Enhancement and Classification of Underwater Images Using LightDehazeNet and Zero-Shot Learning

Shreya Doddamani[1], Pushpa Kalakamb[2], Nikita Shanbhag[3],
Samruddhi Kore[4], Channabasappa Muttal[5], Sneha Varur[6]

[1, 2, 3, 4]Author's Designation, School of Computer Science and Engineering, KLE Technological University, Hubballi
[5, 6]Co-Author's Designation, School of Computer Science and Engineering, KLE Technological University, Hubballi

Email: [1]01fe22bcs207@kletech.ac.in, [2]01fe22bcs205@kletech.ac.in, [3]01fe22bcs111@kletech.ac.in,
[4]01fe22bcs014@kletech.ac.in, [5]channabasappa.muttal@kletech.ac.in, [6]sneha.varur@kletech.ac.in
Contact: 1Author's contact no., 2Co-Author's contact no., etc.

Abstract: Underwater image processing is crucial for various applications, but the quality of underwater images is often degraded due to factors such as light attenuation and dispersion in the aquatic environment. This study proposes a two-stage technique to address these issues, involving image enhancement using LightDehazeNet and classification using Zero-Shot Learning (ZSL). The enhancement stage employs LightDehazeNet to remove haze, adjust contrast and sharpen the images using methods like Contrast Limited Adaptive Histogram Equalisation (CLAHE), gamma correction and sharpening. In the classification stage, a pre-trained CLIP model is used with ZSL to classify images through text-image correspondence, eliminating the need for task-specific training data. The proposed work's performance is evaluated using metrics such as PSNR and SSIM for image enhancement and accuracy for classification. The results demonstrate the work's effectiveness in enhancing underwater images and classifying them accurately. The proposed work achieves a PSNR of 27.39 ± 3.37 dB and an SSIM of 0.81 ± 0.08, comparable to state-of-the-art methods. The Classification accuracy of 86% using ZSL highlights the ability to distinguish between various image classes without extensive labelled data. This research benefits the advancement of underwater imaging applications in fields such as marine biology, navigation, underwater exploration, and environmental monitoring.

Index Terms: Underwater image enhancement, LightDehazeNet, Zero-Shot Learning, CLIP model, PSNR, SSIM, underwater classification.

I. INTRODUCTION

The ocean, encompassing 71% of the Earth's surface and spanning 360 million square kilometres, harbours abundant resources. Human development has long been intertwined with exploring and utilizing the ocean. However, as resource scarcity intensifies, the exploration and development of the ocean have become imperative for human progress[1]. The ocean holding vast resources, makes marine exploration a key focus. Underwater image processing plays a crucial role in gathering and using marine data[2]. Underwater images are important in advancing knowledge and technology in so many fields. Marine biology and ecology analyse the images to identify species and monitor health. This helps archaeologists discover shipwrecks, and preserve submerged cultural heritage. Underwater images also help fisheries and aquaculture monitor fish populations and farm management. Applications of underwater images include navigation, where they help in obstacle detection and pipeline inspections, and underwater robotics for tasks like SLAM and autonomous navigation. Oceanography and geology benefit from seafloor mapping and current studies, while defence uses images for mine detection and border security. These underwater images also help tourism with virtual diving and promote deep-sea exploration of uncharted areas. Autonomous underwater vehicles (AUVs) and remotely operated vehicles (ROVs) are part of these developments, according to research[3], [4]

However, the underwater environment presents several challenges compared to the above-water environment[5]. Underwater images suffer from poor visibility, color distortion, and dispersion as a result of light attenuation in the water, absorption of light of various wavelengths, and light blocking and scattering by some suspended particles[3]. As mentioned in [6] red color in atmospheric light is absorbed early due to its shorter wavelength, whereas the colors like blue and green penetrate deeper into the water due to larger wavelength. As a result, the underwater images appear bluish or greenish in color. Poor image quality, often caused by sensor limitations and suboptimal lighting, can significantly hinder visibility and make tasks like object detection, tracking, and surveillance challenging[7]. These issues will require some advanced visual tasks and will make some research difficult to complete. As a result, a better solution for recovering the color, visibility, clarity, and contrast of underwater images is critically needed. For which there has been plenty of work done with different kinds of models in which the most common and frequently used are the Dark Channel Prior(DCP)[8] and Contrast-Limited Adaptive Histogram Equalization (CLAHE)[9]. Improved underwater image quality allows for better observation and understanding of the marine ecosystem, which is beneficial to the development of marine engineering and resources, as well as the navigation and localization of underwater robots, target identification and sensing, and other tasks[10].

The proposed work deals with the problems of image enhancement and classification using LightDehazeNet[11] and ZSL[12] respectively. In this study, the images are first enhanced by removing the haze and noise and adjusting the contrast and saturation level; second, image classification is done using the ZSL method to identify or detect the images and classify them into the respective classes accordingly.

The paper's organisation is as such that the section II deals with the exploration and study of previously done works, i.e., the Literature Review. The section III presents the methodology. The section IV has the results and discussions of the proposed work. The section V concludes the paper, and the section VI describes the limitations and VIIsuggests potential improvements for the proposed work in the upcoming days.

II. Related Work

Attempts have been made to improve underwater images and resolve issues like haze, low contrast, degradation of spatial resolution, and uneven illumination by using super-resolution techniques. The UIEB dataset containing 950 authentic underwater images has become a common reference for many research projects. An example is Water-Net, which is a CNN-based model trained on UIEB and effectively tackles these obstacles[13]. In the same way, the LSUI dataset, containing more than 5000 image pairs, has been used to create advanced models like LU2Net, which combines LCH color spaces and SSIM loss functions to offer effective underwater image enhancement in a lightweight manner[14]. The combination of Super-Resolution CNN (SRCNN) and Multi-Scale Retinex (MSR) has shown significant improvements in spatial resolution and defogging in recent works[15]. Enhanced Water-Net, which comprises feature transformation blocks has been tested on the UIEB and EUVP datasets, and it was shown that the network can be adapted to different underwater environments[16]

The authors of [6] tried to improve the quality of underwater images by reducing the color distortion and loss of detail generated by water absorption. Their solution consisted of a two-step procedure: color balance correction, to shift the color of the image to such a value that colors got closer to their natural appearance and sharpening, applying a sharpening technique to boost the details of the image and show up the edges and textures. Therefore, after applying these two above-mentioned techniques, the authors were able to generate underwater ages that are both visually appealing and informative. In Research[17] the authors have highlighted various methods such as color correction, contrast enhancements, and dehazing, and discuss datasets and evaluation metrics used in this field. Recent developments such as Deep Reinforcement Learning (DRL) and Variational Autoencoders (VAE) have been developed to improve the quality of image reconstruction underwater by adjusting provides for feedback from the environment in real-time[17]. Techniques such as multi-channel histogram equalization and depth-adaptive corrections are essential for superior image enhancement in real-world scenarios, addressing issues with uneven lighting and depth-dependent distortions[18]. Historical methods like Dark Channel Prior (DCP) and Retinex-based techniques are still important to enhance contrast and remove haze[13]. In addition, ESRGCNN makes use of residual learning and adaptive up-sampling techniques to generate clear images based on low-quality input captures, effectively capturing detailed information from distant contexts[19]. These improvements are evaluated with popular metrics like PSNR, SSIM, UCIQE, and UIQM showing significant improvements in image quality in both subjective and objective assessments. The paper [20] addresses the integration of a three-in-one problem as restoration of color, de-hazing, and enhancement in contrast, and fast algorithms used in enhancing the quality of underwater images without compromising the methods' efficiency to enable real-time applications.

The paper[21] compares different GAN techniques used for underwater image enhancement and concludes that the CycleGAN and UGAN qualitatively and quantitatively improved the image quality in various underwater environments, and FUnIE-GAN had performance differences depending on the underwater environment. The results of [22] reveal that the CNN-based approaches produce satisfactory outcomes as they tend to yield blurry images. In contrast, GAN-based models not only provided clear and detailed images but also demonstrated superior performance in terms of quantitative assessment, underlying the potential of the framework beyond the specific land type investigated. The experimental results of [23] say that the CLAHE method performs better than Contrast Stretching and Histogram Equalisation in both color spaces. The research of [7] noted that deep learning algorithms have made significant advances in picture improvement under low-light circumstances. However, these methods fail to collect information about fine-grained local structures, limiting their performance, this resulted in the development of a generative model LightNet to improve the quality of photos taken in low light. This model will be able to capture both the overall details and the fine-grained textures of an image using a hierarchical generator encompassing an encoder-decoder module.

After exploring these research papers, some limitations were observed where some models failed under challenging conditions like backscatter or extreme lighting conditions, whereas some models struggled with color distortions. The water-net method generally failed under challenging conditions like the above-mentioned backscatter and in the areas of complex lighting. Few models used synthetic datasets, their models did not represent real-world scenarios. The models that used the GAN needed further optimizations to improve performance in practical applications. Research in[17] highlighted gaps such that existing methods often struggle to generalise across different underwater environments and there is a shortage of diverse, high-quality datasets for evaluating enhancement techniques. The major gap is that many methods fail to address real-time, practical challenges like varying light conditions and dynamic water clarity.

III. Methodology

The proposed approach offers a comprehensive and robust method for enhancing underwater images using LightDehazeNet and Zero-Shot Learning techniques.

A. Dataset Description

The UIEB (Underwater picture improvement Benchmark) dataset, already studied in [13], is a specialized collection made for enhancing underwater image enhancement algorithms. Consisting of 890 images underwater from diverse subaquatic environments, as depicted in Figure2., the dataset succeeds in capturing the different conditions in the water, along with the lighting and visibility available. It displays some common underwater image degradation factors, including blur and low contrast, that occur as a result of the light penetration and scattering effects. Being a standardized benchmark, the UIEB dataset can be used for comparing and evaluating the different enhancement techniques. As indicated in [13], qualitative benchmarks for underwater image processing have not been addressed. Consequently, this dataset has become a widely utilized resource in research, particularly for evaluating augmentation algorithms based on deep learning models.

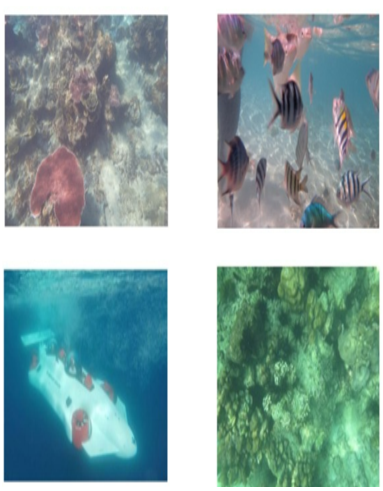

Fig. 1: Sample images from the UIEB dataset which will be used as the input for our work.

The proposed work Figure2. operates in two steps one is image enhancement and the other is image classification. Adjustments to the contrast and saturation of images were also included during the enhancements to address the issue of darkness after the dehazing of images.

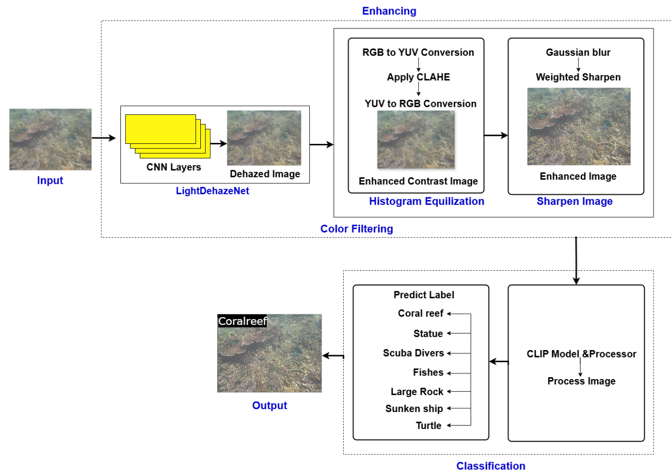

Fig. 2: Model Architecture explains the working of the proposed work, in which the underwater image when given as input to the model is enhanced and then classified to its respective class.

B. Image Enhancement

The proposed work in this part improves the underwater image clarity by addressing haze, contrast, noise, and colour distortion.

Dehazing: During stormy weather, contaminated suspended atmospheric particles reduce image quality, affecting overall surveillance systems. To address these challenges, LightDehazeNet[11] is used.

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

Here, in Equation 1 $J(x)$ is the estimated clean (dehazed) image, which the algorithm aims to reconstruct. $I(x)$ is the input hazy image that has been degraded by atmospheric particles. A represents the global atmospheric light, which is assumed to be a constant value contributing to the haze's brightness. $t(x)$ is the transmission map, which represents the portion of the light that reaches the camera without scattering. It utilizes a convolutional neural network to remove haze while preserving the other details. This architecture includes multiple convolutional layers and ReLU activations.

Contrast Enhancement: Uses Contrast Limited Adaptive Histogram Equalization (CLAHE)[9] which is the adaptive histogram equilisation to enhance local contrast.

$$T(p) = \frac{\text{CDF}(p)}{\text{CDF}_{\text{max}}} \cdot (L - 1) \tag{2}$$

In Equation 2, $T(p)$ is the transformed intensity for a pixel value p, CDF(p) represents the cumulative distribution function up to intensity p, CDF_{max} is the maximum value of the cumulative distribution function, and L denotes the total number of intensity levels (e.g., 256 for an 8-bit image). This formula is applied to redistribute pixel intensities for enhanced contrast while maintaining the original image's overall brightness levels.

Gaussian Blur: Gaussian Blur[24] is a smoothening filter that removes noise and fine details by smoothening high-frequency components of the image. It provides a softened, low-frequency version of the original image. It enhances the original image by selectively amplifying the differences between the original and the blurred version.

Sharpening: Highlights the edges in images to make them appear clearer. It applies a convolutional filter[25][26]to do so.

Metrics Calculation: It evaluates image quality using PSNR(Peak Signal-to-Noise Ratio)[27] and SSIM(Structural Similarity Index Matrix)[28]

C. Image Classification

The model uses a pre-trained CLIP model to perform the classification using the ZSL[12] method. CLIP[29] uses a pre-trained model to learn visual concepts from large-scale datasets of images and text. CLIP can then perform ZSL image classification by leveraging textual descriptions of classes without needing task-specific training data[30]. **Zero Shot Learning**: Traditional supervised learning requires labelled examples for all classes, which is impractical for tasks with a vast or open-ended set of categories. ZSL[12] seeks to address this limitation by enabling classification without requiring labelled examples for every target class.

$$\hat{y} = \arg \max_{y \in \mathcal{Y}_{\text{unseen}}} S(f(x), g(y)) \tag{3}$$

In Equation 3 x is the input data or feature vector (e.g., an image feature), $f(x)$ is the feature extractor that maps x to a shared semantic space, y is the class label, $\mathcal{Y}_{\text{unseen}}$ is the set of unseen class labels, $g(y)$ is the representation of the class y in the shared semantic space (e.g., attribute vector or word embedding), and $S(\cdot, \cdot)$ is the similarity function (e.g., cosine similarity or dot product).

Preprocessing: Converts the images to the required format for CLIP[29] and tokenizes the class labels into text embeddings. Multiple images are processed together for efficiency using Batch Processing.

Feature Extraction: Text-Image Matching extracts features from images using CLIP's image encoder and computes text embeddings for class labels.

Similarity Calculation: Computes cosine similarity[31] between image and text embeddings to predict the most likely class.

Accuracy Calculation: Compares predicted labels with true labels and computes accuracy as the percentage of correct predictions. Ground truth tables are generated randomly for illustration.

Visualisation: Displays images with predicted class labels for easy verification Figure**??**

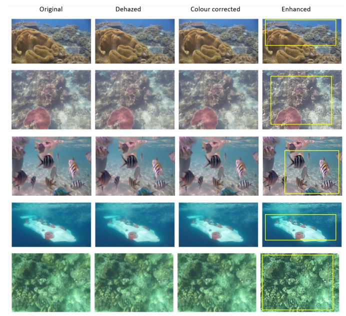

Fig. 3: Image Enhancement Results: Input image is given to the model, where first the haze is removed, then the color is corrected and finally the input image is enhanced

IV. Results and Discussions

The proposed work was carefully examined for both tasks involving underwater image enhancement and classification.

A. Image Enhancement :

To evaluate the performance and effectiveness of the proposed work, several other state-of-the-art techniques have been used for the comparison experiments using the PSNR and the SSIM. The results are summarised in Table **??**

TABLE I: Performance Comparison of Underwater Image Enhancement Models

Model	PSNR	SSIM
Light weight	25.49	0.868
Enhanced Water-Net	23.15	0.81
UIEB PDCFNet	27.37 ± 5.27	0.92 ± 0.06
Proposed work	**27.39 ± 3.37**	**0.81 ± 0.08**

The proposed work achieved a PSNR of 27.39 ± 3.37 dB, closely matching the performance of UIEB PDCFNet[32] (27.37 ± 5.27 dB). This shows the ability of the proposed work that is able to effectively reconstruct images with minimal amount of noise while keeping the quality high. The SSIM of 0.81 ± 0.08 attained by the proposed work, which is lower than UIEB PDCFNet[32], reporting a value of 0.92 ± 0.06. It can be argued however that while the SSIM reflects a slight amount of compromise in terms of structural similarity, the proposed work still provides visually pleasing results for underwater scenes. The results of the Image Enhancement in the current research investigation can be seen in Figure3.

B. Image Classification

The proposed classification method, based on the ZSL CLIP model, achieved an accuracy of 86.67%. This demonstrates the model's ability to effectively classify underwater images as shown in Figure4 without requiring task-specific training data. Visual inspection of predicted class labels supports the consistency of the results. The proposed work, in general, possesses a great ability to work on the two problems of enhancement and classification of underwater images. It can be regarded as a practical solution for underwater tasks given

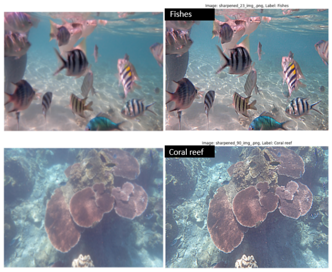

Fig. 4: Image Classification: The Output of LightDehazeNet is given to the ZSL model to classify them to their respective classes.

its balanced performance during execution and its efficiency and versatility.

V. Conclusion

This study aimed to enhance and classify underwater images using modern approaches. The LightDehazeNet [11] model was employed to remove haze, improving visibility and contrast. Additionally, level changes improved visual quality. For classification, the ZSL [12] method was used, requiring minimal training data to classify images. PSNR [27] and SSIM [28] metrics were used to evaluate image enhancements, generating results of 27.39 ± 3.37 dB and 0.81 ± 0.08, respectively, outperforming previous methods. Classification accuracy was 86.67%, validating the ZSL method.

This combination of techniques shows promise for underwater image processing, relevant to marine biology, underwater exploration, and environmental monitoring. However, limitations exist, such as the performance of LightDehazeNet under extreme conditions and the dependence of ZSL on the quality and diversity of the data set. Future work could involve incorporating more advanced enhancement techniques to handle diverse underwater conditions, potentially improving accuracy and reliability beyond the current framework.

VI. Future Work

Despite producing positive results, there will be a few limitations. Extreme underwater conditions, such as severe distortion or loss of color images, could improve LightDehazeNet's performance. in the same sense when we talk about ZSL, the performance may also depend on how good and how diverse the datasets used are. In the future, we could be able to add more sophisticated enhancing techniques in order to handle a wider variety of underwater situations. Perhaps, combining more complex procedures would increase the accuracy and reliability. These challenges may bring opportunities for improvement beyond the results of the proposed framework.

References

[1] K. Hu, C. Weng, Y. Zhang, J. Jin, and Q. Xia, "An overview of underwater vision enhancement: From traditional methods to recent deep learning," *Journal of Marine Science and Engineering*, vol. 10, no. 241, 2022. DOI: 10.3390/jmse10020241.

[2] Q. Xie, X. Gao, Z. Liu, and H. Huang, "Underwater image enhancement based on zero-shot learning and level adjustment," *Heliyon*, vol. 9, no. 4, 2023.

[3] A. Jamadandi, C. Desai, R. Tabib, and U. Mudenagudi, *Underwater image enhancement via group-wise deep whitening and coloring transforms*, 2019.

[4] M. Johnson-Roberson, M. Bryson, A. Friedman, *et al.*, "High-resolution underwater robotic vision-based mapping and three-dimensional reconstruction for archaeology," *Journal of Field Robotics*, vol. 34, no. 4, pp. 625–643, 2017.

[5] C. Desai, R. Tabib, S. Reddy, U. Patil, and U. Mudenagudi, "Ruig: Realistic underwater image generation towards restoration," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, IEEE, Nashville, TN, USA, 2021, pp. 2181–2189. DOI: 10.1109/CVPRW.2021.02719.

[6] Z. Patel, C. Desai, R. A. Tabib, M. Bhat, U. Patil, and U. Mudengudi, "Framework for underwater image enhancement," *Procedia Computer Science*, vol. 171, pp. 491–497, 2020.

[7] C. Desai, N. Akalwadi, A. Joshi, *et al.*, "Lightnet: Generative model for enhancement of low-light images," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2023, pp. 2231–2240.

[8] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1397–1409, 2010.

[9] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics Gems IV*, P. S. Heckbert, Ed., San Diego, CA, USA: Academic Press, 1994, pp. 474–485.

[10] K. Liao and X. Peng, "Underwater image enhancement using multi-task fusion," *PLoS ONE*, vol. 19, no. 2, e0299110, 2024. DOI: 10.1371/journal.pone.0299110.

[11] H. Ullah, K. Muhammad, M. Irfan, *et al.*, "Lightdehazenet: A novel lightweight cnn architecture for single image dehazing," *IEEE Transactions on Image Processing*, vol. 30, pp. 8968–8982, 2021.

[12] M. Palatucci, D. Pomerleau, G. Hinton, and T. Mitchell, "Zero-shot learning with semantic output codes," in *Advances in Neural Information Processing Systems*, vol. 22, 2009.

[13] C. Li, C. Guo, W. Ren, *et al.*, "An underwater image enhancement benchmark dataset and beyond," *IEEE Transactions on Image Processing*, vol. 29, pp. 4376–4389, 2019.

[14] H. Yang, J. Xu, Z. Lin, and J. He, "Lu2net: A lightweight network for real-time underwater image enhancement," *arXiv preprint arXiv:2406.14973*, 2024.

[15] Y. Gogireddy and J. Gogireddy, "Advanced underwater image quality enhancement via hybrid super-resolution convolutional neural networks and multi-scale retinex-based defogging techniques," *arXiv preprint arXiv:2410.14285*, 2024.

[16] R. Priyadharshini, S. Arivazhagan, K. Pavithra, and S. Sowmya, "An ensemble deep learning approach for underwater image enhancement," *e-Prime Advances in Electrical Engineering, Electronics, and Energy*, p. 100 634, 2024.

[17] S. Raveendran, M. Patil, and G. Birajdar, "Underwater image enhancement: A comprehensive review, recent trends, challenges, and applications," *Artificial Intelligence Review*, vol. 54, pp. 5413–5467, 2021.

[18] M. Mousavi, R. Estrada, and A. Ashok, "Idehaze: Supervised underwater image enhancement and dehazing via physically accurate photorealistic simulations," *Electronics*, vol. 12, no. 11, p. 2352, 2023.

[19] C. Tian, Y. Yuan, S. Zhang, C. Lin, W. Zuo, and D. Zhang, "Image super-resolution with an enhanced group convolutional neural network," *Neural Networks*, vol. 153, pp. 373–385, 2022.

[20] M. Islam, Y. Xia, and J. Sattar, "Fast underwater image enhancement for improved visual perception," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 3227–3234, 2020. DOI: 10.1109/LRA.2020.2974710.

[21] H.-G. Kim, J.-M. Seo, and S. M. Kim, "Comparison of gan deep learning methods for underwater optical image enhancement," *Journal of Ocean Engineering and Technology*, vol. 36, no. 1, pp. 32–40, 2022.

[22] P. Kramer, A. Steinhardt, and B. Pedretscher, "Enhancing sentinel-2 image resolution: Evaluating advanced techniques based on convolutional and generative neural networks," *arXiv preprint arXiv:2410.00516*, 2024.

[23] S.-L. Wong, Y.-P. Yu, N. A.-J. Ho, and R. Paramesran, "Comparative analysis of underwater image enhancement methods in different color spaces," in *2014 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, 2014, pp. 034–038. DOI: 10.1109/ISPACS.2014.7024420.

[24] S. Bazeille, I. Quidu, L. Jaulin, and J. Malkasse, "Automatic underwater image pre-processing," in *CMM'06*, 2006.

[25] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[26] Y. LeCun, B. Boser, J. Denker, *et al.*, "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989.

[27] C. Fu, F. Lu, X. Zhang, and G. Zhang, "Joint dedusting and enhancement of top-coal caving face via single-channel retinex-based method with frequency domain prior information," *Symmetry*, vol. 13, no. 11, p. 2097, 2021.

[28] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[29] P. Sahu, N. Gupta, and N. Sharma, "A survey on underwater image enhancement techniques," *International Journal of Computer Applications*, vol. 87, no. 13, pp. 19–23, 2014.

[30] A. Radford, J. Kim, C. Hallacy, *et al.*, "Learning transferable visual models from natural language supervision," in *International Conference on Machine Learning*, PMLR, 2021, pp. 8748–8763.

[31] G. Salton and M. McGill, *Introduction to Modern Information Retrieval*. McGraw-Hill, 1989.

[32] S. Zhang, D. Li, and R. Zhao, "Pdcfnet: Enhancing underwater images through pixel difference convolution," *arXiv preprint arXiv:2409.19269*, 2024.