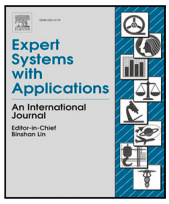
Expert Systems With Applications 268 (2025) 126314

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa



Adaptive deep learning framework for robust unsupervised underwater image enhancement

Alzayat Saleh [a], Marcus Sheaves [a], Dean Jerry [a,b], Mostafa Rahimi Azghadi [a,b,*]

[a] College of Science and Engineering, James Cook University, Townsville, QLD, Australia
[b] ARC Research Hub for Supercharging Tropical Aquaculture through Genetic Solutions, James Cook University, Townsville, QLD, Australia

ARTICLE INFO

Keywords:
Computer vision
Convolutional neural networks
Underwater image enhancement
Variational autoencoder
Machine learning
Deep learning

ABSTRACT

One of the main challenges in deep learning-based underwater image enhancement is the limited availability of high-quality training data. Underwater images are often difficult to capture and typically suffer from distortion, colour loss, and reduced contrast, complicating the training of supervised deep learning models on large and diverse datasets. This limitation can adversely affect the performance of the model. In this paper, we propose an alternative approach to supervised underwater image enhancement. Specifically, we introduce a novel framework called Uncertainty Distribution Network (UDnet), which adapts to uncertainty distribution during its unsupervised reference map (label) generation to produce enhanced output images. UDnet enhances underwater images by adjusting contrast, saturation, and gamma correction. It incorporates a statistically guided multicolour space stretch module (SGMCSS) to generate a reference map, which is utilized by a U-Net-like conditional variational autoencoder module (cVAE) for feature extraction. These features are then processed by a Probabilistic Adaptive Instance Normalization (PAdaIN) block that encodes the feature uncertainties for the final image enhancement. The SGMCSS module ensures visual consistency with the input image and eliminates the need for manual human annotation. Consequently, UDnet can learn effectively with limited data and achieve state-of-the-art results. We evaluated UDnet on eight publicly available datasets, and the results demonstrate that it achieves competitive performance compared to other state-of-the-art methods in both quantitative and qualitative metrics. Our code is publicly available at https://github.com/alzayats/UDnet.

1. Introduction

The enhancement of underwater images is a critical task in computer vision, with applications ranging from underwater robotics to marine biology. However, this task presents unique challenges due to the complex optical properties of water, such as random distortion, low contrast, and wavelength-dependent absorption (Ji et al., 2024). These factors result in colour casts, blurriness, and uneven illumination, making underwater images inherently difficult to process and analyse, see Fig. 1. Addressing these challenges is crucial for improving the accuracy and reliability of tasks like object detection and target recognition in underwater environments.

Over the years, various approaches have been proposed to enhance underwater images. Traditional methods, such as histogram equalization and contrast stretching, attempt to improve image visibility by redistributing pixel intensities or enhancing specific features. While these methods are computationally efficient, they often fail to address the unique complexities of underwater environments, such as non-uniform lighting and scattering effects. In contrast, deep learning-based techniques have shown great promise, leveraging large datasets to learn complex representations for image enhancement. Supervised approaches, such as those employing U-Net architectures and generative adversarial networks (GANs) (Zheng et al., 2024), have achieved significant improvements in underwater image quality. However, these methods rely heavily on paired training data — underwater images and their corresponding ground truth — which are challenging and costly to acquire in underwater scenarios.

Despite these advancements, the existing methods face critical limitations. Traditional approaches lack adaptability to diverse underwater conditions, while supervised learning techniques are constrained by their dependence on annotated datasets and their potential for overfitting to specific domains. Furthermore, many deep learning methods struggle to generalize effectively to new datasets, limiting their applicability in real-world underwater environments (Cheng et al., 2024).

* Correspondence to: College of Science and Engineering, James Cook University, 1 James Cook Drive, QLD 4814, Australia.
E-mail addresses: alzayat.saleh@jcu.edu.au (A. Saleh), marcus.sheaves@jcu.edu.au (M. Sheaves), dean.jerry@jcu.edu.au (D. Jerry), mostafa.rahimiazghadi@jcu.edu.au (M. Rahimi Azghadi).

https://doi.org/10.1016/j.eswa.2024.126314
Received 13 August 2024; Received in revised form 23 December 2024; Accepted 25 December 2024
Available online 2 January 2025

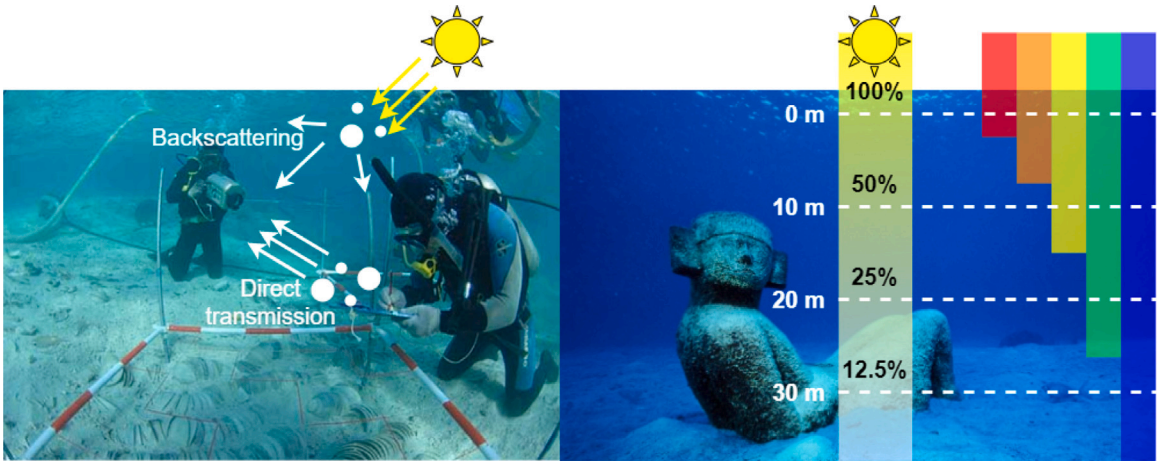


Fig. 1. (Left) Natural light entering the water is scattered multiple times, forming the backscattering for the underwater scene. The light directly reflected off objects in the scene also travels to the camera, and the total light perceived is the sum of these two components, creating the colours and details in underwater images. (Right) Different wavelengths of light are absorbed and scattered differently as they travel through water. Blue light travels the longest distance due to its shorter wavelength, making underwater objects appear blue in colour.

To address these limitations, we propose a novel unsupervised framework for underwater image enhancement, termed the Uncertainty Distribution Network (UDNet). This paper presents UDNet, a novel unsupervised deep learning framework for robust underwater image enhancement. Unlike supervised methods that rely on large datasets of paired raw and enhanced underwater images, UDNet learns without using any ground truth or manually annotated images. Instead, it leverages a probabilistic approach that embraces the uncertainty inherent in underwater image enhancement. Unlike traditional methods, UDNet eliminates the need for paired training data, making it more practical and scalable for underwater applications. Our method introduces adaptive enhancement through uncertainty modelling, leveraging a Statistically Guided Multi-Colour Space Stretch (SGMCSS) module to generate diverse reference maps. During training, the model randomly selects from these maps, enabling it to learn robust and generalized representations of underwater environments. Additionally, UDNet incorporates a Probabilistic Adaptive Instance Normalization (PAdaIN) block, which enhances its ability to adapt to varying image characteristics, further improving its robustness.

While previous work, such as Fu et al. (2022b), has successfully demonstrated the use of uncertainty-inspired methods for underwater image enhancement, these approaches often focus on specific types of uncertainty, such as pixel-level deviations. In contrast, our proposed UDNet introduces a broader framework that models distributional uncertainty through adaptive reference selection and probabilistic feature normalization. These innovations allow our method to generalize effectively across varied underwater environments without requiring paired training data. By addressing key limitations of Fu et al. (2022b) approach, such as its reliance on predefined uncertainty maps, our method achieves state-of-the-art performance while maintaining a lightweight and practical design for real-world applications.

The advantages of UDNet are evident in its performance. It demonstrates strong generalizability across multiple underwater datasets, even those it was not trained on, highlighting its applicability to a wide range of scenarios. Moreover, both qualitative and quantitative evaluations show that UDNet achieves state-of-the-art results, outperforming or matching existing supervised and unsupervised methods.

In summary, the contributions of this work are as follows:

- We introduce UDNet, an unsupervised framework for underwater image enhancement that eliminates the need for paired training data, addressing a key limitation of existing methods.
- We propose an SGMCSS module for generating diverse reference maps, combined with a PAdaIN block for adaptive enhancement, enabling robust and generalized performance across diverse underwater conditions.
- We demonstrate that UDNet achieves state-of-the-art performance on multiple datasets, showcasing its robustness and applicability in various underwater scenarios.

The remainder of this paper is organized as follows: In Section 2, we review related work on underwater image enhancement, highlighting the strengths and limitations of existing methods. Section 3 details the architecture of our proposed framework. In Section 4, we describe our experimental setup and present the results. Finally, we discuss our findings in Section 5 and conclude in Section 6.

2. Related work

Underwater image enhancement is a challenging and active area of research, with various approaches proposed to address issues such as low contrast, distortion, and uneven illumination (Liu et al., 2024). These approaches can be broadly categorized into four main groups: prior-based methods, model-free methods, deep learning-based methods, and probabilistic-based methods.

2.1. Prior-based methods

Prior-based methods rely on physical models of underwater image formation to estimate the optical parameters affecting underwater images. These parameters are then reversed to reconstruct enhanced images. Examples of visual cues used in such methods include the red channel prior (Huang et al., 2018), the underwater dark channel prior (Drews et al., 2013), and the underwater light attenuation prior (Song et al., 2018). While these methods leverage physical insights, their effectiveness can be limited in highly complex underwater environments where the assumptions of the underlying models may not hold.

2.2. Model-free methods

Model-free methods enhance images without explicitly modelling the degradation process, instead focussing on improving image visibility through redistribution of pixel intensity or feature enhancement. Common techniques in this category include contrast-limited adaptive histogram equalization (CLAHE), white balance (WB), and Retinex-based methods. These approaches are computationally efficient and can be extended using fusion-based or multi-scale strategies for improved performance (Drews et al., 2013).

Recently, various enhancements within this category have been proposed, including a hybrid whale optimization algorithm designed for the enhancement of contrast (Braik, 2024) and detail in colour images and a fusion-based approach combining adaptive colour correction with improved contrast enhancement strategies for the improvement of underwater image quality (Raveendran et al., 2024). Other methods include a histogram equalization model specifically developed for colour image contrast enhancement (Wang & Yang, 2024), a technique utilizing interval-valued intuitionistic fuzzy sets to refine colour image quality (Jebadass & Balasubramaniam, 2024), and an intelligent underwater image enhancement method that integrates colour correction with contrast stretching (Lei et al., 2024).

These methods primarily (Lei et al., 2024) aim to improve contrast and colour balance, which are critical for underwater image enhancement. Despite their effectiveness in addressing these aspects, model-free methods often struggle to adapt to the diverse and dynamic underwater conditions. Challenges such as non-uniform lighting, scattering effects, and the varying turbidity of underwater environments remain significant limitations of these approaches. Consequently, while model-free methods are valuable for certain applications, they may benefit from integration with more adaptive or data-driven approaches to handle complex underwater imaging scenarios effectively.

2.3. Deep learning-based methods

Deep learning-based methods utilize training data to automatically learn representations for underwater image enhancement. These methods can be divided into convolutional neural networks (CNNs) and generative adversarial networks (GANs) (Zheng et al., 2024). For example, CNN-based models have been developed using encoder–decoder frameworks to remove noise from underwater images, while lightweight CNN architectures incorporate scene-specific information to synthesize degraded images (Sharma et al., 2023). GANs, on the other hand, have been used to generate synthetic underwater images in an unsupervised manner and to train enhancement networks using synthetic data (Wang et al., 2023). While these methods demonstrate significant improvements, they typically rely on large paired datasets (underwater images and corresponding ground truth), which are challenging to obtain, and their performance may be limited to specific training domains.

2.4. Probabilistic-based methods

Probabilistic-based methods integrate uncertainty modelling into deep learning frameworks, providing a principled way to address disturbances, modelling errors, and uncertainties inherent in underwater environments. Conditional variational autoencoders (cVAEs) represent a notable example in this category. Variational autoencoders (VAEs) are generative models comprising an encoder that maps input data to a low-dimensional latent space and a decoder that reconstructs the data from this latent representation (Kingma & Welling, 2019). VAEs differ from traditional encoders by describing probability distributions for latent variables rather than single-point estimates, enabling them to capture diverse data characteristics. To effectively train VAEs, regularization and reconstruction losses are applied to ensure compact and meaningful representations of the input data (Sohn et al., 2015).

VAEs and cVAEs have found applications in underwater image enhancement and related tasks. For instance, they have been used for background modelling in salient object detection (Li et al., 2019), motion sequence generation (Yan et al., 2018), and image denoising (Balakrishnan et al., 2019). Recent advancements have further combined VAEs with contrastive learning to identify and enhance salient features, showcasing their versatility in various scenarios.

Fu et al. (2022b) introduced an uncertainty-inspired underwater image enhancement framework that leverages learned uncertainty maps for image refinement. While their work effectively improves image quality by modelling pixel-level uncertainty, our approach extends this concept by incorporating adaptive uncertainty modelling through random reference selection during training. Furthermore, our Probabilistic Adaptive Instance Normalization (PAdaIN) layer aligns latent feature distributions dynamically, which enables more robust generalization across diverse underwater datasets. Unlike Fu et al. (2022b), who primarily focused on pixel-wise uncertainty, our method captures and leverages distributional uncertainty to achieve enhanced generalization and adaptability.

In our approach, we extend these probabilistic methods by integrating conditional VAEs with uncertainty modelling techniques, enabling robust underwater image enhancement. Our framework leverages probabilistic adaptive instance normalization to learn diverse and generalized representations of underwater environments. The experimental results demonstrate that this approach significantly improves performance across diverse datasets, addressing key limitations of existing methods.

3. Method

This section describes the various components and concepts utilized to build our Uncertainty Distribution Network (UDnet). As shown in Fig. 2, UDnet is composed of three abstract building blocks including a reference map generation block that uses a statistically guided multi-colour space stretch module, a feature extractor block that uses a cVAE, and a probabilistic adaptive instance normalization block. All of these blocks and their underlying components and concepts will be discussed in detail below, however, the reader is encouraged to investigate the full detail of our implementation code at https://github.com/alzayats/UDnet.

3.1. Uncertainty distribution

UDNet learns to adapt to the uncertainty distribution inherent in underwater image enhancement, where different images require varying degrees of enhancement in terms of contrast, saturation, gamma correction, and other factors. This uncertainty distribution refers to the inherent ambiguity that can exist in the image enhancement process, as different images need different types of enhancements. The main idea behind `UDNet` is to better incorporate this uncertainty in the enhancement process. This is motivated by the fact that the true clean image is often unavailable and that there is a degree of uncertainty in the labels used to train image enhancement models. Existing deterministic learning-guided methods (Sharma et al., 2023) are unable to capture this uncertainty and therefore must make compromises between different possible enhancement results.

To address this, `UDNet` employs a probabilistic framework that introduces uncertainty during the training process. Instead of relying on a fixed set of labels generated from other enhancement techniques, `UDNet` randomly selects one of three enhanced reference maps generated by applying contrast adjustment, saturation adjustment, or gamma correction to the input image. This random selection forces the model to learn a more robust and generalized representation of underwater image enhancement by accounting for the inherent ambiguity in defining the ideal enhancement.

UDnet uses an implicit variable $\mathbf{z}$ to represent the uncertainty in the enhancement process. This variable could represent human subjective preferences, or the parameters of the camera or enhancement algorithms used to capture or generate the ground truth images, which could affect the outcome of the enhancement process. By taking this uncertainty into account, UDnet is able to more accurately capture the range of possible enhancements, rather than trying to determine a single "correct" result. This is particularly useful in situations where the true, unaltered image is not available or cannot be accurately reproduced.

The goal of UDNet is to learn a mapping from the low-quality input image $\mathbf{x}$ to the clean image $\mathbf{y}$ that takes into account the uncertainty

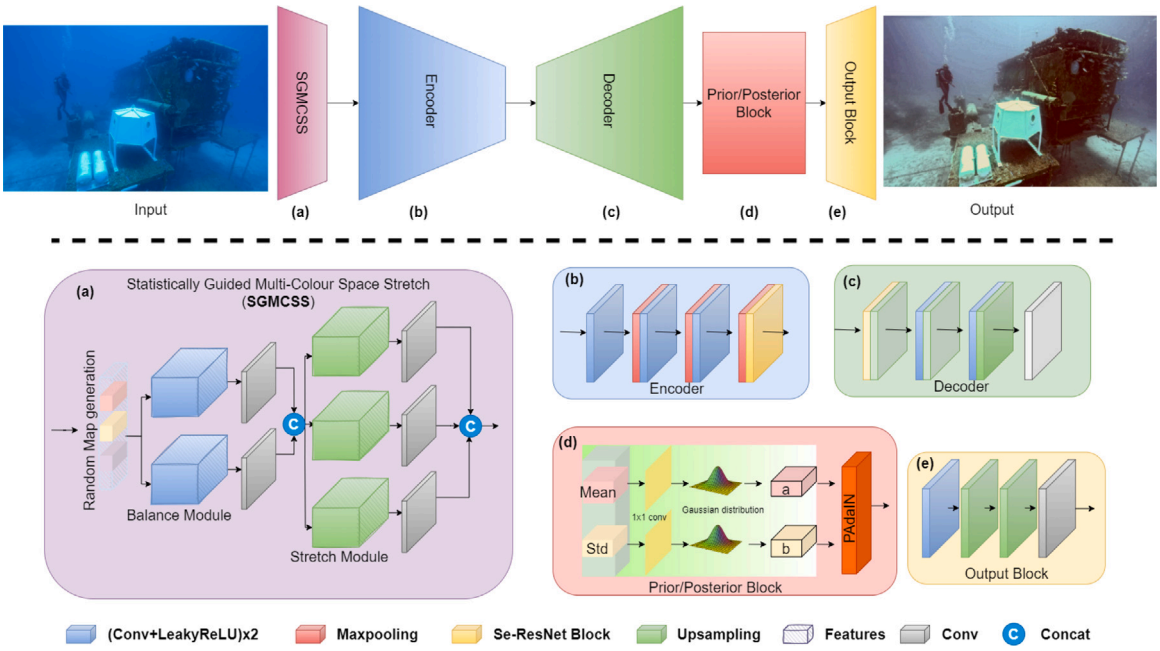


Fig. 2. The architecture of UDnet is composed of five primary components: (a) Statistically Guided Multi-Colour Space Stretch (SGMCSS) for reference maps generation, (b) Encoder for feature extraction, (c) Decoder for image reconstruction, (d) Prior/Posterior Block to calculate and sample Gaussian distributions for feature adaptation, and (e) Output Block that generates the enhanced underwater image. The SGMCSS module transforms the input degraded image into balanced and stretched reference maps, enabling robust feature extraction and enhancement across varying underwater image conditions. The Encoder and Decoder process the input through multi-layer convolutional operations, while the PAdaIN-enabled Prior/Posterior Block ensures stochasticity and flexibility in image enhancement.

represented by $\mathbf{z}$. This can be formalized as follows:

$$p(y|\mathbf{x}) \approx p\left(\mathbf{y}|\mathbf{z}_{\max}, \mathbf{x}\right), \mathbf{z}_{\max} \sim p(\mathbf{z}|\mathbf{x}), \tag{1}$$

where $p(\mathbf{z}|\mathbf{x})$ denotes the distribution of uncertainty, and $\mathbf{z}_{\max}$ denotes the sample with the maximum probability.

Eq. (1) represents the probabilistic framework underlying UDnet. In this equation, $p\left(\mathbf{y}|\mathbf{z}_{\max}, \mathbf{x}\right)$ is the probability of the clean image $\mathbf{y}$ given the sample with the maximum probability $\mathbf{z}_{\max}$ and the low-quality input observation $\mathbf{x}$. $p(\mathbf{z}|\mathbf{x})$ is the probability of the uncertainty variable given the observation. The goal of the model is to learn these probability distributions from the training data and then use them to generate enhanced images that incorporate uncertainty into the enhancement process. By doing so, UDnet is able to (1) provide users with multiple alternative enhancement results to choose from, or (2) improve the accuracy and reliability of the final enhancement result by taking the enhancement sample with the maximum probability as the final estimation, without user intervention.

In the proposed UDnet framework, randomness is introduced in the creation of pseudo-enhanced images through a statistically guided multi-colour space stretch (SGMCSS) module. This module randomly selects one of the contrast, saturation, or gamma-corrected versions of the input image and applies a random colour space stretch to it. The colour space stretch is guided by the statistical properties of the input image, which ensures visual consistency with the raw input image. This process generates a reference map that is used by the U-Net-like conditional variational autoencoder (cVAE) module to extract features for feeding to the probabilistic adaptive instance normalization (PAdaIN) block that encodes feature uncertainties for final enhanced image generation. The randomness introduced in the SGMCSS module helps to create diverse pseudo-enhanced images, which can improve the generalization ability of the model and reduce overfitting. During training, the cVAE network is designed and trained to minimize the reconstruction loss and the KL divergence loss. The reconstruction loss measures the difference between the input image and the reconstructed image, while the KL divergence loss measures the difference between the learned prior distribution of the latent code and the standard normal distribution. During testing, the cVAE network is used to sample from the learned prior distribution of the latent code to generate new images.

3.2. Comparison with uncertainty-inspired underwater image enhancement

This subsection provides a detailed comparison between the proposed UDNet framework and the uncertainty-inspired underwater image enhancement method presented by Fu et al. (2022b) to highlight the unique contributions of our approach. Both methods utilize uncertainty to enhance underwater images; however, their approach to modelling, leveraging, and learning from uncertainty is markedly different.

Similarities

- Both UDNet and Fu et al. (2022b)'s method address the challenge of underwater image enhancement by incorporating **uncertainty modelling**. Both approaches recognize that the enhancement process has inherent uncertainty due to factors such as varied lighting conditions, water turbidity, and the presence of diverse underwater environments.

Differences

- **Uncertainty Modelling**: UDNet employs a framework that models *distributional uncertainty* through adaptive reference selection and probabilistic feature normalization. This approach considers the uncertainty inherent in the potential distributions of enhanced images. In contrast, Fu et al. (2022b)'s method focuses on modelling *pixel-level uncertainty*, using learned uncertainty maps to refine the enhanced images and quantify uncertainty on a per-pixel basis.
- **Reference Map Generation**: UDNet uses a *Statistically Guided Multi-Colour Space Stretch (SGMCSS) module* to generate its reference maps. This module adaptively adjusts the contrast and saturation of the input image and applies gamma correction to generate diverse reference images that are statistically consistent

with the input. In contrast, Fu et al. (2022b)'s method relies on existing UIE algorithms to generate a set of potential reference images.

- **Learning Approach**: UDNet uses a fully *unsupervised approach*, learning from randomly selected, statistically guided reference images. This means it does not require manually annotated or paired training data. Fu et al. (2022b), on the other hand, employ a *supervised learning approach*, which requires a manually created set of potential reference images for training.

3.3. Reference maps generation

The main challenge when training deep learning networks for underwater image enhancement is the limited availability of reference maps (labels) for degraded input images. To address this issue, we autogenerated reference maps based on Underwater Image Enhancement Benchmark Dataset (UIEBD) (Li et al., 2020), which contains real-world underwater images and corresponding reference maps generated using 12 state-of-the-art enhancement algorithms.

3.3.1. Autogeneration of three reference maps from the input image

In the original UIEBD, volunteers were asked to compare the enhanced results and subjectively select the best one as the final reference image. However, our reference map generation process uses the same intuition without human intervention. Using the degraded input image (original image), as shown in the first step of Fig. 2, we generate three enhanced reference maps by three enhancement algorithms, one of which is randomly selected to introduce uncertainty into our training dataset. "randomly selected" here refers to the three enhanced reference maps generated by three enhancement algorithms using the degraded input image. One of these three maps is randomly selected to introduce uncertainty into the training dataset. It is worth mentioning that adding more enhanced reference maps did not increase the model accuracy as discussed in more detail in Section 4.8.

The three methods that we chose to introduce uncertainty into the dataset were contrast and saturation adjustment, as well as gamma correction on the original images. These methods were chosen because they can effectively simulate the distortions commonly found in underwater images, such as changes in contrast, saturation, brightness, and colours. As shown in Eq. (2), the contrast and saturation adjustment was performed using a linear transformation formula, where the adjustment coefficient α was the same for all pixels for contrast adjustment and varied for each pixel for saturation adjustment.

$$y = (x - m) \times \alpha + x, \tag{2}$$

where x and y refer to the degraded and enhanced images, respectively, m denotes the mean of each channel, and α is the adjustment coefficient.

In Eq. (2), the adjustment coefficient α plays distinct roles in contrast and saturation adjustments. For contrast adjustment, α remains constant for all pixels, ensuring a uniform contrast enhancement across the entire image. Conversely, for saturation adjustment, α varies for each pixel, enabling fine-grained and localized adjustments to saturation levels. This pixel-wise variation allows for targeted enhancement of specific image areas without influencing others. This difference in α application is crucial, as it results in either global contrast enhancement or localized saturation adjustments, both contributing to a more precise and controlled image enhancement process.

Our approach has several advantages, including saving time and increasing reliability compared to using human observers to generate reference maps. We evaluated the effectiveness of the generated reference maps in Section 4.5 and Section 4.6 by comparing the enhanced results to the subjective selections made by volunteers in the original UIEBD dataset. Our goal was to create uncertain labels that would reflect the uncertainty in the ground truth recording, rather than significantly altering the original labels. To achieve this, we utilized a Statistically Guided Multi-Colour Space Stretch (SGMCSS) or Colour Correction module.

3.3.2. Statistically guided multi-colour space stretch for colour correction

To improve the visual quality of the reference images used as pseudo-labels, a multi-scale statistically guided multi-colour space stretch module is developed. The term 'multi-scale' refers to the different levels of abstraction in the feature extraction process. The goal of this module is to improve the colour and contrast of the randomly chosen reference maps, which guide the network's unsupervised learning process.

This is obtained by transforming the reference map Red Green Blue (RGB) values to the optimal RGB values, which involves determining the proper camera white-balance for colour-neutral subjects, as well as removing the effects of lens flare and red–green chromatic aberration. This could be useful when dealing with oversaturated images. The SGMCSS is designed for the case where the mean and standard deviation of the red green and blue colour values are known. This module uses a non-parametric approach to colour correction (Xiao et al., 2022), which is able to accommodate new statistical distributions of the pixel values in the red, green and blue colour channels. The SGMCSS consists of two main components: a dual-statistic balance module and a multi-colour space stretch module.

In the dual-statistic balance module, the image is processed by two different modules that use statistics of the image (average and maximum values) to correct its colour balance. The output is then enhanced using two residual-enhancement modules to recover lost details.

The first residual-enhancement module is based on Grey World (GW) theory. The Gray World theory is a method for colour correction in images. It is based on the assumption that the average colour of objects in a perfect image is grey, which means that the average values of the R, G, and B channels are equal. This means that the scale factors for each channel, e_R, e_G, and e_B, can be determined using the GW theory:

$$x^{GW} = Conv_{1\times1}(x) \circ \overline{A}, \tag{3}$$

where $\overline{A} = [\frac{1}{A_R}, \frac{1}{A_G}, \frac{1}{A_B}] \in \mathbb{R}^{3\times1}$, A_c denotes the average value of c channel in the original image, and $\circ$ denotes pixel-wise multiplication.

A 1×1 convolution operation ($\mathrm{Con}_{1\times1}(x)$) is used to reduce the number of channels in the input image or to combine information from different channels. In this case, it is used to adjust the contrast, saturation, and gamma correction of the raw underwater image. and then multiplied element-wise ($\circ$) with a matrix $\bar{A}$ to obtain the output image (x^{GW}).

The second residual-enhancement module is based on the White Patch (WP) algorithm. The White Patch algorithm is another method for colour correction in images. It is based on the assumption that the maximum response of the RGB channels in an image is caused by a white patch in the scene. This white patch is assumed to reflect the colour of the light in the scene, so the largest value in the RGB channels is used as the source of light. Based on this hypothesis, the scale factors for each channel can be expressed as:

$$x^{WP} = Conv_{1\times1}(x) \circ \overline{M}, \tag{4}$$

where $\overline{M} = [\frac{1}{M_R}, \frac{1}{M_G}, \frac{1}{M_B}] \in \mathbb{R}^{3\times1}$, M_c denotes the maximum value of c channel in original image.

The two residual-enhancement results are merged and passed to the stretch module as follows:

$$x^{DSB} = Conv_{3\times3}(x^{GW}) + Conv_{3\times3}(x^{WP}), \tag{5}$$

where x^{DSB} represents the result enhanced by the dual-statistic balance module.

In the multi-colour space stretch module, the image is transformed into different colour spaces (HSI and Lab) and processed by a trainable module to improve contrast. The original image is also enhanced and added to the stretched version as follows:

$$x^{final} = Conv_{3\times3}(x^{r}) + Conv_{3\times3}(x^{h}) + Conv_{3\times3}(x^{l}), \tag{6}$$

Where x^r, x^h, x^l denote the histogram stretched pixel value in RGB, HSI, and Lab colour spaces, respectively.

In the RGB colour space, the red, green, and blue channels are individually stretched based on their statistical properties. In the HSI colour space, only the saturation (S) and intensity (I) channels are stretched, while the hue (H) channel is preserved. For the Lab colour space, the a and b channels, representing colour-opponent dimensions, are stretched, while the L channel, representing lightness, is maintained.

The output is then converted back to the RGB colour space and merged together by going through 3×3 convolutional layer and pixel-wise add up. Overall, this technique can improve the visual quality of the reference map that will be passed to the next building block of UDNet, i.e. the feature extractor module (see Fig. 2), by correcting colour balance and enhancing contrast.

3.4. Feature extraction

The next abstract building block of UDNet, as shown in Fig. 2, is its feature extractor block. UDnet uses a two-branch U-Net-based feature extractor to map the input images to representations. These representations are then fed into the PAdaIN module, which transforms the enhancement statistics of the input to create the enhanced image as explained in 3.5.

The training branch of the feature extractor is used to construct posterior distributions using the raw original underwater image and its corresponding reference map image as inputs. The test branch, on the other hand, is used to estimate the prior distribution of a single raw underwater image.

The PAdaIN block is used to encode the uncertainty in the input image, allowing UDNet to generate multiple enhanced versions of the image that capture the different possible interpretations of the original image. To achieve this, UDnet uses a prior/posterior block to build the distribution of possible enhancements. This block is designed to construct both a mean and a standard deviation distribution, using 1×1 convolutions to transform the input data matrix into a series of distributions that capture the uncertainty in the input image.

In the training stage, the input image and its corresponding reference image are used to learn the `posterior` distributions of the latent codes as follows:

$$\boldsymbol{a} \sim \mathcal{N}_{\text{mean}}\left(\boldsymbol{\mu}(\boldsymbol{y},\mathbf{x}), \boldsymbol{\sigma}^2(\boldsymbol{y},\mathbf{x})\right), \tag{7}$$

$$\boldsymbol{b} \sim \mathcal{N}_{\text{std}}\left(\boldsymbol{m}(\boldsymbol{y},\mathbf{x}), \boldsymbol{v}^2(\boldsymbol{y},\mathbf{x})\right), \tag{8}$$

where $\boldsymbol{a}$ and $\boldsymbol{b}$ are two random samples from the mean and standard deviation posterior distributions, $\mathcal{N}_{\text{mean}}$ and $\mathcal{N}_{\text{std}}$ are the N-dimensional Gaussian distribution of the mean and standard deviation, and $\boldsymbol{y}$ and $\boldsymbol{x}$ are the reference image and the raw input image, respectively.

Eq. (7) represents the distribution of the feature activations (a) in the UDnet's conditional variational autoencoder (cVAE) module. The distribution is assumed to be normal (represented by the symbol $\mathcal{N}$) with mean $\mu(\boldsymbol{y},\mathbf{x})$ and variance $\sigma^2(\boldsymbol{y},\mathbf{x})$. The mean and variance are functions of the input image ($\mathbf{x}$) and the reference map ($\boldsymbol{y}$) generated by the statistically guided multi-colour space stretch (SGMCSS) module.

Eq. (8) represents the distribution of the feature uncertainties (represented by the vector $\boldsymbol{b}$) in the UDnet's probabilistic adaptive instance normalization (PAdaIN) block. The distribution is also assumed to be normal, but with a different subscript s to distinguish it from the distribution in the cVAE module. The mean and variance of the distribution are functions of the input image and the reference map, represented by $\boldsymbol{m}(\boldsymbol{y},\mathbf{x})$ and $\boldsymbol{v}^2(\boldsymbol{y},\mathbf{x})$, respectively.

Overall, these equations describe the probabilistic nature of UDnet, which allows it to model the uncertainty in the input data and generate enhanced images that are consistent with the input image and reference map. The use of probabilistic distributions also enables UDnet to learn from a limited amount of data without the need for manual human annotation.

Once these distributions have been constructed, random samples are extracted from them and injected into the PAdaIN module, where they are used to transform the statistics of the received features.

In the testing stage, the latent codes generated for PAdaIN are determined only by the input image to learn the `prior` distributions of the latent codes as follows:

$$\boldsymbol{a} \sim \mathcal{N}_{\text{mean}}\left(\boldsymbol{\mu}(\mathbf{x}), \boldsymbol{\sigma}^2(\mathbf{x})\right), \tag{9}$$

$$\boldsymbol{b} \sim \mathcal{N}_{\text{std}}\left(\boldsymbol{m}(\mathbf{x}), \boldsymbol{v}^2(\mathbf{x})\right), \tag{10}$$

where $\boldsymbol{a}$ and $\boldsymbol{b}$ are two random samples from the mean and standard deviation prior distributions, $\mathcal{N}_{\text{mean}}$ and $\mathcal{N}_{\text{std}}$ are the N-dimensional Gaussian distribution of the mean and standard deviation, respectively and $\boldsymbol{x}$ is the raw input image.

The UDnet model is applied multiple times to the same input image in order to generate multiple enhancement variants. This is done by re-evaluating only the PAdaIN module and the output block, without retraining the entire model, which makes UDNet very efficient. The resulting diverse enhancement samples are then used for Maximum Probability estimation that takes the enhancement sample with the maximum probability as the final estimation.

3.4.1. Loss function

The training process for UDnet follows the standard procedure for training a cVAE model, which involves minimizing the variational lower bound. However, our approach has an additional step of finding a meaningful embedding of enhancement statistics in the latent space. This is achieved through the use of a posterior network (as shown in Fig. 2), which learns to recognize posterior features and map them to posterior distributions of the mean and standard deviation. Random samples from these distributions can be used to formalize the enhanced results. This approach allows for the incorporation of uncertainty into the enhancement process, which can improve the accuracy and reliability of the resulting images.

During the training process, the PAdaIN module is used to predict the enhanced image by receiving random samples $\boldsymbol{a}$ and $\boldsymbol{b}$ from Eq. (7) and Eq. (8), respectively. The enhancement loss (Eq. (11)) is calculated based on the differences between the predicted image and the reference map, and is used to penalize the model if the output deviates from the reference.

$$L_{\text{e}} = L_{\text{mse}} + \lambda L_{\text{vgg16}}, \tag{11}$$

where L_{mse} denotes the mean square error loss and L_{vgg16} denotes the perceptual loss (Johnson et al., 2016), λ refers to a weight parameter.

The mean square error loss L_{mse} and the perceptual loss L_{vgg16} are two common metrics used to evaluate the performance of image enhancement algorithms. The mean square error loss measures the average squared difference between the predicted and reference images, while the perceptual loss, which was introduced by Johnson et al. (2016), measures the differences between the high-level features of the predicted and reference images. The weight λ is used to control the relative importance of these two loss terms in the overall enhancement loss L_{e}. For example, if λ is set to a high value, the model will be more heavily penalized for large differences between the predicted and reference images, while if λ is set to a low value, the model will be less sensitive to such differences. The specific values of λ used in the training process will depend on the characteristics of the dataset and the desired performance of the model.

In addition to minimizing the enhancement loss L_{e}, the training process for UDnet also involves using Kullback–Leibler (KL) divergences D_{KL} to align the posterior distributions with the prior distributions (Eqs. (12) and (13)).

$$L_m = D_{\text{KL}}\left(\mathcal{N}_{\text{mean}}(\mathbf{x}) \,\|\, \mathcal{N}_{\text{mean}}(\boldsymbol{y},\mathbf{x})\right), \tag{12}$$

$$L_s = D_{\text{KL}}\left(\mathcal{N}_{\text{std}}(\mathbf{x}) \,\|\, \mathcal{N}_{\text{std}}(\boldsymbol{y},\mathbf{x})\right), \tag{13}$$

where $\boldsymbol{m}$ and $\boldsymbol{s}$ are the mean and the standard deviation, respectively. KL divergence is a measure of the difference between two probability distributions and can be used to compare the posterior distributions learned by the model with the prior distributions that are assumed to represent the distribution of latent variables in the training data. By minimizing the KL divergences between the posterior and prior distributions, the model is able to learn a more accurate representation of the latent space, which can improve the quality of the enhanced images.

The total loss function used for training UDnet is the weighted sum of the enhancement loss L_e and the KL divergences D_{KL} between the posterior and prior distributions,

$$L = L_e + \beta(L_m + L_s), \tag{14}$$

where β is a weight parameter, whose value depends on the dataset's characteristics and the model's desired performance. By minimizing this total loss function, L, the model is able to learn an effective mapping from the input degraded images to the corresponding enhanced images, while also aligning the posterior and prior distributions in the latent space. This allows the model to generate high-quality enhanced images while also incorporating uncertainty into the enhancement process.

3.5. Probabilistic adaptive instance normalization (PAdaIN)

The final abstract building block of UDNet, as shown in Fig. 2 is PAdaIN block. The goal of UDnet is to adjust the appearance of underwater images, such as the colours and contrasts, without altering the content of the image. This is important because it allows the enhanced images to be more visually appealing and easier to interpret, without compromising the integrity of the original image. Therefore, We use a probabilistic adaptive instance normalization (PAdaIN) to capture these properties.

The core component of UDNet's probabilistic framework is PAdaIN (Fu et al., 2022b), a modified version of the AdaIN algorithm specifically designed for underwater image enhancement. PAdaIN leverages the uncertainty distribution learned during training to encode feature uncertainties, allowing it to generate multiple enhanced versions of the input image. These multiple versions reflect the inherent ambiguity in underwater image enhancement, as there is no single correct enhancement for a given image. By generating a distribution of possible enhancements, UDNet provides a more comprehensive and flexible approach to underwater image enhancement.

Unlike (Fu et al., 2022b), which utilizes static uncertainty maps to guide image enhancement, our method employs adaptive uncertainty modelling by randomly selecting reference images during training. This approach encourages the network to learn from a diverse set of potential image distributions, resulting in a more robust and flexible enhancement framework. Additionally, our PAdaIN layer further refines feature alignment by accounting for variability in the latent space, enabling our model to generalize effectively to datasets with diverse characteristics. These advancements collectively extend (Fu et al., 2022b) method by addressing its limitations in handling diverse underwater environments and ensuring consistent enhancement quality.

However, AdaIN relies on the availability of known content and style images, which is not always the case in underwater image enhancement processes. To address this issue, PAdaIN introduces random samples from the posterior distributions of the mean and standard deviation as the parameters of the AdaIN operation, which can be formulated as:

$$\mathrm{PAdaIN}(\mathbf{x}) = \boldsymbol{b}\left(\frac{\mathbf{x} - \boldsymbol{\mu}(\mathbf{x})}{\boldsymbol{\sigma}(\mathbf{x})}\right) + \boldsymbol{a}, \tag{15}$$

where $\boldsymbol{b}$ and $\boldsymbol{a}$ are two random samples from the posterior distributions of the mean and standard deviation, respectively.

These posterior distributions are learned using a cVAE, which was described in 3.4. This allows PAdaIN to generalize the AdaIN algorithm and apply it to underwater image enhancement without the need for known content and style images. Overall, PAdaIN is able to capture the important appearance-related features of the input image and use them to generate enhanced images that maintain the integrity of the original image.

It is worth noting that in contrast to other approaches that consider the variance of the image, such as GAN, PAdaIN is based on the statistical distribution of the image features, which are invariant to transformations like colour transformation. This is done by conditioning the network on training images and their reference map, which, along with the use of a differentiable approximation of the uncertainty, make UDnet easily trainable with a single backward pass.

Table 1
The datasets used in our research. The numbers represent the amount of images in sets.

Datasets	Train		Test	
	Paired	Unpaired	Paired	Unpaired
EUVP (Islam, Xia, & Sattar, 2020)	3700	3140	515	–
UFO (Jahidul Islam et al., 2020)	1500	–	120	–
UIEBD (Li et al., 2020)	800	–	90	60
DeepFish (Saleh et al., 2020)	–	3200	–	600
FISHTRAC (Mandel et al., 2023)	–	600	–	71
FishID (Lopez-Marcano et al., 2021)	–	7093	–	6897
RUIE (Liu et al., 2020)	–	2904	–	726
SUIM (Islam, Edge, et al., 2020)	–	1525	–	110

4. Experiments

In this section, we perform several experiments to evaluate the performance of our proposed method. We will first describe the utilized datasets, evaluation metrics and implementation details. Then, we quantitatively and qualitatively evaluate our model against 10 popular image enhancement models on 8 public datasets. Finally, we will demonstrate the significance of our work through a visual perception improvement test.

4.1. Datasets

We used eight publicly available datasets for our model's performance verification. These datasets are: EUVP (Islam, Xia, & Sattar, 2020), UFO (Jahidul Islam et al., 2020), UIEBD (Li et al., 2020), DeepFish (Saleh et al., 2020), FISHTRAC (Mandel et al., 2023), FishID (Lopez-Marcano et al., 2021), RUIE (Liu et al., 2020), SUIM (Islam, Edge, et al., 2020). Details of these datasets can be found in Table 1. In EUVP (Islam, Xia, & Sattar, 2020), UFO (Jahidul Islam et al., 2020), and UIEBD (Li et al., 2020), there are many paired images and unpaired images which were divided as shown in Table 1. The paired images are the ones that have ground truth. The rest of the datasets have only unpaired images. In our experiment, we used only UIEBD (Li et al., 2020) for training in an unsupervised way without the ground truth. We used the other datasets for performance evaluation.

4.2. Evaluation metrics

Evaluation metrics for image enhancement are often based on natural image statistics. Perceptual and structural image qualities can be judged in different ways. We employ four full-reference evaluation metrics and three no-reference evaluation metrics for evaluating the quantitative performance of our image enhancement model. Specifically, (1) The full-reference evaluation metrics consist of Peak Signal-to-Noise Ratio (PSNR) (Wang et al., 2004), Structural Similarity (SSIM) (Wang et al., 2004), Most Apparent Distortion (MAD) (Chandler, 2010), and Gradient Magnitude Similarity Deviation (GMSD) (Xue et al., 2014), which are used for paired test sets (EUVP (Islam, Xia, & Sattar, 2020), UFO (Jahidul Islam et al., 2020), UIEBD (Li et al., 2020)). A higher PSNR or a lower MAD score means that the output image and the label

Table 2
Comparison against published works on three *PAIRED* datasets (EUVP (Islam, Xia, & Sattar, 2020), UFO (Jahidul Islam et al., 2020) and UIEBD (Li et al., 2020)). Underwater image enhancement performance metric in terms of average PSNR (Wang et al., 2004), SSIM (Wang et al., 2004), MAD (Chandler, 2010) and GMSD (Xue et al., 2014) values are shown, where (↑) means higher is better and (↓) means lower is better. We represent the best two results in RED and BLUE colours.
* The model trained on UIEBD (Li et al., 2020) dataset with label.

Method	EUVP				UFO				UIEBD			
	PSNR ↑	SSIM ↑	MAD ↓	GMSD ↓	PSNR ↑	SSIM ↑	MAD ↓	GMSD ↓	PSNR ↑	SSIM ↑	MAD ↓	GMSD ↓
CLAHE (Zuiderveld, 1994)	18.97	0.726	138.6	0.090	18.76	0.701	143.7	0.098	20.64	0.821	100.9	0.053
IBLA (Peng & Cosman, 2017)	22.62	0.719	97.78	0.068	20.71	0.671	122.6	0.082	17.56	0.614	141.2	0.126
PIFM* (Chen et al., 2021)	20.17	0.747	113.5	0.0719	20.63	0.728	118.2	0.076	23.62	0.852	80.80	0.056
PUIEnet* (Fu et al., 2022b)	21.01	0.770	94.55	0.052	21.38	0.737	102.3	0.057	23.74	0.844	79.36	0.057
RGHS (Huang et al., 2018)	21.13	0.753	98.40	0.056	20.74	0.730	112.8	0.066	23.57	0.803	81.02	0.053
UCM (Iqbal et al., 2010)	20.91	0.767	99.37	0.062	20.34	0.743	110.0	0.068	22.03	0.815	92.95	0.067
UDCP (Drews et al., 2013)	15.80	0.572	136.8	0.098	15.95	0.561	148.1	0.111	13.47	0.548	139.0	0.118
ULAP (Song et al., 2018)	21.91	0.730	108.4	0.071	21.98	0.729	116.7	0.071	18.95	0.718	113.0	0.085
USLN* (Xiao et al., 2022)	20.87	0.771	94.62	0.050	20.73	0.749	105.0	0.057	24.04	0.849	78.91	0.057
Wavenet* (Sharma et al., 2023)	20.25	0.753	109.3	0.067	20.98	0.736	115.1	0.071	24.61	0.881	68.08	0.045
UDnet (ours)	22.96	0.771	87.67	0.049	22.43	0.738	99.53	0.053	22.23	0.812	74.21	0.043

image are closer in perceptual content, while a higher SSIM or a lower GMSD score means that the two images are more structurally similar. (2) The no-reference evaluation metrics are: Underwater Image Quality Measure (UIQM) (Panetta et al., 2016), Multi-scale Image Quality Transformer (MUSIQ) (Ke et al., 2021), and Natural Image Quality Evaluator (NIQE) (Mittal et al., 2013) which are used for unpaired test sets (DeepFish (Saleh et al., 2020), FISHTRAC (Mandel et al., 2023), FishID (Lopez-Marcano et al., 2021), RUIE (Liu et al., 2020), SUIM (Islam, Edge, et al., 2020)). The UIQM is the linear combination of three underwater image attribute measures: the underwater image colourfulness measure (UICM), the underwater image sharpness measure (UISM), and the underwater image contrast measure (UIConM). A higher UIQM and MUSIQ or a lower NIQE score suggests a better human visual perception. However, it is worth noting that these no-reference metrics cannot accurately reflect the quality of an image in some cases, so scores of UIQM, MUSIQ, and NIQE are only provided as references for our study. We will present enhanced unpaired images in the visual comparisons section for readers to assess.

We chose these metrics because they collectively provide a comprehensive evaluation of our image enhancement model. The full-reference metrics allow us to compare the enhanced images directly with the original images, providing a measure of the fidelity and structural similarity of our enhancements. The no-reference metrics offer an assessment of the image quality from a human visual perception perspective, which is crucial as the ultimate goal of our work is to improve the visual quality of underwater images for human viewers. These metrics together ensure a robust and thorough evaluation of our method. For example, PSNR measures the fidelity of the enhanced image to the original image, while SSIM takes into account the structural information in the image. MAD measures the most apparent distortion between the enhanced and original images, while GMSD measures the similarity of gradient magnitude between the two images. While we acknowledge that no single metric can fully capture the quality of an enhanced image, we believe that our choice of these seven metrics provides a meaningful and comprehensive evaluation of our proposed method.

4.3. Implementation details

Our models were trained with an input resolution of 256 × 256 pixels. We scale the lowest side of the image to 256 and then extract random crops of size 256 × 256. We found that for this problem set, a learning rate of 1×10^{-4} works the best. It took around 500 epochs for the model to train on this problem and the batch size was set as 10. Our networks were trained on a Linux host with a single NVidia GeForce RTX 2080 Ti GPU with 11 GB of memory, using Pytorch framework. The training is carried out with ADAM optimizer, and the loss function, as explained in 3.4.1, is a combination of the Mean Squared Error (MSE) L_{mse}, the perceptual loss L_{vgg16} (Johnson et al., 2016), and Kullback–Leibler (KL) divergences L_{kl}.

In order to boost network generalization, we augment the training data with rotation, flipping horizontally and vertically. Following Fu et al. (2022b), we adopt 1×1 convolutions to broadcast the samples to the desired number of channels before input to PAdaIN with a latent space of a 20-dimensional N.

From an implementation and computational point of view, our proposed method, UDnet, is feasible. We implemented UDnet using PyTorch and trained it on a single NVIDIA GeForce RTX 2080 Ti GPU. The training time and the number of parameters for each component of UDnet are as follows: the SGMCSS module has 1.2 million parameters and takes 1.5 h to train, the cVAE module has 1.2 million parameters and takes 2.5 h to train, and the PAdaIN block has 0.2 million parameters and takes 1 h to train. The entire UDnet model has 2.6 million parameters and takes 5 h to train.

In terms of computational complexity, the inference time for UDnet is 0.03 s per image on average, which is faster than some existing methods for underwater image enhancement.

In conclusion, our results suggest that UDnet is feasible from both an implementation and computational perspective. However, it is important to note that the computational requirements may vary depending on the size and complexity of the input images, as well as the hardware used for training and inference.

4.4. Compared methods

To have a comprehensive and fair evaluation of our model, we compare it to 10 previous studies including six conventional unsupervised methods (CLAHE (Zuiderveld, 1994), IBLA (Peng & Cosman, 2017), RGHS (Huang et al., 2018), UCM (Iqbal et al., 2010), UDCP (Drews et al., 2013), ULAP (Song et al., 2018)) and four deep-learning-based methods (PIFM (Chen et al., 2021), PUIEnet (Fu et al., 2022b), USLN (Xiao et al., 2022), Wavenet (Sharma et al., 2023)). The comparison with conventional unsupervised methods aims to demonstrate the advantages of our trainable unsupervised deep-learning-based method.

We applied these conventional unsupervised approaches directly to the test sets. We used the respective studies' code and training approach for the deep learning-based methods. To guarantee the experiment's objectivity, we trained the four deep-learning-based methods on UIEBD (Li et al., 2020) and applied the author-provided model and network training parameters.

4.5. Quantitative comparisons

The comparison results for all paired test sets are summarized in Table 2. We report the average scores of the four full-reference metrics (PSNR, SSIM, MAD, GMSD). Table 2 demonstrates that our proposed method outperforms all six conventional unsupervised methods and all four deep-learning-based methods in all four full-reference metrics on

Table 3
Comparison against published works on five *UNPAIRED* datasets (DeepFish (Saleh et al., 2020), FISHTRAC (Mandel et al., 2023), FishID (Lopez-Marcano et al., 2021), RUIE (Liu et al., 2020), AND SUIM (Islam, Edge, et al., 2020)).
Underwater image enhancement performance metric in terms of average UIQM (Panetta et al., 2016), MUSIQ (Ke et al., 2021) AND NIQE (Mittal et al., 2013) values are shown, where (↑) means higher is better, and (↓) means lower is better. We represent the best two results in RED and BLUE colours.

Method	DeepFish			FISHTRAC			FishID			RUIE			SUIM		
	UIQM↑	MUSIQ↑	NIQE↓	UIQM↑	MUSIQ↑	NIQE↓	UIQM↑	MUSIQ↑	NIQE↓	UIQM↑	MUSIQ↑	NIQE↓	UIQM↑	MUSIQ↑	NIQE↓
CLAHE (Zuiderveld, 1994)	3.136	25.67	4.10	2.686	48.94	3.78	2.631	37.01	5.20	3.028	34.98	4.48	2.914	58.40	3.66
IBLA (Peng & Cosman, 2017)	1.993	43.34	6.31	1.704	55.36	6.20	1.745	44.45	6.69	2.577	32.60	4.68	1.839	58.17	4.10
PIFM (Chen et al., 2021)	3.275	27.39	4.29	2.892	48.38	4.17	2.424	41.14	5.12	3.087	33.29	4.51	2.694	58.56	3.82
PUIEnet (Fu et al., 2022b)	3.209	28.82	4.10	3.421	48.60	4.22	2.301	40.08	5.32	3.102	28.32	4.56	2.838	60.03	3.75
RGHS (Huang et al., 2018)	3.150	42.26	6.72	1.913	57.03	5.57	1.823	44.46	5.95	2.991	30.49	4.32	2.317	59.55	3.75
UCM (Iqbal et al., 2010)	2.918	41.37	6.45	2.575	54.39	11.25	2.452	44.32	6.06	3.107	32.59	4.19	2.804	58.51	3.90
UDCP (Drews et al., 2013)	2.391	42.42	5.72	1.622	51.01	6.09	1.320	37.97	6.48	2.159	29.79	4.71	1.731	56.14	4.09
ULAP (Song et al., 2018)	2.814	40.46	6.16	1.763	54.37	6.18	2.176	45.12	5.98	2.396	33.00	4.72	2.232	58.36	3.97
USLN (Xiao et al., 2022)	3.015	32.12	4.24	3.316	49.97	4.30	2.067	44.58	6.32	3.068	32.78	4.44	2.682	61.30	4.03
Wavenet (Sharma et al., 2023)	3.304	40.14	3.00	3.071	56.79	3.70	2.252	46.66	5.15	3.081	30.79	4.56	2.773	61.03	3.67
UDnet(ours)	3.292	24.56	4.02	3.390	50.24	3.41	2.303	37.85	5.11	3.154	26.74	4.18	2.875	57.84	3.65

Table 4
Comparison against published works on two *PAIRED* datasets (UIEBD (Li et al., 2020), ANDEUVP (Islam, Xia, & Sattar, 2020)) and two *UNPAIRED* datasets (UCCS (Liu et al., 2020), and UIQS Liu et al., 2020) .
Underwater image enhancement performance metric in terms of average PSNR (Wang et al., 2004), SSIM (Wang et al., 2004), UIQM (Panetta et al., 2016), UCIQE (Panetta et al., 2016), higher values is better. We represent the best two results in RED and BLUE colours.

Method	UIEBD				EUVP				UCCS		UIQS	
	PSNR	SSIM	UIQM	UCIQE	PSNR	SSIM	UIQM	UCIQE	UIQM	UCIQE	UIQM	UCIQE
FIRUA (Yu & Qin, 2023)	27.80	0.849	3.452	0.534	20.65	0.728	2.985	0.316	3.650	0.652	3.223	0.809
RCA-CycleGAN (Wang et al., 2023)	21.28	0.804	2.987	0.280	21.32	0.829	2.834	0.337	2.942	0.671	2.975	0.832
IEFD (Liu, Dong, et al., 2022)	22.01	0.793	3.234	0.317	21.17	0.705	3.148	0.363	3.199	0.598	3.379	0.827
RFHP (Fu, Lin, et al., 2022)	20.31	0.841	3.189	0.598	20.91	0.649	2.932	0.299	3.345	0.627	3.788	0.801
Two-step DA (Jiang et al., 2022)	21.78	0.802	3.402	0.624	20.37	0.711	3.105	0.208	3.724	0.674	3.418	0.913
MCACE (Zhang et al., 2022)	20.69	0.784	3.293	0.587	20.89	0.698	2.927	0.309	3.933	0.587	2.851	0.856
Twin ACL (Liu, Jiang, et al., 2022)	22.30	0.888	3.595	0.683	21.04	0.724	3.029	0.354	3.953	0.688	3.032	0.897
MNIAM (Ji et al., 2024)	23.11	0.833	3.590	0.607	21.49	0.752	3.251	0.619	3.719	0.653	3.111	0.848
LFT-DGAN (Zheng et al., 2024)	24.45	0.828	3.442	0.593	21.67	0.741	3.109	0.610	3.618	0.645	3.289	0.832
FMTformer (Xiang et al., 2025)	22.89	0.812	3.393	0.621	20.67	0.807	3.014	0.598	3.567	0.633	3.218	0.847
FDCE-Net (Cheng et al., 2024)	23.87	0.917	3.561	0.612	26.17	0.893	3.214	0.645	3.833	0.681	4.253	0.598
Zero-UMSIE (Liu et al., 2024)	23.45	0.841	4.837	0.621	22.89	0.803	3.104	0.649	3.793	0.671	3.672	0.879
UDnet(ours)	22.23	0.812	3.781	0.745	22.96	0.771	3.265	0.749	3.974	0.713	3.154	0.958

the EUVP dataset and shows great performance on the UFO dataset. Our model achieves the highest PSNR, SSIM scores on EUVP, and the lowest MAD, GMSD scores on EUVP, UFO. In addition, we also found that

- Although our model was trained in an unsupervised way, it still outperformed the fully supervised deep-learning-based models trained on UIEBD dataset on EUVP and UFO.
- This result shows that our proposed unsupervised deep-learning-based method is better than conventional ones at preserving structural information and contrast preservation, which suggests the superiority of our trainable model.
- Our model's performance is slightly lower than fully supervised methods on UIEBD dataset. However, this is because fully supervised methods were trained on the images and ground truth labels acquired from the UIEBD, while our method is fully unsupervised and was not trained with ground truth labels.
- Even without any extra labels, our model outperforms models that are trained on the UIEBD dataset on the two metrics of MAD and GMSD for paired datasets.

We also provided quantitative comparisons for unpaired test sets in Table 3, which demonstrate that our model achieves the highest NIQE on FISHTRAC, FishID, RUIE, SUIM, and the second-best UIQM score on DeepFish, FISHTRAC, SUIM. These results also show that

- Deep-learning-based models cannot outperform conventional approaches in no-reference evaluation metrics, in contrast to full-reference evaluation metrics.
- The quantitative results suggest that our method can generalize well on unseen datasets even without ground truth.

Our model shows slight performance variation across different datasets and evaluation metrics in Table 3. This can be attributed to the inherent challenges of underwater image enhancement, including the variability of underwater conditions and the limited availability of high-quality training data. It is crucial to acknowledge that no-reference metrics, including UIQM, MUSIQ, and NIQE, may not consistently provide an accurate representation of image quality under certain circumstances. Consequently, the scores derived from these metrics are utilized solely as reference points within our study. Even despite the minor performance variation, our model still outperforms several studies and is on-par with the state-of-the-art.

According to the results presented in Table 4, our proposed method, UDnet, demonstrates superior performance in underwater image enhancement when compared to other published works on four datasets: UIEBD, EUVP, UCCS and UIQS. Specifically, UDnet achieved the highest average values for PSNR, SSIM, UIQM and UCIQE metrics on these datasets. For instance, on the UIEBD dataset, UDnet achieved the highest UIQM and UCIQE values while on the EUVP dataset, it achieved the highest PSNR, UIQM and UCIQE values. Similarly, on the UCCS dataset, UDnet achieved the highest UIQM and UCIQE values and on the UIQS dataset, it achieved the highest UCIQE value. These results indicate that UDnet is a robust method for enhancing underwater images.

Our method, UDnet, performs better than other methods due to the effective combination of several components. The cVAE module is able to learn a compact and informative representation of the underwater image content, which helps to preserve important details during the enhancement process. The PAdaIN module is able to adaptively adjust the style of the enhanced image to match the target domain, resulting in more natural and visually pleasing results. Finally, the multi-colour

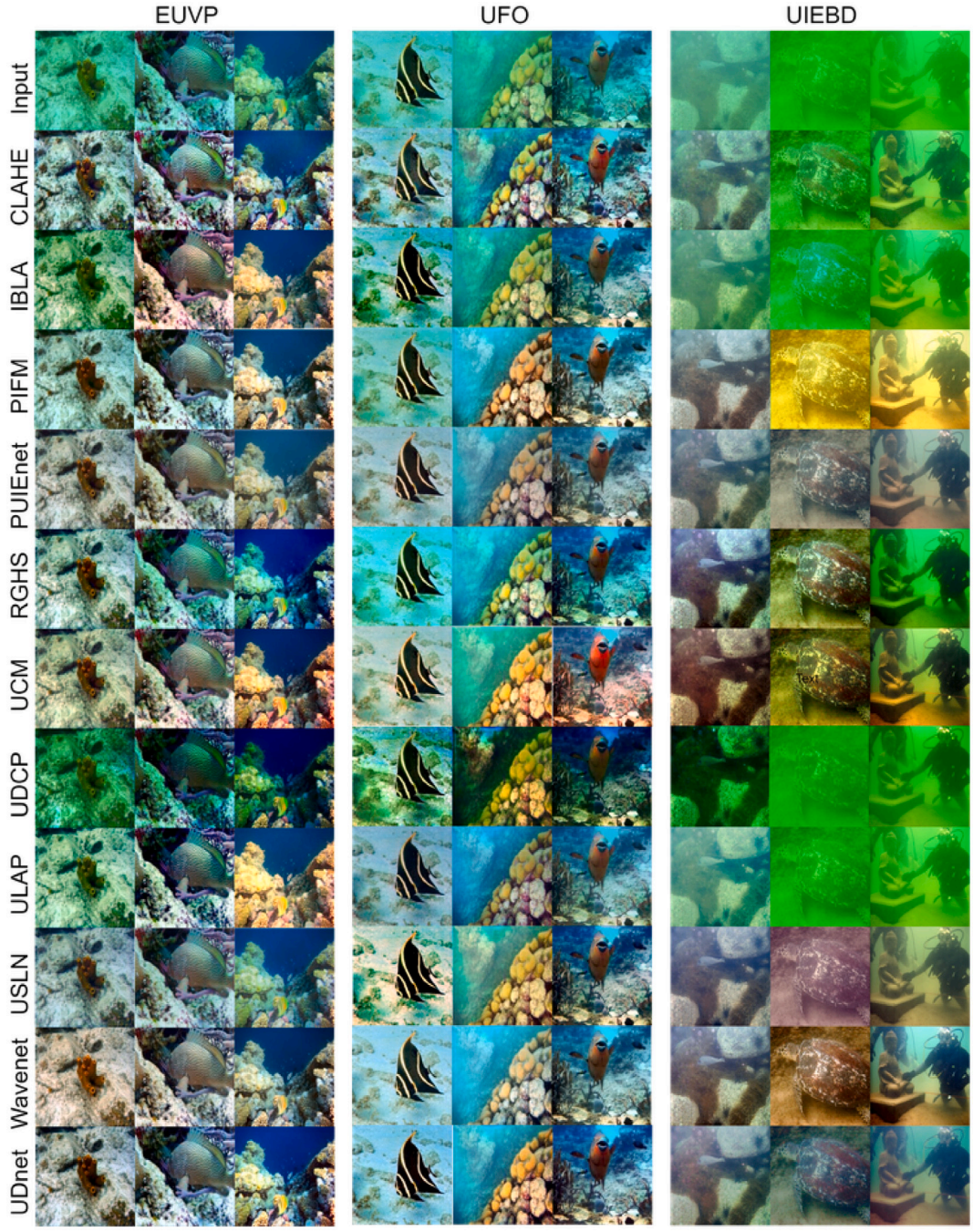


Fig. 3. Visual comparisons on challenging underwater images sampled from paired datasets, i.e. EUVP (Islam, Xia, & Sattar, 2020), UFO (Jahidul Islam et al., 2020), and UIEBD (Li et al., 2020). The name on the right of each row refers to the enhancement method used.

space stretch module is able to effectively enhance the contrast and colour of the underwater images by stretching the colour histogram in multiple colour spaces. These components work together to produce high-quality enhanced underwater images.

4.6. Qualitative comparisons

Underwater images possess several unique characteristics. They have more texture content and low luminance and contrast compared to terrestrial images. Therefore, it is important to assess human visual perception in terms of image content enhancement in underwater images, especially in terms of colour enhancement. To gain more insight into the effectiveness of our proposed UDnet, we performed comprehensive investigations and comparisons among all eight data sets using the ten previous methods introduced.

Fig. 3 demonstrates three example raw input images of each of the three paired datasets in the first row, along with the enhanced image outputs from the 10 aforementioned studies and our UDNet. This comparison has a two-fold purpose: (1) To demonstrate the effectiveness of the deep-learning-based methods in the no-reference settings. (2) To showcase the superiority of our unsupervised method, which has enhanced the underwater scenes without ground truth for training.

Furthermore, to prove the superiority of our model in handling unpaired images, we show visual comparisons of randomly selected underwater images from the five aforementioned unpaired datasets in Figs. 4, and 5. We also include a short video of our model's prediction at https://youtu.be/k4ASsGze5p8 and https://youtu.be/NV5GH-GG_3c.

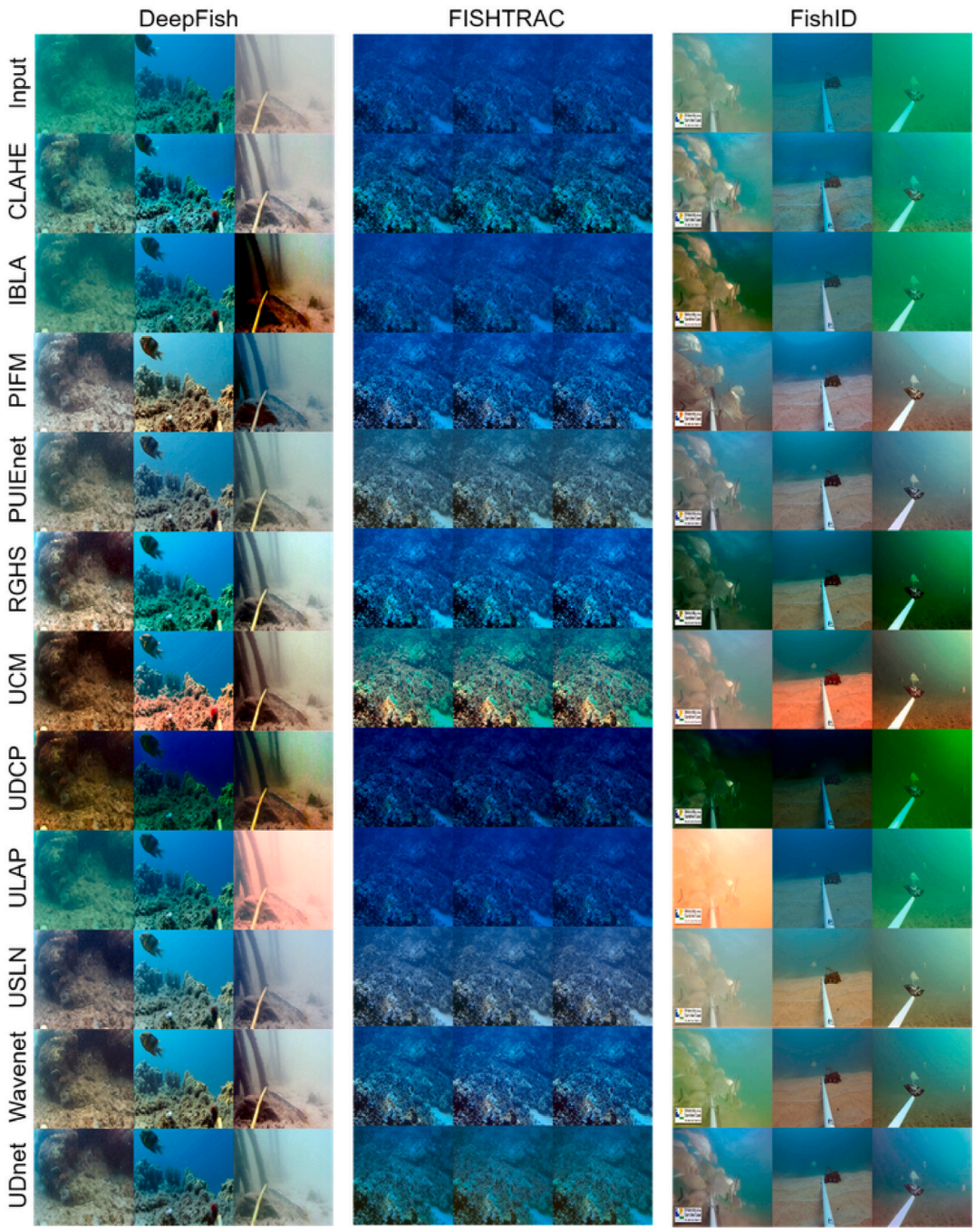


Fig. 4. Visual comparisons on challenging underwater images sampled from DeepFish (Saleh et al., 2020), FISHTRAC (Mandel et al., 2023), and FishID (Lopez-Marcano et al., 2021). The name on the right of each row refers to the method. We also include a short video of our model's prediction at https://youtu.be/k4ASsGze5p8 and https://youtu.be/NV5GH-GG_3c.

As Fig. 4 shows, the obvious light limitation of the raw image results in low contrast. For example, UDCP and ULAP models tend to make the image darker, while others such as UCMeven introduce reddish colour. In comparison, our model increases both brightness and contrast, making the details of the image clear. The input image samples given in Fig. 5 mostly suffer from obvious green deviation, which cannot be resolved by most models. For example, CLAHE, IBLA, and ULAP fail to remove the green deviation. In comparison, our model removes the greenish colour and makes the image colours balanced.

Overall, our qualitative comparison results show:

- Even when the ground truth label of the paired images is added to enhance visual quality, some of the previous methods show problems such as over-enhancement, lack of contrast, and saturation.
- Some of the models' output images from the paired dataset have over- or under-enhanced backgrounds, while some have no change in the background. However, the output image of our model does not show such problems.
- Some of the models' output images' background pixels are saturated. However, our model has not suffered from the over- or under-saturation problem.

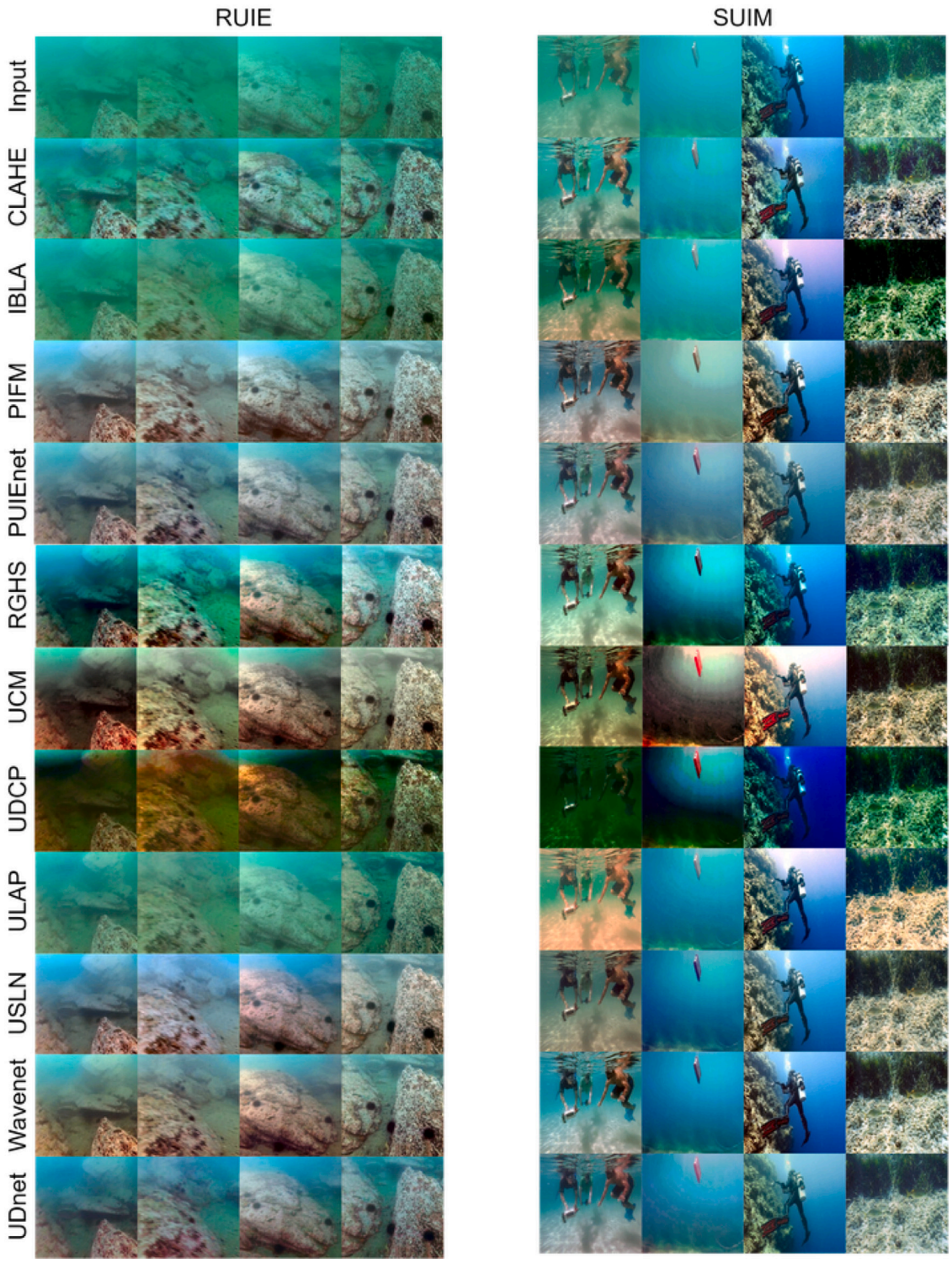


Fig. 5. Visual comparisons on challenging underwater images sampled from RUIE (Liu et al., 2020), and SUIM (Islam, Edge, et al., 2020). The name on the right of each row refers to the method.

4.7. Visual perception improvement

One of the main objectives of underwater image enhancement is to increase underwater robots' capacity to visually perceive their surroundings. This is essential for robots to make autonomous decisions in complex underwater scenarios. To evaluate our model's performance in visual perception improvement, we used feature detection and matching to assess its capability in improving the visual perception of underwater images. Feature detection and matching are commonly used techniques in many computer vision applications, such as structure-from-motion, image retrieval, object detection, and image stitching. Here, we use Scale-Invariant Feature Transform (SIFT), which helps locate the local features in an image (keypoints), and Random Sample Consensus (RANSAC) (Li et al., 2018), which is used to match feature points. These methods are used to compare the visual perception of an underwater image before and after enhancement. Fig. 6

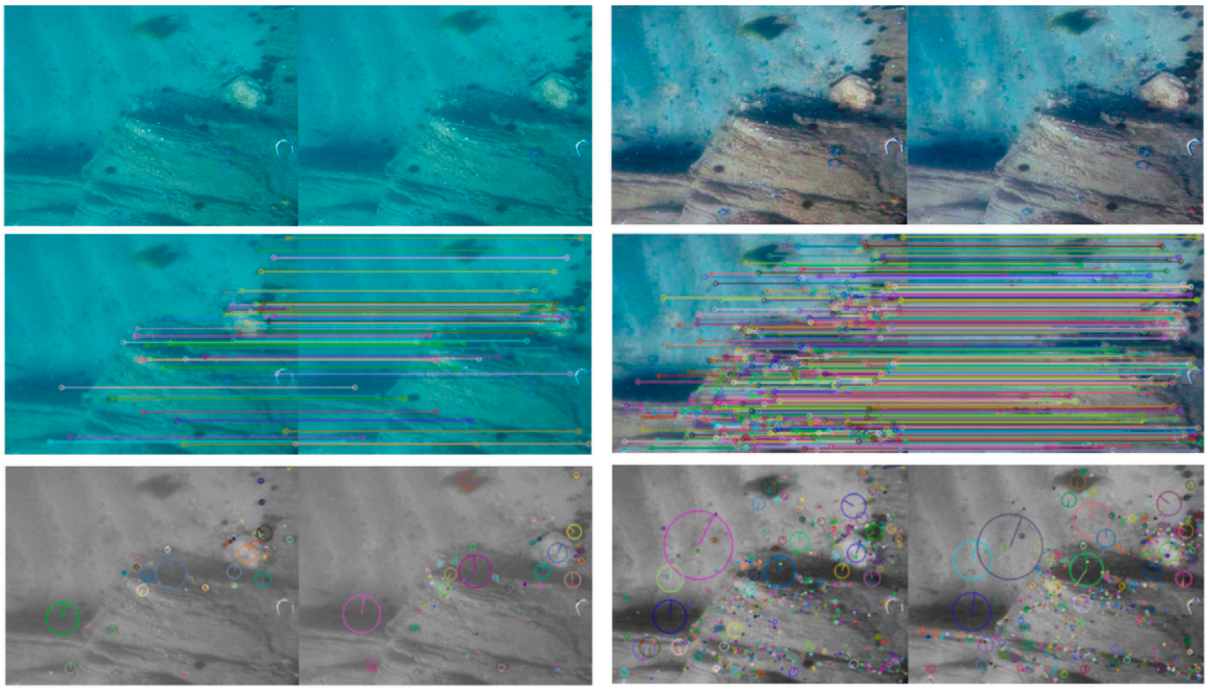


Fig. 6. Comparison of image feature and key points matching before (left) and after (right) image enhancement with our model. From the top: the original images, matched feature points, and SIFT keypoints. The images are from RUIE (Liu et al., 2020) dataset.

depicts the result for two consecutive frames from RUIE (Liu et al., 2020) dataset, (blue_01.jpg) and (blue_02.jpg). These show that the numbers of matched points between the two image frames increase from 74 (before the enhancement) to 594 after the enhancement. At the same time, the number of SIFT keypoints also dramatically increases as a result of the enhancement, significantly improving the visual perception of the environment.

4.8. Ablation study

To better understand how the proposed method works and what are the key factors that contribute to its performance, we conducted an ablation study to examine the impact of its different components and stages. These include the SGMCSS that adjusts the colour balance of the input images, the extra reference maps that are generated by applying different enhancement techniques to the input images, and the VGG loss that measures the perceptual similarity between the output images and the reference maps. We compared the full model with several ablated variants that remove or modify one of these components or stages. The quantitative comparisons are presented in Table 5, where

- **w/o colour** means that UDnet is trained without using the SGMCSS in the reference map generation stage. The input images are directly fed to the cVAE without any colour adjustment.
- **All colour** means that UDnet is trained with applying the SGMCSS to all inputs, including the input images and the reference maps. This means that the colour balance of both the input images and the reference maps are adjusted before feeding them to the cVAE.
- **Multi-label** means that UDnet is trained with using 6 extra enhanced reference maps that are generated by applying different enhancement techniques to the input images, such as histogram equalization, CLAHE, and Retinex. This means that each input image has 9 reference maps in total, including the original 3 generated by contrast and saturation adjustment, and gamma correction.
- **w/o VGG** means that UDnet is trained without using VGG loss in the objective function. The model only minimizes the reconstruction loss between the output images and the reference maps.

Table 5
Ablation study: Comparison against different model variants on UIEBD dataset in terms of average PSNR and SSIM values.

Model variant	PSNR ↑	SSIM ↑
w/o colour	22.01	0.791
All colour	21.73	0.784
Multi-label	22.12	0.795
w/o VGG	21.89	0.789
Full Model	22.23	0.812

We used PSNR and SSIM to evaluate the results on UIEBD, which are shown in Table 5. The results show that the entire model outperforms all variants ablated in both metrics, indicating that each component and stage of the proposed method is essential for achieving high-quality underwater image enhancement. The qualitative comparisons of the output images produced by different variants are presented in Fig. 7. From these comparisons, we can draw the following conclusions:

(1) Without using the SGMCSS in the reference map generation stage, UDnet fails to produce satisfactory results (see Fig. 7). The output images still suffer from low contrast and poor visibility. For example, in the first row of Fig. 7, it is hard to see the details of the sea bed and the green plants at the bottom of the image. This suggests that adjusting the colour balance of the input images is a crucial step for generating realistic reference maps that can guide UDnet to enhance underwater images.
(2) With applying the SGMCSS to all inputs, UDnet performs slightly better than without using it at all, but still worse than using it only in the reference map generation stage (see Fig. 7(d)). The output images are brighter than w/o colour, but they also lose some details and colours. For example, in the second row of Fig. 7, some parts of the fish are over-exposed and washed

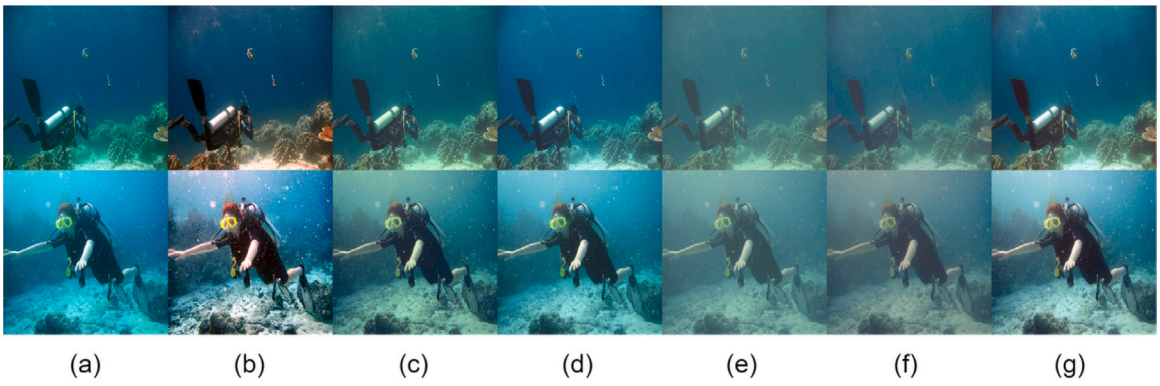


Fig. 7. ABLATION STUDY: The qualitative comparison of the contributions of multiple stages of the proposed framework on the UIEBD dataset. (a) Input, (b) ground truth, (c) w/o colour, (d) All colour, (e) Multi-label, (f) w/o VGG, (g) Full Model.

Table 6
Quantitative comparisons for the PAdaIN module and baseline model against the full UDnet model on the UIEBD dataset. Metrics include PSNR (dB) and SSIM. Higher values indicate better performance.

Model Variant	PSNR (dB)	SSIM
Baseline Model	19.42	0.712
w/o PAdaIN	21.35	0.743
Full Model (UDnet)	**22.23**	**0.812**

out. This implies that adjusting the colour balance of both the input images and the reference maps may introduce some inconsistency and distortion that can affect UDnet's ability to learn from them.

(3) Adding more reference maps by using different enhancement techniques does not improve UDnet's performance, but rather degrades it (see Fig. 7(e)). The output images are over-enhanced and have unnatural colours. For example, in the third row of Fig. 7, some parts of the coral are too bright and have a pinkish hue. This indicates that adding more reference maps does not necessarily provide more useful information for UDnet, but may introduce more noise and ambiguity that can confuse UDnet and make it harder to learn from them.

(4) When UDnet is trained without using VGG loss, the quality of the output images is significantly reduced (see Fig. 7(f)). The output images have low contrast, poor visibility, and distorted colours. For example, in the fourth row of Fig. 7, the output image is very dark and has a bluish tint. This demonstrates that VGG loss is an important component of the objective function that can help UDnet to learn more perceptual features and semantic information from the reference maps and improve the visual quality of the output images.

Effectiveness of the PAdaIN Module and Baseline Model Comparison:

To further validate the significance of the PAdaIN module, we conducted additional experiments comparing its contribution to the performance of the proposed framework. Table 6 presents the quantitative results of these experiments on the UIEBD dataset, using PSNR and SSIM as evaluation metrics. The results confirm that each component of the proposed method plays a crucial role in achieving superior performance.

Baseline Model: This variant, which excludes all proposed enhancements, achieves the lowest PSNR and SSIM values, highlighting the need for the advanced components integrated into the full UDnet model. w/o PAdaIN: Removing the PAdaIN module causes a noticeable drop in performance compared to the full model, demonstrating that PAdaIN effectively encodes feature uncertainties and contributes to high-quality underwater image enhancement. Full Model (UDnet): The complete model, including SGMCSS, cVAE, and PAdaIN modules, achieves the best results with a PSNR of 22.23 dB and SSIM of 0.812, highlighting the synergy of these components in the overall framework. These findings underscore the importance of the PAdaIN module, particularly in refining the network's ability to enhance underwater images.

5. Discussion

Enhancing underwater images is challenging due to the complex and diverse nature of underwater environments. Our proposed method, UDNet, addresses these challenges by adopting an unsupervised framework that leverages probabilistic uncertainty modelling during training. This novel approach enables UDNet to adaptively enhance underwater images with varying characteristics, setting it apart from traditional supervised methods that rely on large datasets of paired raw and enhanced images.

UDNet's key strengths include its unsupervised learning capability, which eliminates the need for ground truth data, and its innovative use of statistical information through the Statistically Guided Multi-Colour Space Stretch (SGMCSS) and Probabilistic Adaptive Instance Normalization (PAdaIN) modules. These modules improve robustness and enhance image quality by addressing variations in contrast, colour balance, and illumination. Experimental results confirm UDNet's competitive performance across eight public datasets, demonstrating its ability to outperform or match state-of-the-art methods quantitatively and qualitatively.

Despite these advancements, UDNet has limitations that warrant further exploration. Backscatter, particularly at greater distances, remains a significant challenge, as it affects the visual clarity of enhanced images. While our approach mitigates some of these issues, its reliance on statistical models can occasionally result in unrealistic enhancements. Additionally, the environmental variability of underwater settings — ranging from oceans to lakes — means that the model's generalization may not always produce optimal results across all scenarios.

To address these limitations, future work will focus on enhancing UDNet's robustness by exploring alternative CNN architectures and integrating multi-resolution approaches to capture finer details. Improved methods for reference map generation could also further reduce dependence on statistical assumptions, resulting in higher-quality enhancements.

The potential applications of UDNet are vast. Enhanced underwater images can significantly benefit environmental monitoring, providing insights into marine ecosystems and supporting conservation efforts. In marine biology, UDNet's improvements can aid in the study of species behaviour and habitats. Moreover, in underwater archaeology, the model's ability to clarify images can facilitate the study of submerged

artifacts and structures.

Beyond its current scope, UDNet has the potential to serve as a framework for generating high-quality reference maps for other domains, such as medical imaging or satellite image enhancement, where ground truth data is challenging to obtain.

6. Conclusion

In this work, we introduced UDNet, an unsupervised deep learning framework designed for underwater image enhancement. By leveraging probabilistic uncertainty modelling and an encoder–decoder architecture, UDNet effectively addresses challenges such as random distortion and low contrast inherent in underwater images. Its innovative design — featuring the SGMCSS and PAdaIN modules — enables robust image enhancement without relying on manually labelled data, marking a significant advancement in the field. Our experimental results demonstrate that UDNet outperforms ten state-of-the-art underwater image enhancement methods across seven metrics and eight datasets, underscoring its versatility and effectiveness. UDNet's strong generalization ability, particularly with unpaired datasets, positions it as a practical tool for diverse underwater applications.

CRediT authorship contribution statement

Alzayat Saleh: Conceptualisation, Data curation, Data analysis, Software development, DL algorithm design, Visualization, Writing – original draft. **Marcus Sheaves:** Writing – review and editing, Supervision. **Dean Jerry:** Writing – review and editing, Supervision. **Mostafa Rahimi Azghadi:** Conceptualisation, Ph.D. supervision, Reviewing /editing the draft, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research is supported by the Australian Research Training Program (RTP) Scholarship and Food Agility HDR Top-Up Scholarship. D. Jerry and M. Rahimi Azghadi acknowledge the Australian Research Council through their Industrial Transformation Research Hub program.

Data availability

Data will be made available on request.

References

Balakrishnan, G., Dalca, A. V., Zhao, A., Guttag, J. V., Durand, F., & Freeman, W. T. (2019). Visual deprojection: Probabilistic recovery of collapsed dimensions. In *Proceedings of the IEEE international conference on computer vision* (pp. 171–180).

Braik, M. (2024). Hybrid enhanced whale optimization algorithm for contrast and detail enhancement of color images. *Cluster Computing*, *27*(1), 231–267.

Chandler, D. M. (2010). Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of Electronic Imaging*, *19*(1), Article 011006.

Chen, X., Zhang, P., Quan, L., Yi, C., & Lu, C. (2021). Underwater image enhancement based on deep learning and image formation model. *Computers & Electrical Engineering*.

Cheng, Z., Fan, G., Zhou, J., Gan, M., & Chen, C. P. (2024). FDCE-net: underwater image enhancement with embedding frequency and dual color encoder. *IEEE Transactions on Circuits and Systems for Video Technology*.

Drews, P., Nascimento, E., Moraes, F., Botelho, S., & Campos, M. (2013). Transmission estimation in underwater single images. In *Proceedings of the IEEE international conference on computer vision workshops* (pp. 825–830).

Fu, Z., Lin, H., Yang, Y., Chai, S., Sun, L., Huang, Y., & Ding, X. (2022). Unsupervised underwater image restoration: From a homology perspective. In *Proceedings of the AAAI conference on artificial intelligence, vol. 36, no. 1* (pp. 643–651). Association for the Advancement of Artificial Intelligence.

Fu, Z., Wang, W., Huang, Y., Ding, X., & Ma, K.-K. (2022). Uncertainty inspired underwater image enhancement. In *Computer vision–ECCV 2022: 17th European conference, Tel Aviv, Israel, October 23–27, 2022, proceedings, part XVIII* (pp. 465–482). Springer.

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 multimedia modeling* (pp. 453–465).

Iqbal, K., Odetayo, M., James, A., Salam, R. A., & Talib, A. Z. H. (2010). Enhancing the low quality images using unsupervised colour correction method. In *2010 IEEE international conference on systems, man and cybernetics* (pp. 1703–1709).

Islam, M. J., Edge, C., Xiao, Y., Luo, P., Mehtaz, M., Morse, C., Enan, S. S., & Sattar, J. (2020). Semantic segmentation of underwater imagery: Dataset and benchmark. In *2020 IEEE/RSJ international conference on intelligent robots and systems* (pp. 1769–1776). IEEE.

Islam, M. J., Xia, Y., & Sattar, J. (2020). Fast underwater image enhancement for improved visual perception. *IEEE Robotics and Automation Letters*, *5*(2), 3227–3234.

Jahidul Islam, M., Luo, P., & Sattar, J. (2020). Simultaneous enhancement and super-resolution of underwater imagery for improved visual perception. In *Robotics: Dcience and systems XVI*. Corvalis, Oregon, USA: Robotics: Science and Systems Foundation.

Jebadass, J. R., & Balasubramaniam, P. (2024). Color image enhancement technique based on interval-valued intuitionistic fuzzy set. *Information Sciences*, *653*, Article 119811.

Ji, X., Wang, X., Leng, N., Hao, L.-Y., & Guo, H. (2024). Dual-branch underwater image enhancement network via multiscale neighborhood interaction attention learning. *Image and Vision Computing*, *151*, Article 105256.

Jiang, Q., Zhang, Y., Bao, F., Zhao, X., Zhang, C., & Liu, P. (2022). Two-step domain adaptation for underwater image enhancement. *Pattern Recognition*, *122*, Article 108324.

Johnson, J., Alahi, A., & Fei-Fei, L. (2016). Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision* (pp. 694–711).

Ke, J., Wang, Q., Wang, Y., Milanfar, P., & Yang, F. (2021). Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 5148–5157).

Kingma, D. P., & Welling, M. (2019). An introduction to variational autoencoders. *Foundations and Trends® in Machine Learning*, *12*(4), 307–392.

Lei, X., Wang, H., Shen, J., Chen, Z., & Zhang, W. (2024). A novel intelligent underwater image enhancement method via color correction and contrast stretching. *Microprocessors and Microsystems*, *107*, Article 104040.

Li, C., Guo, C., Ren, W., Cong, R., Hou, J., Kwong, S., & Tao, D. (2020). An underwater image enhancement benchmark dataset and beyond. *IEEE Transactions on Image Processing*, *29*, 4376–4389.

Li, H., Qin, J., Xiang, X., Pan, L., Ma, W., & Xiong, N. N. (2018). An efficient image matching algorithm based on adaptive threshold and RANSAC. *IEEE Access*, *6*, 66963–66971.

Li, B., Sun, Z., & Guo, Y. (2019). Supervae: Superpixelwise variational autoencoder for salient object detection. In *Proceedings of the AAAI conference on artificial intelligence, vol. 33* (pp. 8569–8576).

Liu, Y., Dong, Z., Zhu, P., & Liu, S. (2022). Unsupervised underwater image enhancement based on feature disentanglement. *Dianzi Yu Xinxi Xuebao*, (Journal of Electronics and Information Technology) *44*(10), 3389–3398.

Liu, R., Fan, X., Zhu, M., Hou, M., & Luo, Z. (2020). Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light. *IEEE Transactions on Circuits and Systems for Video Technology*, *30*(12), 4861–4875.

Liu, R., Jiang, Z., Yang, S., & Fan, X. (2022). Twin adversarial contrastive learning for underwater image enhancement and beyond. *IEEE Transactions on Image Processing*, *31*, 4922–4936.

Liu, T., Zhu, K., Cao, W., Shan, B., & Guo, F. (2024). Zero-UMSIE: a zero-shot underwater multi-scale image enhancement method based on isomorphic features. *Optics Express*, *32*(23), 40398–40415.

Lopez-Marcano, S., Jinks, E., Buelow, C. A., Brown, C. J., Wang, D., Kusy, B., Ditria, E., & Connolly, R. M. (2021). Automatic detection of fish and tracking of movement for ecology. *Ecology and Evolution*, *11*(12), 8254–8263.

Mandel, T., Jimenez, M., Risley, E., Nammoto, T., Williams, R., Panoff, M., Ballesteros, M., & Suarez, B. (2023). Detection confidence driven multi-object tracking to recover reliable tracks from unreliable detections. *Pattern Recognition*, *135*, Article 109107.

Mittal, A., Soundararajan, R., & Bovik, A. C. (2013). Making a "completely blind" image quality analyzer. *IEEE Signal Processing Letters*, *20*(3), 209–212.

Panetta, K., Gao, C., & Agaian, S. (2016). Human-visual-system-inspired underwater image quality measures. *IEEE Journal of Oceanic Engineering*, *41*(3), 541–551.

Peng, Y.-T., & Cosman, P. C. (2017). Underwater image restoration based on image blurriness and light absorption. *IEEE Transactions on Image Processing*, *26*(4), 1579–1594.

Raveendran, S., Patil, M. D., & Birajdar, G. K. (2024). Underwater image quality enhancement using fusion of adaptive colour correction and improved contrast enhancement strategy. *International Journal of Image and Data Fusion*, 1–29.

Saleh, A., Laradji, I. H., Konovalov, D. A., Bradley, M., Vazquez, D., & Sheaves, M. (2020). A realistic fish-habitat dataset to evaluate algorithms for underwater visual analysis. *Scientific Reports*, *10*(1), 14671.

Sharma, P., Bisht, I., & Sur, A. (2023). Wavelength-based attributed deep neural network for underwater image restoration. *ACM Transactions on Multimedia Computing, Communications and Applications*, *19*(1), 1–23.

Sohn, K., Lee, H., & Yan, X. (2015). Learning structured output representation using deep conditional generative models. *Advances in Neural Information Processing Systems (NeurIPS)*, *28*, 3483–3491.

Song, W., Wang, Y., Huang, D., & Tjondronegoro, D. (2018). A rapid scene depth estimation model based on underwater light attenuation prior for underwater image restoration. In *Pacific rim conference on multimedia* (pp. 678–688).

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, *13*(4), 600–612.

Wang, Z., Li, C., Mo, Y., & Shang, S. (2023). RCA-CycleGAN: Unsupervised underwater image enhancement using red channel attention optimized CycleGAN. *Displays*, *76*, Article 102359.

Wang, W., & Yang, Y. (2024). A histogram equalization model for color image contrast enhancement. *Signal, Image and Video Processing*, *18*(2), 1725–1732.

Xiang, D., Zhou, Z., Yang, W., Wang, H., Gao, P., Xiao, M., Zhang, J., & Zhu, X. (2025). A fusion framework with multi-scale convolution and triple-branch cascaded transformer for underwater image enhancement. *Optics and Lasers in Engineering*, *184*, Article 108640.

Xiao, Z., Han, Y., Rahardja, S., & Ma, Y. (2022). USLN: A statistically guided lightweight network for underwater image enhancement via dual-statistic white balance and multi-color space stretch. *IEEE Transactions on Neural Networks Learning System*.

Xue, W., Zhang, L., Mou, X., & Bovik, A. C. (2014). Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE Transactions on Image Processing*, *23*(2), 684–695.

Yan, X., Rastogi, A., Villegas, R., Sunkavalli, K., Shechtman, E., Hadap, S., Yumer, E., & Lee, H. (2018). Mt-vae: Learning motion transformations to generate multimodal human dynamics. In *Proceedings of the European conference on computer vision* (pp. 265–281).

Yu, Y., & Qin, C. (2023). An end-to-end underwater-image-enhancement framework based on fractional integral retinex and unsupervised autoencoder. *Fractal and Fractional*, *7*(1), 70.

Zhang, W., Zhuang, P., Sun, H. H., Li, G., Kwong, S., & Li, C. (2022). Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement. *IEEE Transactions on Image Processing*, *31*, 3997–4010.

Zheng, S., Wang, R., Zheng, S., Wang, L., & Liu, Z. (2024). A learnable full-frequency transformer dual generative adversarial network for underwater image enhancement. *Frontiers in Marine Science*, *11*, Article 1321549.

Zuiderveld, K. (1994). Contrast limited adaptive histogram equalization. In *Graphic gems IV* (pp. 474–485). San Diego: Academic Press Professional.