CS4240 Deep Learning Replication Project – Group 60

Multi-view underwater image enhancement method via embedded fusion mechanism

Emran Yasser Moustafa - e.y.moustafa@student.tudelft.nl - 6078028

Rezoan Hassan Alom - ralom@student.tudelft.nl - 6062466

Shreya Sebastian - sasebastian@tudelft.nl - 5957451

Christina Vogel - c.vogel@student.tudelft.nl - 6119093

Name	**Contributions**
Emran	Building the training loop, loss function and model
Rezoan	Building evaluation metrics, performing inference, debugging/fixing WB and CLAHE functions
Shreya	Building the model, some data preprocessing, writing the blog post
Christina	Implementing WB and CLAHE functions, making the poster, editing the blog post

We presented at the poster presentation session.

1. Overview

In this blog post, we delve into the fascinating realm of underwater image enhancement, exploring the challenges faced by underwater imagery and the existing solutions. Building upon this foundation, we introduce the novel approach proposed by Jingchun Zhou, Jiaming Sun, Weishi Zhang, and Zifan Lin in their 2023 paper, “Multi-view underwater image enhancement method via embedded fusion mechanism.” Within this paper, the authors present the Multi-Feature Underwater Image Enhancement Method via Embedded Fusion Mechanism (MFEF), offering a promising solution to the complexities of underwater imaging.

We worked on replicating the network of this paper and will explain our reproduction approach, the challenges we faced, and present our results.

2. Introduction

Underwater imagery plays a crucial role in marine research, resource exploration, and environmental monitoring, providing valuable insights into a mostly unexplored underwater world. However, capturing clear and detailed underwater images is challenging due to the complex light absorption and scattering properties of water, which often result in images that have a colour cast, are blurred, or distorted.

Model-free methods are effective but typically only for specific types of underwater images, often overlooking or irreparably ignoring image details. Physical model-based methods analyse factors that degrade image quality and use physical models to estimate background light and transmission maps, aiming to restore clear images. However, they rely on simplified models of image degradation, leading to inaccurate restorations. Moreover, the effectiveness of these methods is limited by uncertainties related to their prior assumptions, impacting their overall performance.

The Multi-Feature Underwater Image Enhancement Method via Embedded Fusion Mechanism (MFEF) presents a novel approach to address these challenges, making use of advanced deep learning techniques to significantly enhance the quality of underwater images.

3. Abbreviations

MFEF - Multi-Feature Underwater Image Enhancement Method via Embedded Fusion Mechanism

WB - White Balance

CLAHE - Contrast-Limited Adaptive Histogram Equalization

MFF - Multi-Feature Fusion

PCAM - Pixel-Weighted Channel Attention Module

REM - Residual Enhancement Module

UIEB - Underwater Image Enhancement Benchmark

PSNR - Peak Signal-To-Noise Ratio

SSIM - Structural Similarity Index Metric

PCQI - Patch-Based Contrast Quality Index

UIQM - Underwater Image Quality Measure

UIConM - Underwater Image Contrast Measure

UISM - Underwater Image Sharpness Measure

UICM - Underwater Image Colorfulness Measure

4. The MFEF Model

Traditional underwater image enhancement techniques often struggle to effectively mitigate the multitude of issues that plague underwater imaging. Scattering and absorption of light causes images taken in this environment are often blurred and colour cast with a blue/green tint.Deep learning-based methods have shown promise in overcoming these obstacles by automatically learning the optimal transformations for enhancing underwater image quality. Despite their advancements, many of these methods focus on single-feature content extraction.The MFEF model is innovative due to incorporating multiple input pathways and an embedded fusion mechanism, utilising the power of both the White Balance (WB) and Contrast-Limited Adaptive Histogram Equalization (CLAHE) algorithms to extract and fuse diverse feature forms.

In image processing, white balance is a technique used to modify colours to make the image appear more natural. The primary objective of this operation is to rectify colour shifts in the image. This is especially important for images taken underwater, where the varied wavelengths of light absorbed by the water can distort colours, oftentimes suppressing values in the red channel. Histogram equalisation is an image processing in which the contrast of the image is enhanced. CLAHE is an advanced version of histogram equalisation in which the image is divided into small rectangles and this technique is applied section by section. It works especially well at highlighting features in images taken underwater when visibility is low.

Within the MFEF model architecture, both WB and CLAHE processed images are used as distinct input pathways along with the original image, the WB branch providing a “colour corrected” version of the input and the CLAHE branch providing a version of the image with enhanced details and higher contrast.

Each input is processed through separate pathways within the MFEF model, where the features are extracted using a Residual Enhancement Module (REM). The outputs from these REMs are then integrated using the Multi-Feature Fusion (MFF) module.

The Multi-Feature Fusion (MFF) module is a component designed to integrate and enhance features extracted from different paths of an image processing network. The outputs of these REM modules are combined, making sure that several feature representations are included in the finished feature map through addition or concatenation. The MFF module does this by closely connecting the feature encoding of several viewpoints by passing the features through a channel shuffle procedure.

The output is passed through a second REM module and then a series of convolutional layers before being passed through the Pixel-Weighted Channel Attention Module (PCAM).

By highlighting significant features and suppressing less valuable ones, PCAM is intended to recalibrate the channel-wise feature responses. Using a weighting matrix, this module dynamically modifies the contributions of the various channels according to the input image's content. To improve the clarity and quality of the reconstructed underwater pictures, PCAM aims to increase the representation of important elements while decreasing the impact of less useful ones. Convolutional operations are used in the implementation to adaptively modify channel weights in response to the channel-wise and spatial information found in the feature maps.

Then a third set of Residual Enhancement Modules (REMs) is applied to these feature maps and the improved features go through one last convolutional layer after the REMs. By combining all of the refined features into a single output image, this layer serves as an aggregator. A visually coherent underwater image is produced by aggregating all the refined features from the feature maps through convolutional operations.

Additionally, the paper proposes a novel loss function for training the model. The loss is a combination of two loss functions; an L1 loss as well as a loss called perceptual loss. To calculate the perceptual loss, a VGG-16 neural network pretrained on the ImageNet dataset is used. The prediction and ground truth image are passed through the network and the total distance between the outputs at each layer of the network are measured. The perceptual loss measures the difference between the high-level features of enhanced image and the target image, capturing discrepancies in texture, colour, and structural details that are often missed by traditional pixel-based loss functions.

5. Methodology

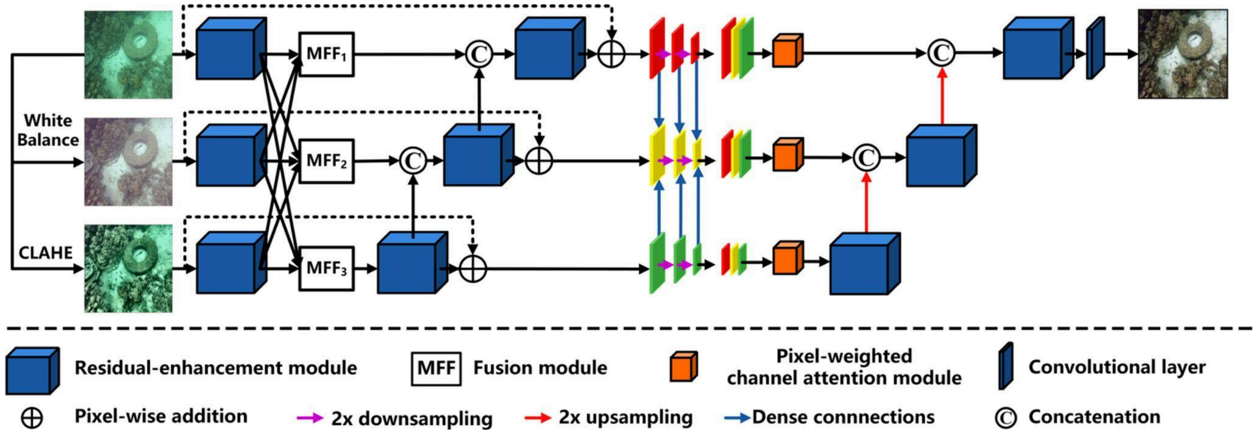


Fig.: Overview of the MFEF architecture

We tried to replicate the model proposed in the paper as closely as possible. There were however some areas of ambiguity, as well as some areas in which we were not able to replicate the model exactly as proposed because of various constraints.

For example, one such area of ambiguity in the paper is in the number of channels in the data change throughout the network. Specifically, after the second set of residual enhancement blocks, we were unsure as to how many channels the data should have. The output, we believe, has to be three channels given that we perform a pixel-wise addition of the input image and the output. However, we concatenate the data in the different branches before feeding it to the corresponding residual-enhancement block, indicating that the number of channels should increase to a number larger than three. To solve this, we concatenate the inputs of the three channels. However, we added a convolutional layer after each residual enhancement module to return the channel size to what is desired by the rest of the network (i.e. three channel data).

There was also ambiguity as to what the red, yellow, and green convolutional section of the network (see figure above) are. Moreover, there was no description of what the “dense connections” between these branches is meant to indicate. Our interpretation of this figure was that we perform three convolution operations, using the same kernels on the data at each branch, downsampling after each operation. We then concatenate the data from each branch of the same size. This newly formed data is what is passed onto the PCAM section of the network.

Finally, there was a blocker during the replication process that caused us to deviate from what was described in the paper. In the MFF module description, the authors call for a channel shuffle operation to be performed. We found however that the backward pass for this operation in PyTorch was not implemented. Finding no solution to this, we opted instead to omit this operation from the model.

6. Experimental Setup

We trained our model using a basic PyTorch training procedure. Every 150 images in our training set, we would perform an evaluation on the test set. We used a simple checkpoint method of saving the model weights corresponding to the model with the lowest observed loss overall on the test set. As proposed in the paper we trained the model over 200 epochs at a learning rate of 1e-4.

The study uses multiple assessment metrics to evaluate the performance of the MFEF model, guaranteeing statistical and perceptual quality of the enhanced images.

Peak Signal-to-Noise Ratio (PSNR): The average error between the input and output is measured using PSNR, which is based on error-sensitive image quality assessment. The more similar the output image is to the reference image, the higher the PSNR value.

Structural Similarity Index Metric (SSIM): Three components of an image are often assessed using SSIM: structure, contrast, and illumination. In terms of image structure, the higher the SSIM score, the higher the reliability and the closer the outcome is to the actual data.

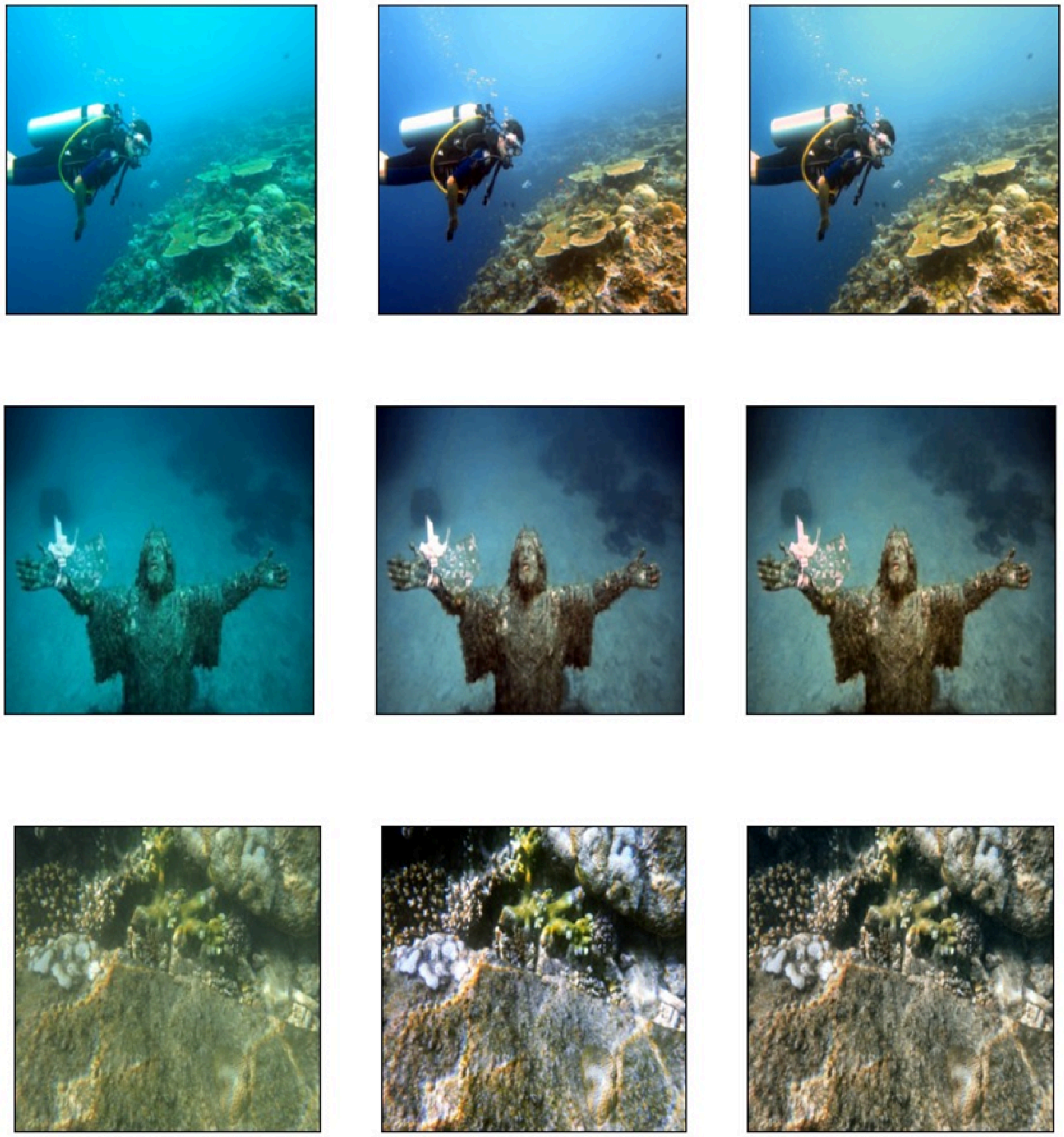
Patch-based Contrast Quality Index (PCQI): A large portion of the visual perception of image mass is occupied by PCQI, which is utilised to measure contrast in images. Three factors are taken into consideration by PCQI when evaluating image distortion: average intensity, signal intensity, and signal structure. A higher PCQI score corresponds to a stronger contrast and sharper image.

Underwater Colour Image Quality Evaluation (UCIQE) and Underwater Image Quality Measure (UIQM): The comprehensive indicators of the image's colour intensity, saturation level, and contrast are then measured using the UCIQE and UIQM scores as a reference. The better the image processing outcome, the higher the UCIQE and UIQM index scores. UCIQE is used to objectively assess whether an image is, for example, colour cast or blurry. Underwater image contrast measure (UIConM), underwater image sharpness measure (UISM), and underwater image colourfulness measure (UICM) are the three underwater image attribute measures that make up the UIQM.

7. Experimental Results

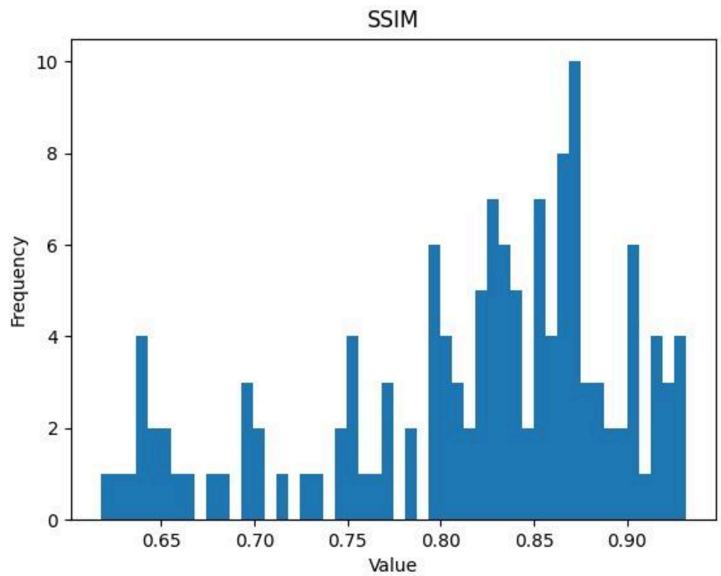
Our model converged at 89 epochs, with a total loss of 73996 on the test set. Training for 200 epochs took approximately 12 hours on a single Nvidia P100 16GB GPU.

Below, you will find some visualisations of how our model performed at enhancing the underwater images in the UIEB dataset. The input to the model is on the left, ground truth in the middle, and the prediction of our model is on the right.

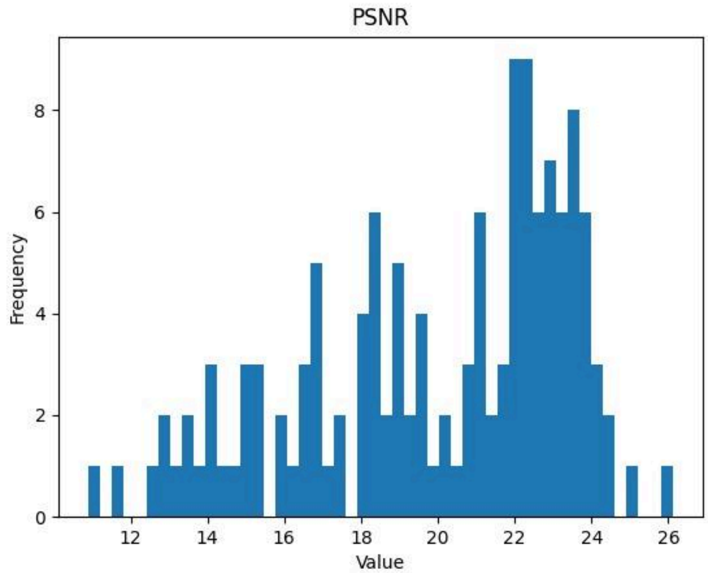


Evaluation Metrics

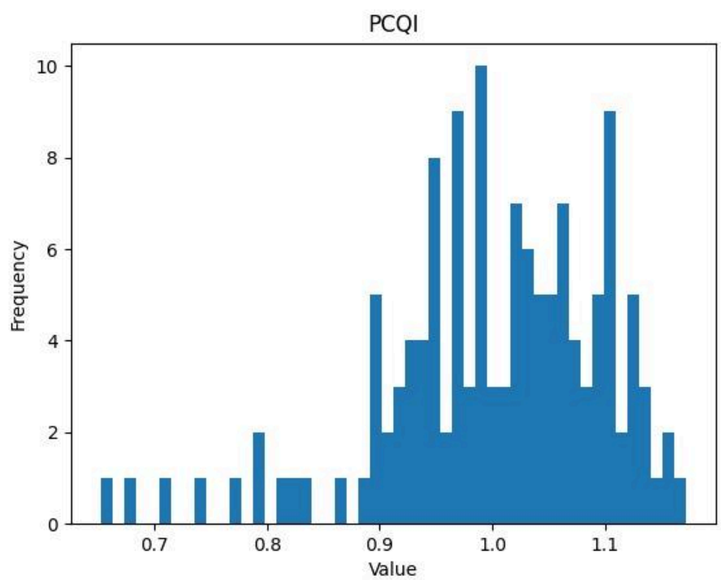
The following graphs show a series of histograms representing the distribution of values for the different image quality metrics.



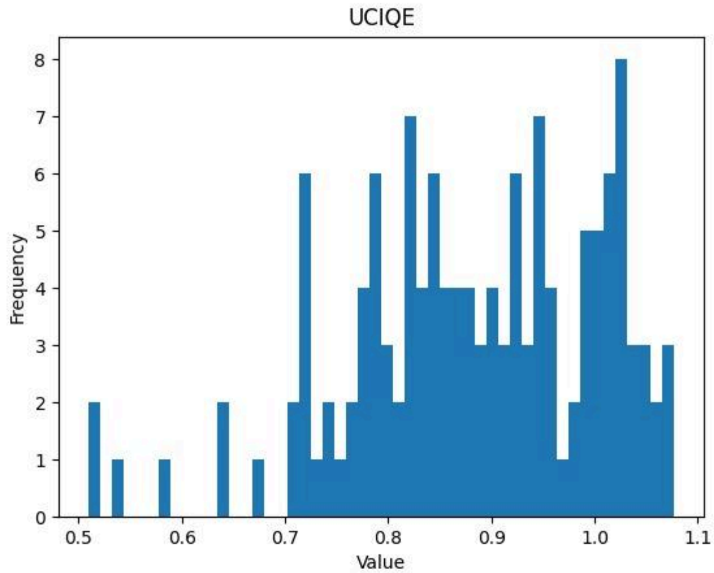
The **SSIM** histogram suggests a concentration of values around a particular range, with few outliers. This implies that the structural integrity of the images is mostly well-preserved after image enhancement, indicating the model's consistency in maintaining image structures across various samples.



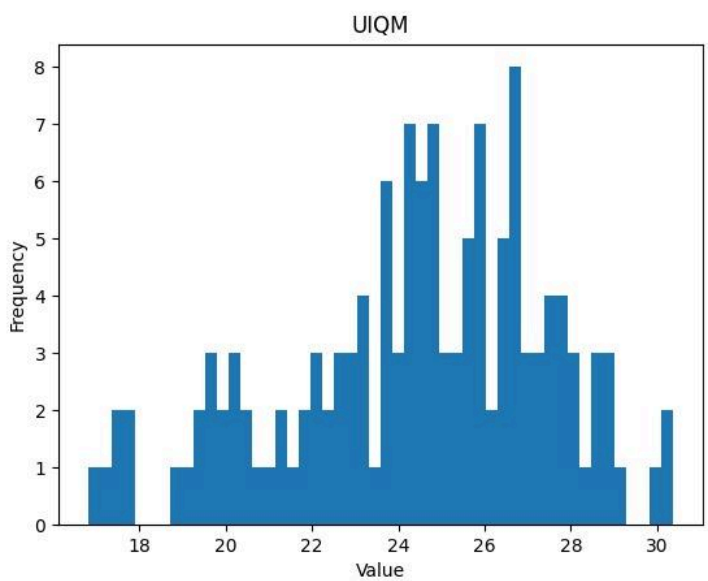
The **PSNR** histogram is more spread out, indicating variability in the noise reduction capabilities of the model. However, there are peaks at the higher values suggesting good performance for most of the images.



The **PCQI** histogram shows a skew towards higher values, indicating that the model consistently enhances the contrast of underwater images. This finding is particularly noteworthy considering the importance of contrast enhancement in underwater imagery.



The **UCIQE** histogram's distribution suggests a high performance in terms of colour correction as the graph is skewed to the right. The values are dispersed but not too widely, indicating a level of consistency in colour enhancement but with room for improvement.



The **UIQM** histogram exhibits a sharp peak at a high value, which denotes that the model achieves high-quality enhancements with regards to overall underwater image quality, considering factors like colourfulness, sharpness, and contrast.

Metric	Results from paper	Our results
mean PSNR	23.352	19.553
mean SSIM	0.910	0.811
mean PCQI	0.881	0.999
mean UCIQE	0.602	0.889
mean UIQM	1.333	24.340

Our PCQI and UIQM are significantly higher than the values seen in the paper. We are unsure as to why this behaviour is being exhibited. There could be a number of causes related to the changed image size, our implemented training procedure or an incorrect implementation of the metric in our project.

When selecting the model based on test loss, we monitored the test loss during training and chose the model with the lowest value. However, this approach may unintentionally favour the characteristics of the test data, potentially restricting the model's ability to generalise.

8. Future Work

Improved Model Generalisation: In the next iterations of our research, we will add a validation set consisting of entirely unseen data to make sure that our model is not overfitting to particular features of the data it has been trained on. This validation set, which offers an intermediate assessment apart from the final test set, will function as a reliable checkpoint during the training phase. This way, instead of having our model memorise the training set, we can keep track of it and make sure it is learning to generalise.

In addition, we intend to expand our assessment framework to incorporate an alternative dataset, enabling us to assess the MFEF model's generalisability across various categories of underwater images. This dataset will be chosen to reflect a wider range of underwater conditions and environments, which may include various water types, depths, and biomes.

Feature Selection Analysis: Our own comparison of the features of the input and enhanced images will be conducted. Experiments will be conducted by first observing how many features can automatically be selected from the original images and then comparing that result to the enhanced images. The model's performance can be evaluated by taking this difference into consideration.

9. Conclusion

In conclusion, our replication project aimed to reproduce the results of the “Multi-view underwater image enhancement method via embedded fusion mechanism” proposed by Jingchun Zhou et al. While our reproduced mean PSNR and SSIM values were lower than those documented for the MFEF model in the original paper, our results align with expectations for the more commonly used PSNR, SSIM, and UCIQE metrics. Qualitatively, we are confident in the success of our replication, as the images generated by our network visually closely resemble the ground truth.

The discrepancies in quantitative results could stem from variations in testing conditions, dataset characteristics, or differences in the model architecture between our replication and the original study. This underscores the importance of comprehensive explanations for newly proposed methods in academic papers. The practical utility of such approaches diminishes when they cannot be replicated due to ambiguities in their descriptions.

10. References

- Jingchun Zhou, Jiaming Sun, Weishi Zhang, Zifan Lin, Multi-view underwater image enhancement method via embedded fusion mechanism, Engineering Applications of Artificial Intelligence, Volume 121, 2023, 105946, ISSN 0952-1976, https://doi.org/10.1016/j.engappai.2023.105946.