The Role of Image Enhancement in Underwater Image Classification

David Sousa Almeida

University of Porto (UP)

Abstract. This study delves examines the critical role of image quality in the classification of underwater images using machine learning models. Focusing on the Enhanced Underwater Vision Perception (EUVP) dataset, we explore how image enhancement influences classification accuracy. Our research highlights the impact of high-quality datasets in training effective machine learning models, particularly in challenging environments like underwater settings. Surprisingly, models trained on lower-quality images exhibited unique characteristics, suggesting new avenues for future research. This study contributes to underwater image analysis and underscores the broader implications of dataset quality in machine learning. Future directions include exploring more complex model architectures and advanced techniques like transfer learning, as well as analyzing misclassified examples for further iterative model improvement.

Keywords: Underwater Imaging · Image Enhancement · Machine Learning · Convolutional Neural Networks

1 Introduction

In the domain of underwater exploration and research, the task of automatic image classification plays an important part and, with the increasing adoption of algorithms based on machine learning (ML), an ever more relevant role. The aim of this project (undertaken in the scope of the course of "Foundations and Applications in Machine Learning") is to study some of the intricacies of ML-based underwater image classification, with a particular focus on assessing the impact of image quality and enhanced visual perception on model development. The Enhanced Underwater Vision Perception (EUVP) dataset [1] was employed to reach these research objectives, a rich collection of underwater images that present a diverse range of visual characteristics.

This project seeks to address the challenges posed by underwater image classification, a domain that is known to pose challenges such as poor visibility, varying light conditions, and distortions caused by water properties. These challenges significantly impact the performance of conventional image processing and machine learning techniques, demanding specialized approaches for effective analysis.

Central to this research is the development and comparison of various machine learning models, tailored to effectively handle the unique properties of underwater images. The project's methodology encompasses custom labeling of the dataset, an essential step for accurate classification and assessment, followed by the deployment of advanced machine learning techniques for categorizing and evaluating image quality.

In summary, this project aims to contribute to the field of underwater image analysis by employing a popular dataset in an unprecedented way. Thus, this work seeks to advance the understanding of how image quality affects classification accuracy by developing robust models capable of handling the complex nature of underwater imagery.

2 Literature Review

The field of underwater image analysis has been an area of active research, addressing the unique challenges posed by aquatic environments. This literature review synthesizes the key findings and theories relevant to underwater image classification and image quality assessment.

Recent studies have highlighted the primary issues in underwater imaging, such as light absorption and scattering, which lead to color distortion and reduced visibility. Researchers have explored various methods to counteract these effects, including color correction algorithms and image enhancement techniques specifically designed for underwater conditions [2, 3].

In the domain of machine learning, the focus has been on developing models that can adapt to the unpredictable and varied nature of underwater imagery. This includes the use of convolutional neural networks (CNNs) for image classification, which have shown promise in recognizing patterns and features in visually complex underwater scenes [4, 5].

The literature also emphasizes the importance of dataset quality and diversity in training effective machine learning models. Studies have shown that the accuracy of image classification is heavily dependent on the representativeness and quality of the training dataset, highlighting the need for well-labeled, high-quality underwater image datasets [6].

In summary, the existing body of research lays a foundational understanding of the challenges and solutions in underwater image processing and classification. This project builds upon these insights, aiming to further advance the field through a comprehensive analysis of the EUVP dataset and the development of optimized machine learning models for underwater image classification.

3 Dataset Description

The Enhanced Underwater Vision Perception (EUVP) dataset [1], elected for this research, is a comprehensive collection of underwater images. This dataset is unique due to its diverse range of underwater environments and conditions it represents, making it a valuable resource for training and testing ML models in the field of underwater image analysis.

The EUVP dataset comprises thousands of images that exhibit a wide array of challenges typical in underwater settings, such as varying degrees of visibility, light conditions, and color distortions. This diversity is crucial for developing robust models capable of handling real-world underwater scenarios. The dataset includes images from both natural underwater environments and artificial settings, providing a broad spectrum of data for analysis.

The selection of the EUVP dataset for this project is based on its relevance to the research objectives, which include understanding the impact of image quality on classification results. Essentially, every image in the dataset contains a paired version of enhanced quality (see Figure 1). The dataset's comprehensive nature allows for a thorough examination of these aspects, contributing to more generalized and effective ML solutions for underwater image processing.

Furthermore, the application of the EUVP dataset for image classification represents a novel shift in its usage, since the dataset was initially developed for image enhancement. This adaptation not only broadens the dataset's utility but also contributes a new perspective to the field of underwater image processing.

Fig. 1. Examples of paired images of the EUVP dataset - with original quality (left) and enhanced quality (right)

4 Methodology

The methodology of this research is anchored in a comprehensive approach towards underwater image classification, leveraging the EUVP dataset. This section details the steps taken, from data preparation to model development and evaluation. This methodology reflects a holistic and detailed approach to tackling the challenges in underwater image classification, aiming to develop robust and accurate models suited for diverse underwater environments and imbalanced datasets.

4.1 Data Preparation and Labeling

The labeling process was streamlined into two simple binary categories: 'Fish' (for images that contain fish) and 'None' (for images that do not contain fish). A total of 1306 images from the enhanced section were labeled using *Roboflow*, with 70% under 'Fish' and 30% under 'None'. Table 1 shows the chosen distribution of the whole labeled dataset between training, validation, and testing data.

Table 1. Division of Labeled Dataset (with percentages referent to the subsets of their respective table rows)

Data Subset	Total of Images	Images of class 'Fish'	Images of class 'None'
Complete	1306	910 (70%)	396 (30%)
Training	947	685 (72%)	262 (28%)
Validation	150	76 (51%)	74 (49%)
Testing	209	149 (71%)	60 (29%)

The validation subset was intentionally balanced between the two classes to aid in consistent model evaluation throughout the training process. Notably, the testing subset includes more images than the validation subset. This decision was made to ensure a statistically significant number of images, separate from the training process, despite the limited data availability. Additionally, the class imbalance inherent in the entire dataset was deliberately maintained in the testing set to provide a realistic assessment scenario.

To enhance the dataset and tackle the class imbalance, data augmentation was employed for the training data subset. The original images, resized to 128 x 128 pixels, were augmented to create different versions - one new version for each image labeled 'Fish' and four new versions for each 'None' image (on average). This led to a final augmented training data subset of 2612 images with a much more balanced class distribution (with 52 % of images belonging to the 'Fish' class and 48% to the 'None' class). Augmentation techniques included rotations, flips, image shearing, and elastic transformations to mimic underwater distortions. Figure 2 illustrates the resizing of a given image and its subsequent utilization for data augmentation.

4.2 Model Development

From the beginning, special emphasis was placed on CNNs, recognized for their effectiveness in image classification tasks. The models were designed to adapt to

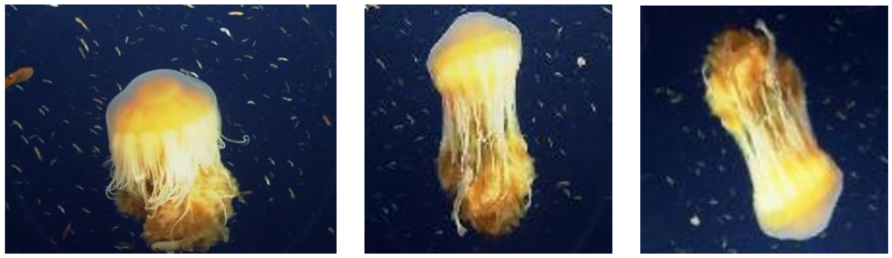

Fig. 2. Original Image (Left), Resized Image (Center), Augmented Version (Right)

the unique characteristics of underwater images, accounting for common issues like color distortion and reduced visibility. The models were developed and tested in Python with extensive use of *TensorFlow*. Iteratively, manual hyperparameter tuning was employed to optimize model performance (since computational limitations halted automated tuning methods).

The architecture of the final CNN was meticulously designed, incorporating convolutional layers with 32, 64, and 128 filters to capture image details. Each convolutional layer is followed by ReLU activations for non-linear processing, MaxPooling layers for reducing dimensionality, and Dropout layers (where the model randomly ignores a 25% of neurons) to prevent overfitting and avoid overreliance on particular features. A key aspect of the network's design was the use of L2 regularization within the convolutional layers, which helped to penalize larger weight values, thus steering the model towards identifying simpler patterns that could be more generalizable [4].

The design of the Convolutional Neural Network for this project culminated in a structure that transitions from 2D feature maps to a 1D vector through a flattening layer. This is essential for preparing the data for the final classification stages, which involve dense layers, including a prominent layer with 512 neurons. Another Dropout layer was incorporated before final activation (randomly ignoring 32% of neurons). The culmination of the network architecture is marked by the use of a sigmoid activation function in the final layer, which is particularly adept for binary classification tasks such as distinguishing 'Fish' from 'None'. This choice of activation function is deliberate, as it maps the final layer outputs to a probability between 0 and 1, thereby facilitating a clear and definitive classification.

The model's training was meticulously planned to ensure computational efficiency and minimize overfitting. Spanning 50 epochs and utilizing a batch size of 48, it employed binary cross-entropy as the loss function for its binary classification objective. Advanced methods like *ReduceLROnPlateau* and Early Stopping were integrated into the training regimen. *ReduceLROnPlateau* reduced the learning rate by half if no validation loss improvement was noted for five epochs, allowing for more nuanced model adjustments. Early Stopping terminated training if no improvement in validation loss was observed over twelve epochs, streamlining the training process and preventing overfitting.

This CNN's design and training approach underscored the project's commitment to creating a robust model capable of accurately classifying underwater images, with particular emphasis on the detection of fish. The iterative testing and crafting of the model's architecture was crucial in addressing the unique challenges presented by underwater imagery.

4.3 Model Evaluation

Appropriate metric selection was crucial due to the class imbalance present in the dataset. Using accuracy as the sole performance indicator in such scenarios would be inadequate. Thus, the usage of the F1 score, which combines precision and recall, poses as vital for an adequate assessment of our imbalanced dataset. Additionally, the ROC-AUC score is highlighted for its effectiveness in measuring the model's capability to differentiate between classes. A crucial part of the evaluation process is the use of a confusion matrix, which provides detailed insights into the model's performance, specifically in terms of error types and misclassifications.

This comprehensive approach to model evaluation is essential for a nuanced understanding of performance in the context of imbalanced datasets. Details on the employed metrics and evaluation methods may be found in [4].

5 Results and Discussion

Two models underwent evaluation. The first was trained using 947 enhanced images, all properly labeled as described in section 4.1. The second model was similarly trained but with 947 original quality images, naturally maintaining identical labeling. This training approach was also implemented for the validation process, using 150 labeled images from the validation subset for each model.

Both models were subjected to tests using two distinct subsets of 209 images each, one subset containing enhanced quality images and the other consisting of images in their original quality. The results of these tests are summarized in Table 2. The optimal thresholds were calculated by determining the point closest to the top-left corner of the plot on the respective Receiver Operating Characteristic (ROC) curves.

Figure 3 shows the plotted ROC curves for the model trained with enhanced data. Figures 4 and 5 show the Confusion Matrices for the same model, tested with enhanced and original data, respectively. All metrics and graphs indicate that this first model performed better when tested with the enhanced images.

On the other hand, Figure 6 shows the (ROC) curves for the model trained with the original data. Figures 7 and 8 show the Confusion Matrices for the same model, tested with enhanced and original data, respectively. The ROC-AUC scores in this case were worse than in the first model. However, when using the respective optimal thresholds, this second model performed better than the

Table 2. Overview of results - with Accuracy, Precision, Recall, and F1 calculated at the optimal thresholds (Opt. Thresholds).

Train Quality	Test Quality	ROC-AUC	Opt. Threshold	Accuracy	Precision	Recall	F1
Enhanced	Enhanced	0.69	0.77	0.63	0.90	0.54	0.67
Enhanced	Original	0.66	0.71	0.50	0.88	0.34	0.50
Original	Enhanced	0.64	0.35	0.71	0.78	0.83	0.80
Enhanced	Original	0.65	0.80	0.57	0.85	0.49	0.62

first one; with notably improved F1 scores when tested with either the enhanced or the original data.

Model performance across different training and testing configurations was variable, reflecting the complexities of image classification in underwater conditions. The optimal thresholds and metric scores such as F1, accuracy, and precision indicate a model that has learned to an extent but also displays notable room for improvement. Precision generally exceeded recall, indicating a model that is cautious in predicting the presence of fish, but at the cost of missing several true positives; which could be due to a variety of factors, including possible class imbalance, feature extraction limitations, or data quality issues.

Essentially the disparity in model performance suggests that image quality has a tangible impact on model efficacy, aligning with the initial hypothesis. However, the performance of the model trained with original data and tested with enhanced data was surprising. This observation might indicate that the model becomes more robust when trained with lower-quality data, or it could simply be an unwanted product of image resizing (performed only due to computational constraints) invalidating our image quality-based assessment.

6 Conclusion

In this report, we have explored the critical role of image quality in the classification of underwater images using machine learning. Our analysis, utilizing the Enhanced Underwater Vision Perception (EUVP) dataset, has demonstrated that image enhancement can significantly impact the accuracy of classification models. This finding underscores the importance of high-quality datasets in training effective machine learning algorithms, especially in challenging environments like underwater scenery. Surprisingly, models trained on lower-quality images exhibited unique characteristics, suggesting potential areas for future research.

This study not only contributes a novel approach to the study of underwater image analysis but also highlights the broader implications of dataset quality in machine learning. Future work should focus on addressing the encountered challenges, possibly through further data augmentation, exploring more complex model architectures, or employing advanced techniques like transfer learning to leverage pre-trained models. Additionally, specifically investigating the misclassified examples could yield insights for further iterative model improvement.

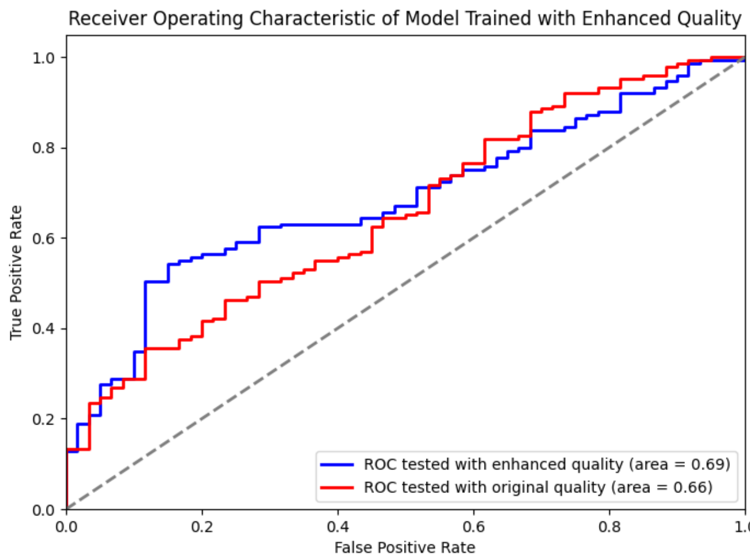

Fig. 3. ROC curves of model trained with enhanced data and tested with either enhanced or original data.

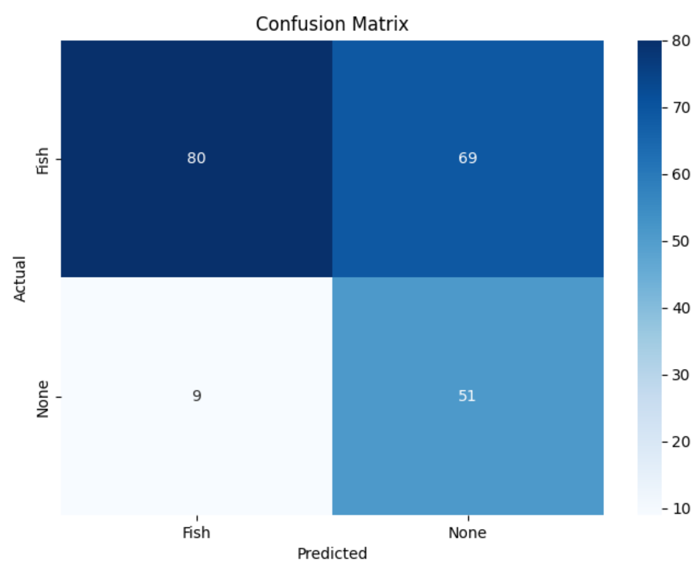

Fig. 4. Confusion Matrix of Model trained and tested with enhanced data (while using the respective optimal threshold)

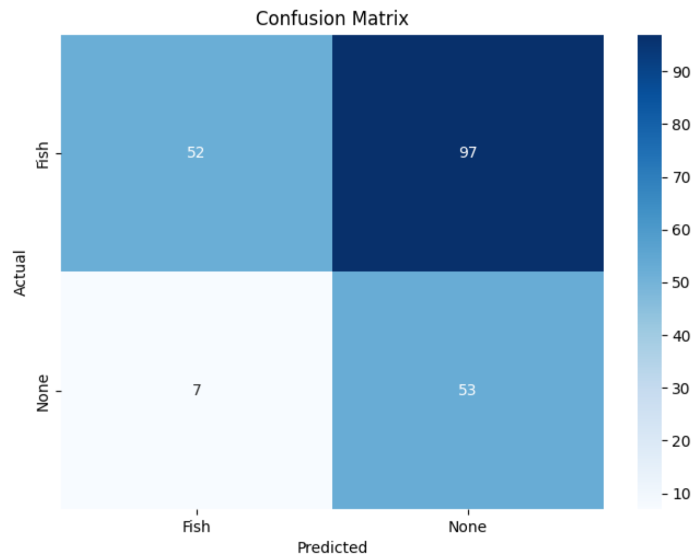

Fig. 5. Confusion Matrix of Model trained with enhanced data and tested with original data (while using the respective optimal threshold)

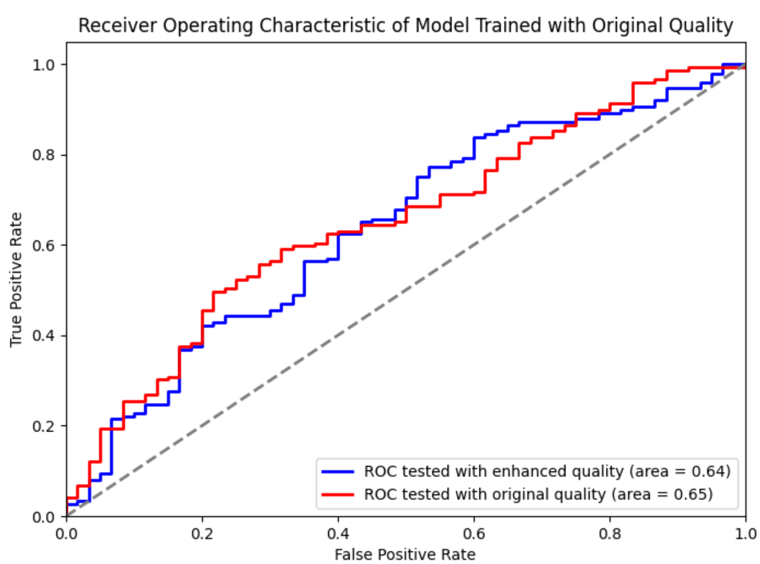

Fig. 6. ROC curves of model trained with original data and tested with either enhanced or original data.

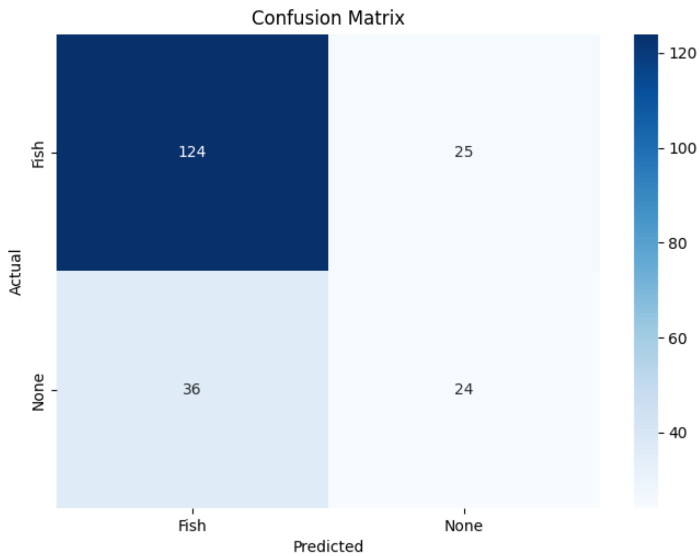

Fig. 7. Confusion Matrix of Model trained with original data and tested with enhanced data (while using the respective optimal threshold)

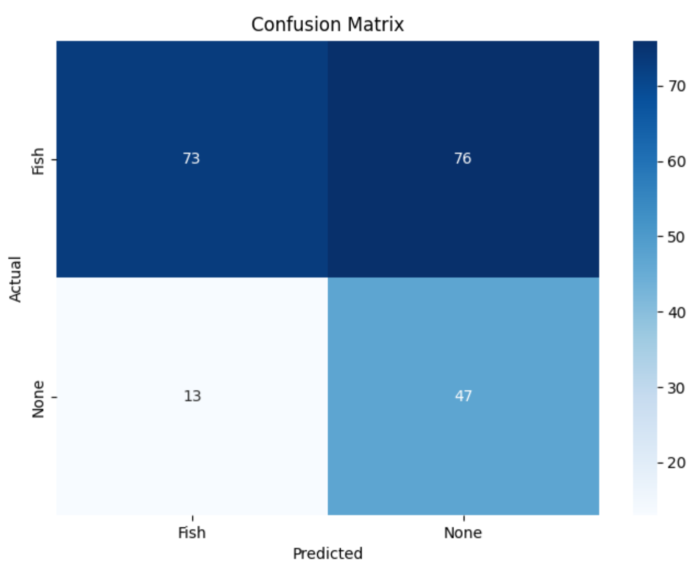

Fig. 8. Confusion Matrix of Model trained and tested with original data (while using the respective optimal threshold)

References

1. M. J. Islam, Y. Xia and J. Sattar, "Fast Underwater Image Enhancement for Improved Visual Perception," in IEEE Robotics and Automation Letters, vol. 5, no. 2, pp. 3227-3234, April 2020, doi: 10.1109/LRA.2020.2974710.
2. Mahmoud, Yasmin & Sakr, Nehal & El-Sappagh, Shaker & Abuhmed, Tamer & Elmogy, Mohammed. (2023). Underwater Image Restoration and Enhancement: A Comprehensive Review of Recent Trends, Challenges, and Applications. 10.20944/preprints202307.0585.v1.
3. Kumar, Ram and Patil, Manoj, Improved the Image Enhancement Using Filtering and Wavelet Transformation Methodologies (July 22, 2022). Available at SSRN: https://ssrn.com/abstract=4182372 or http://dx.doi.org/10.2139/ssrn.4182372
4. Ian Goodfellow, Yoshua Bengio, Aaron Courville, "Deep Learning", MIT Press, 2016.
5. Mittal, Sparsh & Srivastava, Srishti & Jayanth, J. (2021). A Survey of Deep Learning Techniques for Underwater Image Classification. 10.13140/RG.2.2.25098.59846.
6. Drews-Jr, Paulo & Souza, Isadora & Maurell, Igor & Protas, Eglen & Botelho, Silvia. (2021). Underwater image segmentation in the wild using deep learning. Journal of the Brazilian Computer Society. 27. 10.1186/s13173-021-00117-7.