Underwater Image Enhancement Using Multiple Deep Learning Methods

Seif Kassab
Nile University
6 October, Cairo, Egypt
s.Usama2251@nu.edu.eg

NourEldin Ayman
Nile University
6 October, Cairo, Egypt
N.Ayman2201@nu.edu.eg

Karim Raouf
Nile University
6 October, Cairo, Egypt
karim@nu.edu.eg

Omar Radwan
Nile University
6 October, Cairo, Egypt
omar@nu.edu.eg

Danial Maged
Nile University
6 October, Cairo, Egypt
daniel@nu.edu.eg

Abstract

Underwater images are crucial for marine exploration and research, but they often suffer from challenges such as light scattering, color distortion, and blurring. This study evaluates the effectiveness of five state-of-the-art deep learning models—U-Net, EnhanceNet, ResNet, FC-DenseNet, and UICE^2-Net—in addressing these issues and enhancing underwater image quality. Using three datasets—the Large-Scale Underwater Image (LSUI) dataset, the EUVP dataset, and an ImageNet-derived dataset—comprehensive experiments were conducted to assess model performance and overcome the limitations of existing methods. The findings reveal that these models improve image clarity and color fidelity to varying extents, offering unique strengths suited for different underwater scenarios. This work contributes valuable insights into optimizing deep learning approaches for real-world underwater applications.

1. Introduction

High-quality underwater images are essential for exploration and research of marine ecosystems [8], [4], [10]. However, the complex underwater environment leads to issues such as light scattering, color distortion, structural blurring, and noise interference, which ultimately obstructs images and hinder their quality [10]. Underwater images often appear blurry and lack clear contrast because light fades quickly as it moves through the water, making everything look hazy and distorted.

These challenges make it difficult to interpret underwater scenes and limit the effectiveness of tasks such as segmentation, and vision navigation systems for autonomous underwater vehicles [6]. Therefore, Underwater Image Enhancement (UIE) is necessary to combat these issues and improve the quality of the images to better understand the mysterious world of underwater.

Recently, UIE has gained a lot of attention, and significant progress has been made insert table summarizing the work, year, scope, limitation, mainly due to the progress in deep learning technologies. Despite these advancements, existing data-driven methods often face the limitation of small datasets., few underwater scenes, or even real-world images, which limits their performance. [7]. In this paper we aim to evaluate renowned deep-learning models such as the U-Net, EnhanceNet, ResNet, FCDenseNet, UICE^2-Net

In this paper we utilize three datasets:

1. Large scale underwater image (LSUI) dataset [7], which contains 4279 image pairs that covers different water types, lighting, and overall scenes. And the image targets are generated using voting or something
2. The EUVP (Enhancing Underwater Visual Perception) dataset [3] that ... we only used the Image net and scenes
3. The Image net ... which encompasses more images from imagenet

2. Related Work

UIE has been an active field of research due to its crucial role in marine exploration, deep-sea monitoring, and its use in autonomous underwater robots. Prior works can be categorized into traditional and data-driven approaches.

First, UIE methods based on traditional methods primarily relied on image processing. These methods adjust pixel values rather than establishing a mathematical model to simulate the image optical imaging characteristics [5]. Among these, Histogram Equalization [2] [13] is commonly used for contrast enhancement. Through nonlinear pixel value redistribution, the algorithm transforms an arbitrary histogram distribution into a more uniform one. White

Ancuti et al. [1] proposed a fusion-based method that combines multiple image derivatives to enhance contrast and color. Similarly, Galdran et al. [9] introduced Red Channel Prior, addressing the wavelength-dependent attenuation characteristic of underwater environments. (2) Another fundamental approach is White Balance [7], [3], which addresses color distortions caused by varying underwater lighting conditions. This technique works by adjusting the ratios between RGB channels to compensate for color shifts in different underwater environments. Additionally, researchers have developed enhancement methods that leverage human visual perception principles. Notable among these are Retinex-based algorithms, which have made significant contributions to the field by modeling how the human visual system processes brightness and color. These perception-inspired approaches continue to influence modern underwater image enhancement techniques.

These methods, while computationally efficient, often struggle with varying water types and illumination conditions and doesn't perform well in a real-world setting with different water types and enviroments.

With the advent of deep learning, these methods have evolved significantly, transforming various fields from natural language processing to computer vision. These methods can also be used to enhance underwater images.

UnSupervised Underwater Image Restoration method (USUIR), which addresses the scarcity of real-world paired training data. USUIR leverages the homology property between raw underwater images and re-degraded images, decomposing the enhancement process into three key components: global background light estimation, transmission map calculation, and scene radiance recovery. This approach demonstrates how deep learning can overcome traditional limitations in underwater image enhancement while maintaining computational efficiency.

PUIE-Net takes a different approach by treating underwater image enhancement as a distribution estimation problem. The method combines a conditional variational autoencoder with adaptive instance normalization to model possible enhancement outcomes, followed by a consensus process to generate the final enhanced image. This approach effectively handles the common challenge of reference map ambiguity in underwater image enhancement.

Fabbri et al. [11] proposed a method using Generative Adversarial Networks (GANs) to rectify underwater image distortions and enhance input quality for vision-driven tasks in autonomous underwater vehicles. Additionally, a weakly supervised underwater color transmission model based on CycleGAN [14, 13] was introduced, which capitalizes on the adversarial network structure and various loss functions. This model can be trained with unpaired underwater images,ents. Nonetheless, the training da improving the network's adaptability to different underwater environmtasets employed in these methods do not consist of perfectly matched real-world underwater images, resulting in limited enhancement performance across diverse underwater scenes.

Currently most of the deep learning methods used for UIE are weakly supervised learning methods based on generative adversial networks, i.e. [12] and since they're generative, they sometimes produce unstable results.

3. Dataset and Methods

As mentioned in the introduction, this section outlines the datasets utilized and the architectures of our fully connected convolutional networks.

In this paper we leverage three datasets:

1. Large scale underwater image (LSUI) dataset [7], which contains 4279 image pairs that covers different water types, lighting, and overall scenes. And the reference images were selected with two rounds of subjective and objective evaluations.

2. The EUVP (Enhancing Underwater Visual Perception) dataset . The dataset contains three paired datasets: Underwater Dark (5550 pairs), Underwater ImageNet (3700 pairs), Underwater Scenes (2185 pairs). We only used the Underwater ImageNet and Underwater Scenes.The target images are generated using CycleGAN.

3. The third and final dataset consists of 6,128 image pairs sampled from ImageNet https://github.com/cameronfabbri/Underwater-Color-Correction. The target images are also generated using CycleGan.

By combining these datasets we obtain a total of 16,292 image pairs. These include evaluated images from LSUI and the semi-realistic generated data from CycleGAN.

3.1. U-Net

U-Net is a convolutional neural network architecture originally developed for biomedical image segmentation. It is characterized by a symmetric encoder-decoder structure with skip connections, enabling the model to capture both local and global context. This makes U-Net well-suited for pixel-level prediction tasks like underwater image reconstruction.

3.1.1 Architecture

The architecture of U-Net consists of three main components: an encoder, a bottleneck, and a decoder. Each com-

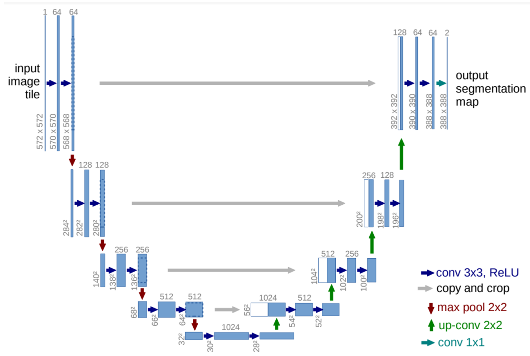

Figure 1. Your descriptive caption here.

ponent contributes to extracting, compressing, and reconstructing features.

Encoder The encoder captures feature representations at multiple scales through a series of convolutional and max-pooling layers. Spatial dimensions are reduced while feature depth increases, enabling the extraction of hierarchical features.

- **Stage 1:** Two 3×3 convolutional layers with 64 filters, followed by 2×2 max pooling.
- **Stage 2:** Two 3×3 convolutional layers with 128 filters, followed by 2×2 max pooling.
- **Stage 3:** Two 3×3 convolutional layers with 256 filters, followed by 2×2 max pooling.

Bottleneck The bottleneck consists of a single 3×3 convolutional layer with 512 filters, capturing high-level, abstract features of the input image.

Decoder The decoder restores spatial resolution using upsampling layers and concatenates features from the encoder via skip connections to improve reconstruction accuracy.

- **Stage 1:** Upsampling by a factor of 2, concatenation with encoder stage 3 output, followed by two 3×3 convolutional layers with 256 filters.
- **Stage 2:** Upsampling by a factor of 2, concatenation with encoder stage 2 output, followed by two 3×3 convolutional layers with 128 filters.
- **Stage 3:** Upsampling by a factor of 2, concatenation with encoder stage 1 output, followed by two 3×3 convolutional layers with 64 filters.

The final layer is a 1×1 convolution with three filters and a sigmoid activation function to output normalized pixel values.

3.2. Full Convolutional DenseNet

Fully Convolutional DenseNet FC-DenseNet, a modified/expanded version of DenseNet which is originally for classification, is used here for image reconstruction. DenseNet connects each layer to every other layer in a feed forward, are built from dense blocks and pooling operations, where inside each dense block is an iterative concatenation of previously outputted feature maps. This architecture is similar to ResNets, which performs summation of previous feature maps instead of concatenation. This modification allows the DenseNet to be parameter efficient and the ability to use previously computed feature maps, alleviating the vanishing-gradient problem. Making it somewhat suitable for real-time ROVs. In each dense block, the output dimension of each layer l has k feature maps, where k here is referred to as the growth rate. This causes the DenseNet to grow linearly with its depth, therefore a pooling operation is added after every dense block referred to as transition down, to reduce the spatial dimension of the feature maps.

The final layer of the downsampling path known as he bottleneck. captures the high-level features of the image. To restore the input spatial resolution, an upsampling path is added, which replaces the convolution operation with a dense block and an upsampling process called transition up. Transition up modules involve a transposed convolution that up samples the preceding feature maps. These upsampled feature maps are then concatenated with those from the skip connection from the corresponding resolutions that are in the downsampling path to serve as the input for a new dense block. Note that The upsampling path increases the spatial resolution of feature maps, leading to linear growth in the number of features, which can become memory-intensive. To avoid this, only the feature maps from the last dense block are used for transposed convolution, avoiding feature map explosion.

The final layer in the network is a convolution followed by a sigmoid activation to provide the normalized pixel values

3.2.1 Architecture

The main architecture of FCDenseNet103, consists of a downsampling path, a bottleneck, and an upsampling path..

3.3. ResNet

This model employs a ResNet architecture designed to capture fine-grained details through residual learning. Each residual block consists of two convolutional layers, each followed by batch normalization and ReLU activation. A shortcut connection between the input and output of each block facilitates gradient flow and prevents vanishing gradients. The architecture includes an initial feature extraction

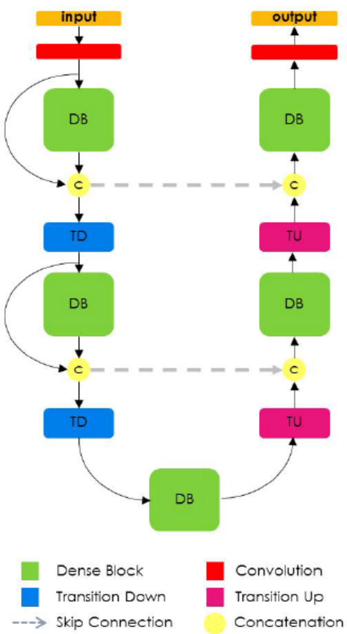

Figure 2. Diagram of the Fully Convolutional DenseNet The diagram is composed of a downsampling path with 2 Transitions Down (TD) and an upsampling path with 2 Transitions Up (TU). A circle represents concatenation and arrows represent connectivity patterns in the network. Gray horizontal arrows represent skip connections, the feature maps from the downsampling path are concatenated with the corresponding feature maps in the upsampling path. Note that the connectivity pattern in the upsampling and the downsampling paths are different. In the downsampling path, the input to a dense block is concatenated with its output, leading to a linear growth of the number of feature maps, whereas in the upsampling path, it is not

layer, followed by multiple residual blocks and an upsampling module to reconstruct high-resolution images. The output layer uses a sigmoid activation function to normalize pixel values. The model is trained with a custom configuration to balance computational efficiency and model performance.

3.4. EnhanceNet

EnhanceNet is a deep convolutional neural network designed for high-quality image enhancement and restoration. The model leverages skip connections, non-local attention, and multi-scale feature extraction to improve the reconstruction of degraded images. This architecture is particularly effective for applications such as underwater image reconstruction, where complex distortions like haze and color degradation must be corrected.

3.4.1 Architecture

The EnhanceNet architecture consists of five main components: input processing, feature extraction, skip connec-

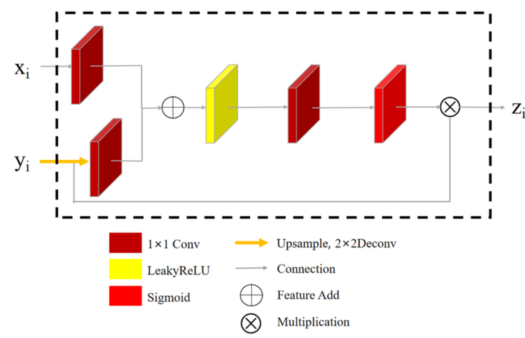

Figure 3. Diagram of the EnhanceNet architecture, showcasing its skip connections, non-local attention, and multi-scale feature extraction.

tions, non-local attention, and upsampling with multi-scale feature reconstruction.

Input Processing and Feature Extraction The input layer receives images of size $128 \times 128 \times 3$. The initial feature extraction involves two 3×3 convolutional layers with 128 filters and ReLU activation, followed by a 2×2 max pooling layer to reduce spatial dimensions. A dropout layer with a rate of 0.3 is applied to prevent overfitting.

Skip Connections The core of the architecture includes three residual blocks with skip connections. Each block performs the following operations:

- extbfConvolutional Layers: Two 3×3 convolutional layers with 128 filters and ReLU activation, followed by batch normalization.
- extbfSkip Connection: The output of the second convolutional layer is added to the input of the block to form a residual connection.
- extbfDropout: A dropout layer with a rate of 0.3 is applied after each residual block.

The skip connections enable the model to learn residual mappings, improving gradient flow and preventing vanishing gradients.

Non-Local Attention To capture global context and improve feature representation, a non-local attention mechanism is approximated using a 1×1 convolutional layer with 128 filters and ReLU activation. A dropout layer with a rate of 0.3 follows this operation.

Multi-Scale Feature Extraction EnhanceNet incorporates multi-scale feature extraction to address complex image distortions. Two dilated convolutional layers with dilation rates of 2 and 4, respectively, are used. Both layers have 128 filters and ReLU activation. Dropout is applied after these layers to enhance generalization.

Upsampling and Output Reconstruction The upsampling process restores the image to its original resolution using a transposed convolutional layer with 128 filters, a 3×3 kernel, and a stride of 2. Finally, a 3×3 convolutional layer with 3 filters and a sigmoid activation function produces the enhanced output image.

- extbfUpsampling: A transposed convolution layer increases the spatial resolution by a factor of 2.
- extbfOutput Layer: A 3×3 convolutional layer with 3 filters and sigmoid activation ensures the output is in the range [0, 1].

3.5. UICE^2-Net

The UICE^2-Net architecture is specifically designed to enhance underwater images by integrating complementary features extracted from the RGB and HSV color spaces. This dual-path architecture leverages the strengths of each color space to produce a more robust enhancement process.

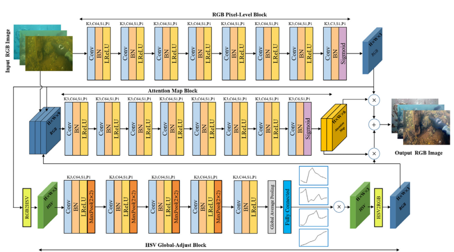

Figure 4. The UICE^2-Net Architecture

3.5.1 RGB Pathway

The RGB pathway processes the three-channel RGB input image to extract spatial and chromatic features. The processing begins with a series of convolutional layers:

Feature Extraction: The first layer applies a convolution operation with 256 filters to capture high-level features from the input image.

Down sampling: Subsequent layers reduce the spatial dimensions progressively, using stride-based convolutions, while increasing the number of filters to enhance feature representation. This step effectively condenses the image's information while retaining its essential details.

Up sampling and Reconstruction: After feature extraction, transposed convolution layers are used to up sample the feature maps back to the original resolution. This operation ensures that the enhanced RGB features contribute directly to the final image reconstruction.

3.5.2 HSV Pathway

The HSV pathway focuses primarily on the Value (V) channel, which represents image brightness, providing a complementary perspective to the RGB pathway:

Input Preprocessing: The input to this pathway is a single-channel HSV representation, derived from the original image by isolating the V channel. The first layer starts with 128 filters

Feature Extraction and Down sampling: Similar to the RGB pathway, convolutional layers extract brightness-related features and down sample the spatial dimensions to condense the representation.

Up sampling and Reconstruction: The pathway reconstructs the processed V channel using transposed convolutional layers. The output is upscaled to match the resolution of the RGB features.

3.5.3 Fusion of RGB and HSV Features

After processing in their respective pathways, the enhanced features from the RGB and HSV channels are combined to produce the final enhanced image:

Broadcasting and Addition: The single-channel HSV output is broadcasted to match the three-channel RGB format, ensuring compatibility for fusion.

Additive Fusion: The processed RGB features and the broadcasted HSV features are added elementwise. This fusion step allows the model to synergize chromatic and brightness information, generating a more visually appealing and balanced output

3.6. Evaluation Criterion and Loss Function

To evaluate the performance of our models, we utilized Mean Squared Error (MSE) as the loss function and some models used Structural Similarity Index Measure (SSIM) for model evaluation.

Loss Function: The Mean Squared Error (MSE) is defined as:

$$\mathcal{L}\text{MSE} = \sum_{i=1}^{M} (I_{ref} - I_{enh})^2 \tag{1}$$

where I_{ref} and I_{enh} are the reference and enhanced images, respectively, and M is the total number of pixels in the image. This loss penalizes large deviations between the predicted and target images.

Evaluation Metrics: SSIM: Some models used the SSIM that evaluates the similarity between two images in terms of luminance, contrast, and structure:

$$\mathrm{SSIM}(I_{\text{ref}}, I_{\text{enh}}) = \frac{(2\mu_{\text{ref}}\mu_{\text{enh}} + C_1)(2\sigma_{\text{ref,enh}} + C_2)}{(\mu_{\text{ref}}^2 + \mu_{\text{enh}}^2 + C_1)(\sigma_{\text{ref}}^2 + \sigma_{\text{enh}}^2 + C_2)}, \tag{2}$$

where μ_{ref} and μ_{enh} are the mean pixel intensities of the reference and enhanced images, σ_{ref}^2 and σ_{enh}^2 are the variances, and $\sigma_{\text{ref,enh}}$ is the covariance. Constants C_1 and C_2 stabilize the division.

4. Experiments

All the models are trained using ADAM optimizer with learning rate of 0.001 on the three datasets [section 3]. The seed was set to 41 for all operations that required randomization, i.e. shuffling.

Table 1. Comparison of Loss and Validation Loss Across Models

Model	Training Loss	Validation Loss
U-Net	0.0024	0.0060
EnhanceNet	0.0614	0.0841
ResNet	0.0069	0.0065
FC-DenseNet	0.0055	0.0139
UICE^2-Net	0.0072	0.0077

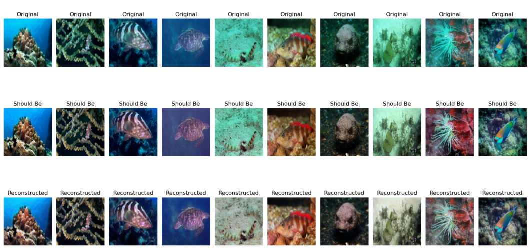

Figure 5. ResNet

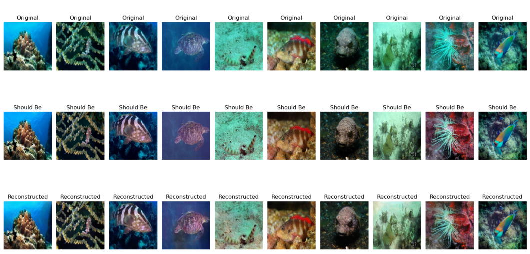

Figure 6. U-Net

Discussion

The results of our experiments demonstrate the varying performance of different deep learning models for underwater image enhancement, reflecting their design principles and underlying mechanisms.

U-Net showcased a robust capability to generalize, with relatively low training and validation losses. This can be attributed to its encoder-decoder architecture and skip connections, which effectively preserve spatial and contextual features. Its strong performance highlights its potential for underwater image enhancement, particularly in scenarios requiring detailed reconstruction.

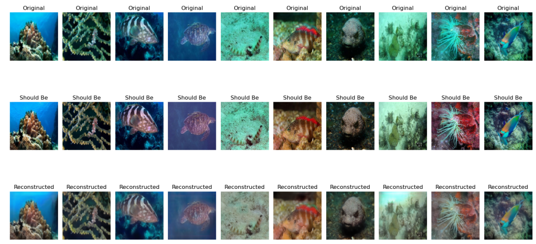

Figure 7. EnhanceNet

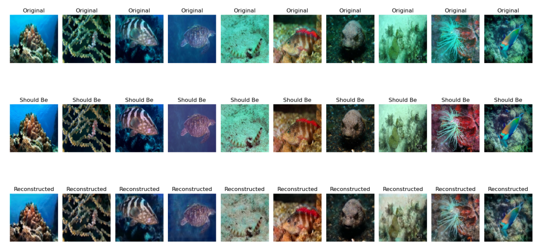

Figure 8. Seif

Figure 9. FCDenseNet

EnhanceNet, despite its more complex architecture, exhibited higher losses during both training and validation. This could be due to its focus on perceptual features and high-frequency details, which may have struggled with the subtle color and lighting variations inherent in underwater images. While it may excel in enhancing certain aesthetic aspects, its general effectiveness appears more limited in this context.

ResNet performed consistently across training and validation datasets, indicating its strength in feature extraction. Its relatively lower loss suggests that its residual connections helped address vanishing gradient issues, allowing for more effective learning even in challenging underwater conditions.

FC-DenseNet achieved promising results, particularly in training, due to its efficient parameter usage and feature concatenation strategy. However, its higher validation loss hints at potential overfitting or challenges in generalizing to unseen underwater scenes. This points to the need for further fine-tuning or additional data augmentation techniques to enhance its adaptability.

Lastly, UICE^2-Net demonstrated competitive performance by leveraging both RGB and HSV color spaces. This unique dual-pathway architecture allowed it to capture complementary features, resulting in well-balanced enhancements. Its consistent training and validation losses highlight its potential for real-world applications, especially where both brightness and chromatic adjustments are crucial.

Overall, the results suggest that while all models contribute uniquely to underwater image enhancement, their suitability varies depending on the specific requirements of the application, such as real-time processing, visual fidelity, or computational efficiency.

Conclusion

This study evaluated several state-of-the-art deep learning models for underwater image enhancement, focusing on their effectiveness across diverse datasets. The findings reveal that each model offers distinct advantages: U-Net excels in spatial preservation, ResNet in robust feature extraction, FC-DenseNet in parameter efficiency, and UICE^2-Net in comprehensive color and brightness adjustment.

While significant progress has been made in underwater image enhancement, challenges such as dataset diversity and real-world generalization remain. Future work could explore hybrid architectures, leveraging the strengths of multiple models, or focus on more extensive datasets encompassing a broader range of underwater conditions. Additionally, incorporating real-time constraints and computational efficiency into model design would further enhance their applicability in autonomous underwater exploration and marine research.

By addressing these challenges, we move closer to fully unlocking the potential of deep learning for understanding and exploring the underwater world.

References

[1] C. O. Ancuti, C. Ancuti, C. D. Vleeschouwer, and P. Bekaert. Color balance and fusion for underwater image enhancement. *IEEE Transactions on Image Processing*, 27:379–393, 2018. 2

[2] R. Hummel. Image enhancement by histogram transformation. *Computer Graphics and Image Processing*, 6(2):184–195, 1977. 1

[3] M. J. Islam, Y. Xia, and J. Sattar. Fast underwater image enhancement for improved visual perception. *IEEE Robotics and Automation Letters*, 5(2):3227–3234, 2020. 1, 2

[4] A. Koschinsky, L. Heinrich, K. Boehnke, J. C. Cohrs, T. Markus, M. Shani, P. Singh, K. Smith Stegen, and W. Werner. Deep-sea mining: Interdisciplinary research on potential environmental, legal, economic, and societal implications. *Integrated Environmental Assessment and Management*, 14(6):672–691, 2018. 1

[5] P. Liu, G. Wang, H. Qi, C. Zhang, H. Zheng, and Z. Yu. Underwater image enhancement with a deep residual framework. *IEEE Access*, 7:94614–94629, 2019. 1

[6] J. Manley. Autonomous underwater vehicles for ocean exploration. In *Oceans 2003. Celebrating the Past ... Teaming Toward the Future (IEEE Cat. No.03CH37492)*, volume 1, pages 327–331 Vol.1, 2003. 1

[7] L. Peng, C. Zhu, and L. Bian. U-shape transformer for underwater image enhancement. *IEEE Transactions on Image Processing*, 32:3066–3079, 2023. 1, 2

[8] Y. Rathore and S. Sain. *Exploration and Exploitation of the Sea for Research and Scientific Development*. PhD thesis, Your Institution Name, 09 2024. 1

[9] R. Schettini and S. Corchs. Underwater image processing: State of the art of restoration and image enhancement methods. *EURASIP Journal on Advances in Signal Processing*, 2010, 12 2010. 2

[10] X. Xu, H. Cai, M. Wang, W. Chen, R. Zhang, and T. Zhao. Exploring underwater image quality: A review of current methodologies and emerging trends. *Image and Vision Computing*, 154:105389, 2025. 1

[11] M. Zhang, Y. Li, and W. Yu. Underwater image enhancement algorithm based on adversarial training. *Electronics*, 13(11), 2024. 2

[12] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks, 2020. 2

[13] K. J. Zuiderveld. Contrast limited adaptive histogram equalization. In *Graphics gems*, 1994. 1