7.16	0.25	3.23
CAViT_G	Test-U90	7.89	**7.89**	**0.28**	**3.29**

Presenting evaluation metrics (UICM, UISM, UIConM & UIQM) for both models CAViT and CAViT_G
The top results are bold

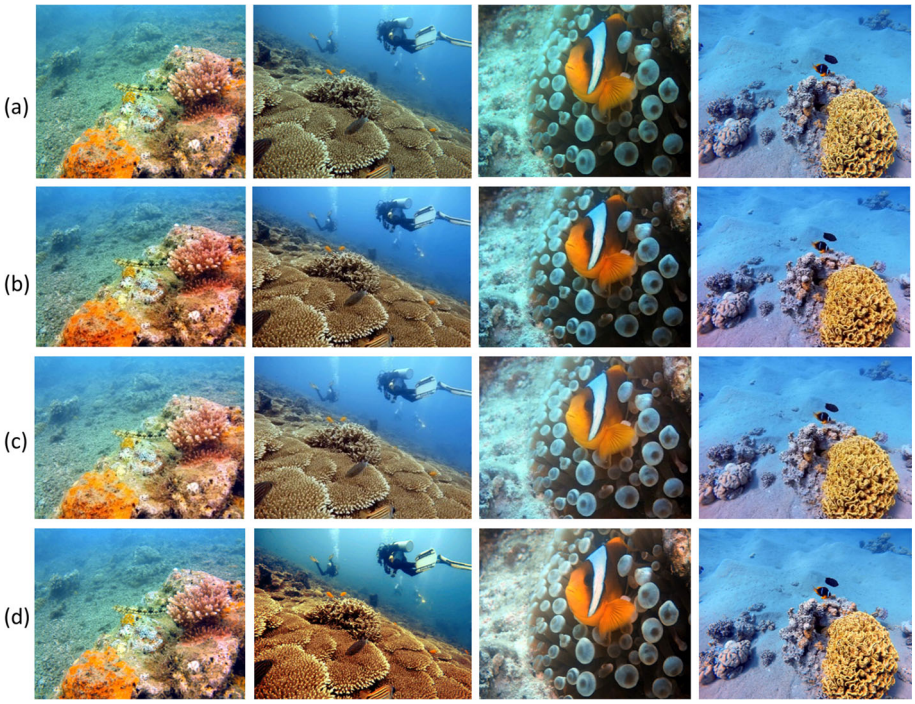

Fig. 9 Enhancement results of CAViT and CAViT_G trained on UIEB underwater datasets. (a): Input images. (b): Enhanced results using the model trained on the Train-U dataset. (c): Enhanced results using the model trained with the Gamma correction component on the Train-U dataset. (d): Reference images (recognized as ground truth (GT))

Table 5 Inference Results on both Datsets (LUSI and UIEB) on CAViT and $CAViT_G$ Models stating (Latency, Number of Paramenters, PSNR, SSIM and MSE Vaules)

Model	Dataset	Latency (s) ↓	Parameters (K) ↓	PSNR (dB) ↑	SSIM ↑	MSE ↓
CAViT	LUSI	0.0034 s	21.87 K	**23.81**	**0.967**	270.69
CAViT_G	LUSI	**0.0025** s	21.87 K	23.89	0.969	**265.19**
CAViT	UIEB	**0.0032** s	21.87 K	**25.49**	**0.981**	183.83
CAViT_G	UIEB	0.0034 s	21.87 K	26.36	0.981	**150.18**

The top results are bold

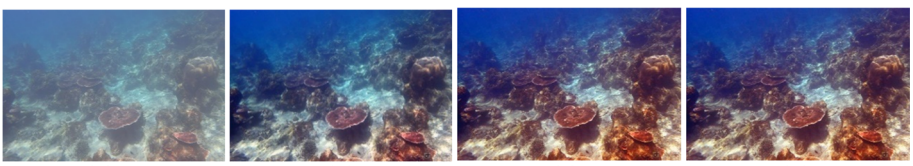

Fig. 10 Inference results. (a): Input images. (b): Enhanced results using the model trained on the Train-L dataset. (c): Enhanced results using the model trained with the Gamma Correction component on the Train-L dataset. (d): Reference images (recognized as ground truth (GT)

Fig. 11 Inference results. (a): Input images. (b): Enhanced results using the model trained on the Train-U dataset. (c): Enhanced results using the model trained with the Gamma Correction component on the Train-U dataset. (d): Reference images (recognized as ground truth (GT)

Table 6 Quantitative comparison of traditional UIE methods (CLAHE, UICM, RGHS and UDCP) on the full-reference testing dataset UEIB test dataset U-90 presenting metrics (PSNR, SSIM, MAD and inference time

Methods	PSNR ↑	SSIM ↑	MAD↓	Time (s)
CLAHE [24]	20.64	0.82	100.0	x
UCM [25]	22.03	0.81	92.95	x
RGHS [26]	23.57	0.80	81.02	8.92s
UDCP [20]	13.47	0.54	139.0	30.82s
Ours	**23.54**	**0.96**	**80.54**	**0.0025s**

The top results are bold

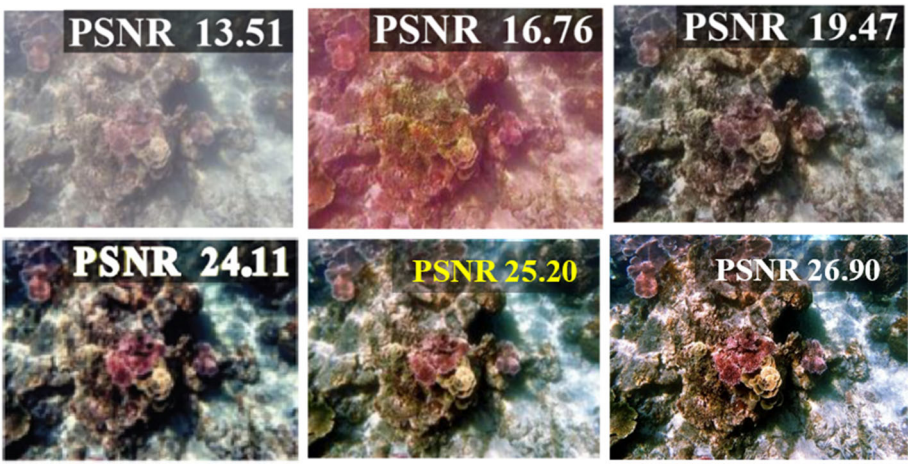

Fig. 12 Visual comparison of enhancement results sampled from the Test-U90 (UIEB) dataset. From left to right are raw underwater images, FUnIE [33], UGAN [41], Ucolor [27], U-Trans [35] and our CAViT. the reference image recognized as ground truth (GT). The highest PSNR value from the mentioned methods is marked in yellow

Table 7 Quantitative comparison of deep learning based UIE methods on the full-reference testing dataset UEIB Test Dataset U-90 presenting (PSNR, SSIM, Number of Parameters, Inference Time and Flops)

Methods	Technique	Test U-90 PSNR	SSIM	Parameters (K)	Time (s)	Flops (G)
WaterNet [29]	CNNs	19.81	0.86	24810	0.61s	193.7
Ucolor [27]	CNNs	20.78	0.87	157400	2.75s	443.8
FUnIE [33]	GANs	19.45	0.85	7019	0.09s	10.23
UIE-DAL [31]	GANs	16.37	0.78	18820	0.07s	29.32
UGAN [41]	GANs	20.68	0.84	57170	0.05s	38.97
U-Trans [35]	ViTs	22.91	0.91	65600	0.07s	66.2
Ours	ViTs	**23.54**	**0.96**	**21.87**	**0.0025s**	**0.04**

The top results are bold

5.3.4 Comparative analysis of various UIE methods

We report the model size and corresponding average PSNR, SSIM, MAD and UIQM on LUSI and UIEB evaluation datasets. To demonstrate the performance of our proposed model, we compare the CAViT Model with 6 deep learning UIE approaches & 4 traditional UIE techniques. It contains comparison of different deep learning techniques: WaterNet [29], U-Trans [35], Ucolor [27], FUnIE [33], UIE-DAL [31] and UGAN [41]. And other traditional UIE methods like CLAHE [24], UCM [25], RGHS [26] and UDCP [20]. The top results are bold. This analysis includes both the Non-Reference and Full Reference evaluation techniques. As well as the visual results are presented at the end to fully understand the metrics and demerits of each method.

For the evaluation of the traditional techniques in Table 6, we have presented the metrics on PSNR, SSIM, MAD. For the comparison of the models performance we have also added the

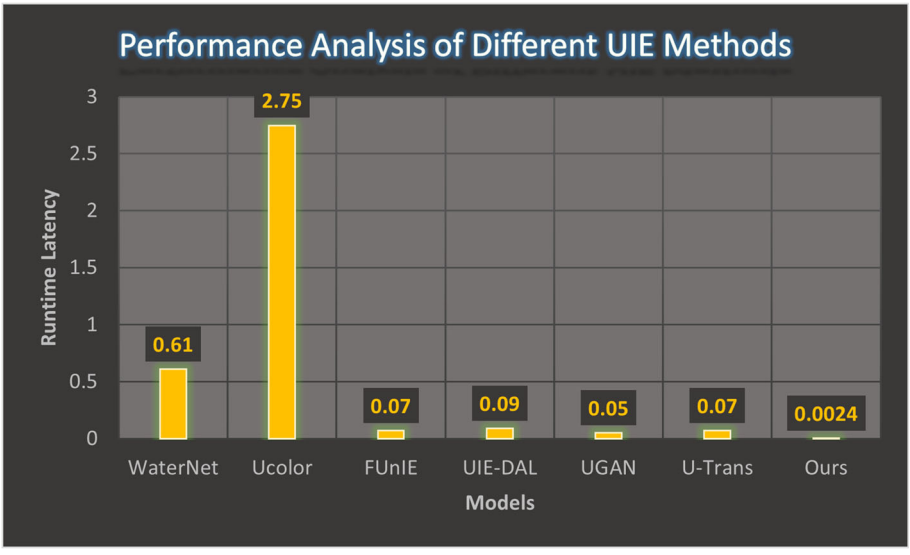

Fig. 13 Graphical of comparison our CAViT inference (run-time) efficiency against WaterNet [29] FUnIE [33], UGAN [41], Ucolor [27], U-Trans [35] results sampled from the Test-U90 (UIEB) dataset

Table 8 Quantitative comparison among different deep learning UIE methods on the non-reference testing dataset U-60 presenting the values of UIQM

Methods	Test U-60 UIQM↑
WaterNet [29]	0.92
U-Trans [35]	0.85
Ucolor [27]	0.84
FUnIE [33]	1.03
UIE-DAL [31]	0.72
UGAN [41]	0.86
Ours	**2.37**

The top results are bold

inference time. It is evident that our model outperforms the traditional methods like CLAHE [24], UCM [25], RGHS [26] and UDCP [20].

The statistical results and visual comparisons are summarized for the Full-Reference UIEB testset U-90 in Table 7 and Fig. 12. As in Table 7, our Context-Aware Vision Transformer demonstrates the best performance on both PSNR and SSIM metrics with relatively few parameters, FLOPs, and running time.

The 6 other deep learning approaches are restricted as explained: The advantage of FUnIE is that it can produce models that are quick, light, and require few parameters, but this inherently restricts its capacity to handle complex and distorted test data. Both UGAN and UIE-DAL did not take into account the varied qualities of the underwater photos. Simply introducing the idea of multi-color space into the network's encoder part cannot effectively take advantage of it, which results in unsatisfactory results in terms of contrast, brightness, and detailed textures. Ucolor's media transmission map prior cannot effectively represent the attenuation of each area. U-trans achieved comparative results but the size of their model is relatively large. A graphical comparison of the inference results of the 6 deep learning approaches is also given in Fig. 13.

Fig. 14 Enhancement results for Test-U60. The images represent underwater scenes of yellowish, greenish-bluish colors. (a) Raw images. (b) Enhanced results. (c) Raw images. (d) Enhanced results

The statistical results and visual comparisons are summarized for the Non-Reference UIEB testset U-60 in Table 8 and Fig. 14. In Table 8, our Context-Aware Vision Transformer demonstrates the best performance on UIQM metric.

6 Conclusions

For many practical purposes, such as underwater exploration, monitoring, and recovery operations carried out by semi- or fully autonomous robots, underwater image improvement is crucial. This provides a substantial challenge for computer vision and image processing. Scientific study, environmental preservation, and industrial applications can all benefit from the capacity to enhance image quality in difficult underwater environments. For marine scientists researching aquatic ecosystems, accurately visualizing underwater landscapes is essential because it makes it easier for them to monitor and record sensitive marine species and their environments.

In this research, we present a lightweight deep learning model for enhancing underwater image quality called Context-Aware Vision Transformer (CAViT). The development of a compact and effective model like CAViT also meets the expanding demand for resource-effective deep learning solutions across numerous disciplines. In addition to being appropriate for resource-constrained underwater robotic platforms, its minimal memory usage and rapid inference also pave the door for more effective deep learning applications in other resource-limited contexts.

We carried out a number of tests to evaluate the suggested model settings. By comparing the suggested model to prior state-of-the-art work, quantitative and qualitative image quality assessment findings demonstrate the proposed model's efficiency and effectiveness in underwater image quality enhancement in terms of color distortion, low visibility, and poor contrast.

As we move forward, new avenues for study and invention open up with the addition of CAViT to the field of low-level vision tasks. The adaptability of this strategy is demonstrated by the incorporation of the transformer design into various tasks, which suggests that it could be a potent tool for overcoming difficulties in the field of computer vision. In the future, perception-related loss functions can be used to train deep learning-based networks, which can then be used to incorporate factors that are congruent with how people interpret aesthetics. It will consequently improve the network's ability to alter visual sensitivity and contrast.

Funding The authors did not receive support from any organization for the submitted work.

Data Availability Following datasets are used for analysis during the current study:
UIEB (Underwater Image Enhancement Benchmark Dataset) available at https://paperswithcode.com/dataset/uieb
LSUI (Large Scale Underwater Image Dataset) available at https://paperswithcode.com/dataset/lsui

Declarations

Conflicts of interests/Competing interests The authors have no relevant financial or non-financial interests to disclose.

References

1. Wang RW (2015) Review on underwater image restoration and enhancement algorithms. In Proceedings of the 7th international conference on internet multimedia computing and service, pp 1-6
2. A survey on underwater images enhancement techniques (2020) IEEE 9th international conference on communication systems and network technologies (CSNT) pp 333-338. IEEE
3. Hu K, Weng C (2022) An overview of underwater vision enhancement: from traditional methods to recent deep learning. J Mar Sci Eng, 241
4. Qian J, Wu D, Li L, Cheng D, Wang X (2014) Image quality assessment based on multi-scale representation of structure. Digit Signal Process 33:125–133
5. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612. https://doi.org/10.1109/TIP.2003.819861
6. Larson EC, Chandler DM (2010) Most apparent distortion: full-reference image quality assessment and the role of strategy. J Electron, Imaging, p 011006
7. Panetta K, Gao C, Agaian S (2015) Human-Visual-System-Inspired Underwater Image Quality Measures. IEEE J Ocean Eng 41(3):541–551. https://doi.org/10.1109/JOE.2015.2469915
8. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. In NIPS
9. Dosovitskiy AB (2020) An image is worth 16x16 words: transformers for image recognition at scale
10. Ruan B, Shuai H, Cheng W (2022) Vision Transformers: State of the Art and Research Challenges. ArXiv:2207.03041
11. Tay Y, Dehghani M, Bahri D, Metzler D (2020) Efficient transformers: A survey. arXiv:2009.06732
12. Khan S, Naseer M, Hayat M, Zamir SW, Khan FS, Shah M (2021) Transformers in vision: a survey. https://doi.org/10.1145/3505244
13. Touvron H, Cord M, Douze M, Massa F, Sablayrolles A, Jegou H (2021) Training data-efficient image transformers & distillation through attention. Proceedings of the 38th international conference on machine learning
14. Yuan L, Chen Y, Wang T, Yu W, Shi Y, Jiang ZH, Tay FE, Feng J, Yan S (2021) Tokens-to-token vit: training vision transformers from scratch on imagenet. arXiv:2101.11986
15. Wu H, Xiao B, Codella N, Liu M, Dai X, Yuan L, Zhang L (2021) Cvt: introducing convolutions to vision transformers. arXiv:2103.15808
16. Trucco E, Olmos-Antillon AT (2006) Self-tuning underwater image restoration. Oceanic Engineering, IEEE Journal, pp 511–519
17. Meng H, Yan Y, Cai C, Qiao R, Wang F (2020) A hybrid algorithm for underwater image restoration based on color correcgtion and image sharpening. Multimed Syst, 1-11
18. Cho M, Javidi B (2010) Three-dimensional visualization of objects in turbid water using integral imaging. J Disp Technol 6:544–547
19. Li Z, Zhou H, Li Z, Yan Z, Hu C, Gao J, Jin X (2021) Thresholded single-photon underwater imaging and detection. Opt. Express, 28124-28133
20. Drews PL, Nascimento ER, Botelho SS, Campos MFM (2016) Underwater depth estimation and image restoration based on single images. IEEE Comput Graph Appl 36:24–35
21. Zhuang P, Li C, Wu J (2021) Bayesian retinex underwater image enhancement. Eng Appl Artif Intell, 104171
22. Song H, Wang R (2021) Underwater image enhancement based on multi-scale fusion and global stretching of dual-model. Mathematics, 595
23. Li X, Hou G, Tan L, Liu W (2020) A hybrid framework for underwater image enhancement. IEEE Access, 197448-197462
24. Zuiderveld K (1994) Contrast limited adaptive histogram equalization. Graph Gems, 474-485
25. Iqbal K, Odetayo M, James A, Salam RA, Talib AZH (2010) Enhancing the low quality images using unsupervised colour correction method. IEEE Int Conf Syst Man Cybern 1703-1709
26. Huang D, Wang Y, Song W, Sequeira J, Mavromatis S (2018) Shallow-water image enhancement using relative global histogram stretching based on adaptive parameter acquisition, In: International conference on multi media modeling, pp 453-465
27. Li C, Anwar S (2021) Underwater image enhancement via medium transmission-guided multi-color space embedding. IEEE T Image Process, 4985-5000
28. Wang K (2019) Underwater image restoration based on a parallel convolutional neural network. Remote Sens, 1591
29. Li C (2019) An underwater image enhancement benchmark dataset and beyond. IEEE Trans Image Process, 4376-4389

30. Guo YH (2019) Underwater image enhancement using a multiscale dense generative adversarial network. IEEE J Ocean Eng, 862-870
31. Uplavikar PM, Wu Z, Wang Z (2019) All-in-one underwater image enhancement using domain-adversarial learning. CVPR Workshops, pp 1-8
32. Zhang H, Sun L, Wu L, Gu KD (2021) An effective framework for underwater image enhancement. IET Image Process
33. Islam MJ, Xia Y, Sattar J (2020) Fast underwater image enhancement for improved visual perception. IEEE Robot Autom, 3227-3234
34. Chen H, Wang Y (2020) Pre-Trained Image Processing Transformer
35. Peng LC (2021) U-shape Transformer for Underwater Image Enhancement
36. Shen ZX (2022) UDAformer: underwater image enhancement based on dual attention transformer. SSRN, p 4162641
37. Huang ZL (2022) Underwater image enhancement via adaptive group attention-based multiscale cascade transformer. IEEE Trans Instrum Meas, 1-18
38. Sun JD (2022) Swin transformer and fusion for underwater image enhancement. In: International workshop on advanced imaging technology (IWAIT) 2022 vol 12177, SPIE, pp 627-631
39. Hu J, Shen L (2018) Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7132-7141
40. Guo C, Li C (2020) Zero-reference deep curve estimation for low-light image enhancement. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 1780-1789
41. Fabbri C, Islam MJ, Sattar J (2018) Enhancing underwater imagery using generative adversarial networks. ICRA, pp 7159-7165
42. Yuan X, Guo L, Luo C, Zhou X, Yu C (2022) A survey of target detection and recognition methods in underwater turbid areas. Appl Sci 12:4898. https://doi.org/10.3390/app12104898
43. Nair RS, Agrawal R, Domnic S, Kumar A (2021) Image mining applications for underwater environment management - A review and research agenda. https://doi.org/10.1016/j.jjimei.2021.100023