# A NEW DATASET FOR BENCHMARKING UNDERWATER IMAGE ENHANCEMENT METHODOLOGIES

(SUALTI GÖRSELLERDE İYILEŞTIRME YÖNTEMLERI KALITE TESTI IÇIN BIR VERI SETI)

by

**ALP YÜCESOY, B.S.**

**Thesis**

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

**MASTER OF SCIENCE**

in

**COMPUTER ENGINEERING**

in the

**GRADUATE SCHOOL OF SCIENCE AND ENGINEERING**

of

**GALATASARAY UNIVERSITY**

**NOVEMBER 2025**

Approval of the thesis:

**A NEW DATASET FOR BENCHMARKING UNDERWATER IMAGE ENHANCEMENT METHODOLOGIES**

submitted by **ALP YÜCESOY** in partial fulfillment of the requirements for the degree of **Master of Science in Computer Engineering Department, Galatasaray University** by,

**Examining Committee Members:**

Supervisor, __________

__________

__________

**Date:** __________

## ACKNOWLEDGMENTS

Special thanks to Sultan Turhan, Murat Egi, Burak Arslan and Güneş Arık for their unending help and patience with me throughout this process.

November 2025
Alp YÜCESOY

# TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

# ABSTRACT

Underwater images serve as an invaluable resource for researchers across diverse scientific disciplines. These images provide critical insights into aquatic ecosystems, underwater archaeology, marine biology, geology, and more. However, capturing high-quality underwater images is inherently challenging due to the unique and often unpredictable environmental conditions. Factors such as light absorption, scattering, turbidity, and the presence of particles in the water can severely distort the colors and details of these images, making it difficult to accurately represent the underwater reality. These limitations pose significant challenges, particularly in tasks that require precision, such as underwater photogrammetry, where accurate three-dimensional reconstructions of submerged terrains or objects are essential.

Photogrammetry itself is a highly specialized methodology that involves capturing a large number of overlapping images of an object or location from various angles to construct a detailed and accurate three-dimensional model. When applied underwater, photogrammetry becomes even more critical, as it enables researchers to map, document, and analyze objects or environments that are otherwise difficult or impossible to access. Examples include shipwrecks, coral reefs, underwater caves, and submerged archaeological sites. The success of underwater photogrammetry relies heavily on the quality of the images, which must retain sufficient detail and color accuracy despite the challenges posed by the aquatic environment.

In recent years, significant research efforts have been dedicated to improving the quality of underwater images through advanced restoration and color correction techniques. These efforts aim to mitigate the effects of light distortion and loss of detail, thereby enhancing the utility of the images for scientific and practical applications. A wide array of approaches has been developed, ranging from traditional mathematical models that address specific aspects of image degradation to state-of-the-art artificial intelligence (AI) and computer vision-based methods. These modern techniques leverage machine learning algorithms and deep neural networks to adaptively correct distortions, enhance image clarity, and restore accurate colors, often surpassing the capabilities of conventional methods.

The integration of advanced image processing techniques with underwater research is paving the way for more precise and detailed studies, enabling scientists to better the underwater world with unprecedented accuracy and efficiency.
This study is presents an analysis of currently available underwater image datasets for image restoration and enhancement, a thorough analysis of the state-of-the art enhancement methodologies, as well a new dataset of underwater images taken specifically for photogrammetry purposes. This dataset was then used to train and compare currently available methodologies in a different context than what they are designed for, which is a generalist approach to all underwater environments as opposed to presenting a model that is specifically trained for certain types of environments.
This study provides a comprehensive analysis of the currently available underwater image datasets that are widely used for image restoration and enhancement tasks. It thoroughly examines the most advanced enhancement methodologies and their capabilities, identifying their strengths and limitations in addressing the challenges posed by underwater environments. Additionally, the study introduces a new dataset specifically designed for photogrammetry applications. This dataset consists of high-resolution underwater images, meticulously captured in controlled conditions to ensure suitability for three-dimensional reconstructions and precise mapping tasks.
The newly developed dataset serves as a unique resource for testing the performance of existing enhancement methodologies in a context distinct from their original design objectives. While many current approaches are generalized to accommodate a wide variety of underwater conditions, this study aims to explore their adaptability and effectiveness in a specialized context. By applying these generalist methodologies to the photogrammetry-specific dataset, the research evaluates their capacity to handle unique challenges such as maintaining structural integrity, color accuracy, and clarity in images intended for precise 3D modeling. The results offer valuable insights into the trade-offs between generalist and specialist models for underwater image processing, paving the way for future advancements tailored to specific use cases like photogrammetry.
This study also examines the effectiveness of domain-specific model design in contrast to generalized approaches, using the aforementioned dataset for training and evaluation. Our results demonstrate that a smaller, task-oriented model not only achieves a higher PSNR score but also offers significant advantages in terms of implementation simplicity and computational efficiency. These findings support the conclusion that lightweight, use-case-specific models are more suitable for underwater photogrammetry and image

enhancement tasks. By focusing on the specific requirements and constraints of underwater imaging, we show that it is possible to outperform larger, more generalized architectures while reducing both the cost and complexity of deployment.

**Keywords :** Image Enhancement, Image Processing, Machine Learning, Deep Learning, Underwater Imaging, Photogrammetry

# RÉSUMÉ

Les images sous-marines constituent une ressource inestimable pour les chercheurs dans divers domaines scientifiques. Elles offrent des informations cruciales sur les écosystèmes aquatiques, l'archéologie sous-marine, la biologie marine, la géologie, et bien d'autres domaines. Cependant, capturer des images sous-marines de haute qualité est un défi inhérent en raison des conditions environnementales uniques et souvent imprévisibles. Des facteurs tels que l'absorption de la lumière, la diffusion, la turbidité et la présence de particules dans l'eau peuvent considérablement altérer les couleurs et les détails des images, rendant difficile une représentation fidèle de la réalité sous-marine. Ces limitations posent des défis importants, notamment dans les tâches exigeant une grande précision, comme la photogrammétrie sous-marine, où des reconstructions tridimensionnelles précises de terrains ou d'objets immergés sont essentielles.

La photogrammétrie est une méthodologie hautement spécialisée qui consiste à capturer un grand nombre d'images superposées d'un objet ou d'un lieu sous différents angles afin de construire un modèle tridimensionnel détaillé et précis. Lorsqu'elle est appliquée sous l'eau, la photogrammétrie revêt une importance encore plus grande, car elle permet de cartographier, documenter et analyser des objets ou des environnements qui seraient autrement difficiles, voire impossibles, à atteindre. Cela inclut des exemples comme les épaves, les récifs coralliens, les grottes sous-marines ou les sites archéologiques immergés. Le succès de la photogrammétrie sous-marine dépend fortement de la qualité des images, qui doivent conserver suffisamment de détails et de précision des couleurs malgré les défis posés par l'environnement aquatique.

Ces dernières années, des efforts de recherche considérables ont été consacrés à l'amélioration de la qualité des images sous-marines grâce à des techniques avancées de restauration et de correction des couleurs. Ces efforts visent à atténuer les effets des distorsions lumineuses et de la perte de détails, améliorant ainsi l'utilité des images pour des applications scientifiques et pratiques. Une large gamme d'approches a été développée, allant des modèles mathématiques traditionnels traitant des aspects spécifiques de la dégradation des images aux méthodes modernes basées sur l'intelligence artificielle (IA) et la vision par ordinateur. Ces techniques modernes exploitent des

algorithmes d'apprentissage automatique et des réseaux de neurones profonds pour corriger de manière adaptative les distorsions, améliorer la clarté des images et restaurer des couleurs précises, souvent avec des performances supérieures à celles des méthodes conventionnelles.
L'intégration de techniques avancées de traitement d'images à la recherche sous-marine ouvre la voie à des études plus précises et détaillées, permettant aux scientifiques de mieux comprendre le monde sous-marin avec une précision et une efficacité sans précédent.
Cette étude propose une analyse approfondie des ensembles de données d'images sous-marines actuellement disponibles, largement utilisés pour les tâches de restauration et d'amélioration d'images. Elle examine en détail les méthodologies d'amélioration les plus avancées et leurs capacités, en identifiant leurs points forts et leurs limites face aux défis posés par les environnements sous-marins. De plus, l'étude présente un nouvel ensemble de données spécifiquement conçu pour des applications de photogrammétrie. Cet ensemble comprend des images sous-marines haute résolution, soigneusement capturées dans des conditions contrôlées pour garantir leur adéquation aux reconstructions tridimensionnelles et aux tâches de cartographie précise.
Le nouvel ensemble de données sert de ressource unique pour tester les performances des méthodologies d'amélioration existantes dans un contexte distinct de leurs objectifs de conception d'origine. Alors que de nombreuses approches actuelles sont généralisées pour s'adapter à une grande variété de conditions sous-marines, cette étude vise à explorer leur adaptabilité et leur efficacité dans un contexte spécialisé. En appliquant ces méthodologies généralistes à l'ensemble de données spécifique à la photogrammétrie, la recherche évalue leur capacité à relever des défis uniques, tels que le maintien de l'intégrité structurelle, de la précision des couleurs et de la clarté dans des images destinées à la modélisation 3D précise. Les résultats offrent des informations précieuses sur les compromis entre modèles généralistes et spécialisés pour le traitement des images sous-marines, ouvrant la voie à de futures avancées adaptées à des cas d'utilisation spécifiques comme la photogrammétrie.
Cette étude examine également l'efficacité de la conception de modèles spécifiques à un domaine par rapport aux approches généralisées, en utilisant le jeu de données mentionné pour l'entraînement et l'évaluation. Nos résultats démontrent qu'un modèle plus petit, orienté vers la tâche, obtient non seulement un score PSNR plus élevé, mais offre également des avantages significatifs en termes de simplicité de mise en œuvre et d'efficacité computationnelle. Ces résultats soutiennent la conclusion selon laquelle

des modèles légers et spécifiques à un cas d'utilisation sont mieux adaptés aux tâches de photogrammétrie sous-marine et d'amélioration d'image. En se concentrant sur les exigences et contraintes spécifiques de l'imagerie sous-marine, nous montrons qu'il est possible de surpasser des architectures plus larges et plus générales tout en réduisant à la fois le coût et la complexité du déploiement.

**Mots Clés :** Amélioration d'images, Traitement d'images, Apprentissage automatique, Apprentissage profond, Imagerie sous-marine, Photogrammétrie

## ÖZET

Sualtı görüntüleri, farklı bilimsel disiplinlerdeki araştırmacılar için son derece değerli bir kaynaktır. Bu görüntüler, su ekosistemleri, sualtı arkeolojisi, deniz biyolojisi, jeoloji ve daha birçok alanda kritik bilgiler sunar. Ancak, yüksek kaliteli sualtı görüntüleri elde etmek, çevresel koşulların benzersizliği ve genellikle öngörülemezliği nedeniyle doğal olarak zordur. Işık emilimi, saçılma, bulanıklık ve sudaki partiküllerin varlığı gibi faktörler, görüntülerin renklerini ve ayrıntılarını ciddi şekilde bozabilir ve sualtı gerçekliğini doğru bir şekilde temsil etmeyi zorlaştırır. Bu sınırlamalar, özellikle batık araziler veya nesneler gibi sualtı ortamlarının üç boyutlu yeniden yapılandırılmasının gerekli olduğu fotogrametri gibi hassasiyet gerektiren görevlerde önemli zorluklar yaratır.

Fotogrametri, bir nesnenin veya konumun farklı açılardan alınan birçok örtüşen görüntüsünü kullanarak ayrıntılı ve doğru bir üç boyutlu model oluşturmayı amaçlayan son derece uzmanlaşmış bir yöntemdir. Sualtı fotogrametrisi, erişilmesi zor nesnelerin veya bölgelerin haritalanması, belgelenmesi ve analiz edilmesi için sıklıkla kullanılmaktadır. Örnekler arasında batıklar, mercan resifleri, sualtı mağaraları ve su altına gömülü arkeolojik alanlar yer alır. Sualtı fotogrametrisinin başarısı büyük ölçüde görüntülerin kalitesine bağlıdır; bu görüntülerin, sualtı ortamının yarattığı zorluklara rağmen yeterli ayrıntı ve renk doğruluğunu koruması gerekir.

Son yıllarda, gelişmiş restorasyon ve renk düzeltme teknikleriyle sualtı görüntülerinin kalitesini artırmaya yönelik önemli araştırmalar yapılmıştır. Bu çalışmalar, ışık bozulmalarını ve ayrıntı kaybını azaltmayı, dolayısıyla görüntülerin bilimsel ve pratik uygulamalar için kullanılabilirliğini artırmayı amaçlamaktadır. Geleneksel matematiksel modellerden, yapay zeka (AI) ve bilgisayarla görme tabanlı modern yöntemlere kadar çok çeşitli yaklaşımlar geliştirilmiştir. Bu modern teknikler, bozulmaları adaptif bir şekilde düzeltmek, görüntü netliğini artırmak ve doğru renkleri geri kazandırmak için makine öğrenimi algoritmalarını ve derin sinir ağlarını kullanır ve genellikle geleneksel yöntemlerin ötesinde performans sergiler.

Gelişmiş görüntü işleme tekniklerinin sualtı araştırmalarıyla entegrasyonu, daha hassas ve detaylı çalışmalara olanak tanıyarak bilim insanlarının sualtı dünyasını benzersiz bir doğruluk ve verimlilikle daha iyi anlamalarını sağlamaktadır.

Bu çalışma, görüntü restorasyonu ve iyileştirme görevleri için yaygın olarak kullanılan mevcut sualtı görüntü veri kümelerinin kapsamlı bir analizini sunmaktadır. Çalışma, en gelişmiş iyileştirme yöntemlerini ve bunların yeteneklerini ayrıntılı bir şekilde inceleyerek, sualtı ortamlarının yarattığı zorluklara yanıt verme konusundaki güçlü ve zayıf yönlerini ortaya koymaktadır. Ayrıca, fotogrametri uygulamaları için özel olarak tasarlanmış yeni bir veri kümesi sunulmaktadır. Bu veri kümesi, üç boyutlu yeniden yapılandırma ve hassas haritalama görevleri için uygunluk sağlamak amacıyla kontrollü koşullarda titizlikle çekilmiş yüksek çözünürlüklü sualtı görüntülerinden oluşmaktadır. Geliştirilen bu yeni veri kümesi, mevcut iyileştirme yöntemlerinin performansını, orijinal tasarım hedeflerinden farklı bir bağlamda test etmek için benzersiz bir kaynak olarak hizmet etmektedir. Günümüzdeki birçok yaklaşım, çeşitli sualtı koşullarına uyum sağlamak için genelleştirilmiş olsa da, bu çalışma, bunların özel bir bağlamdaki uyarlanabilirliğini ve etkinliğini keşfetmeyi amaçlamaktadır. Bu genelleştirilmiş yöntemler, fotogrametriye özgü veri kümesine uygulandığında, yapısal bütünlük, renk doğruluğu ve netlik gibi 3D modelleme için önemli olan benzersiz zorluklarla başa çıkma yetenekleri değerlendirilmiştir. Sonuçlar, sualtı görüntü işleme için genelci ve uzmanlaşmış modeller arasındaki ödünleşimler hakkında değerli bilgiler sunarak, fotogrametri gibi özel kullanım durumlarına yönelik gelecekteki ilerlemelerin yolunu açmaktadır.
Bu çalışma, yukarıda belirtilen veri kümesini eğitim ve değerlendirme amacıyla kullanarak, alana özgü model tasarımının genelleştirilmiş yaklaşımlara kıyasla etkinliğini de incelemektedir. Sonuçlarımız, daha küçük ve göreve yönelik bir modelin yalnızca daha yüksek PSNR puanı elde etmekle kalmayıp, aynı zamanda uygulama basitliği ve hesaplama verimliliği açısından da önemli avantajlar sunduğunu göstermektedir. Bu bulgular, hafif ve kullanım amacına özel modellerin su altı fotogrametrisi ve görüntü iyileştirme görevleri için daha uygun olduğunu desteklemektedir. Su altı görüntülemenin özel gereksinimlerine ve kısıtlamalarına odaklanarak, daha büyük ve genelleştirilmiş mimarileri hem maliyet hem de uygulama karmaşıklığı açısından azaltarak geride bırakmanın mümkün olduğunu gösteriyoruz.

**Anahtar Kelimeler :** Görüntü İyileştirme, Görüntü İşleme, Makine Öğrenimi, Derin Öğrenme, Sualtı Görüntüleme, Fotogrametri

## LIST OF ABBREVIATIONS

| | |
|---|---|
| ResNet | Residual Network |

# 1 INTRODUCTION

Underwater images are a very significant asset for researchers in different fields. However, due to varied conditions and circumstances of the photos taken, achieving a high quality representation of reality is often difficult. This is especially crucial for accurate photogrammetry, as in 3-D reconstructions of underwater terrain.

Although image enhancement models have been researched thoroughly in a general context, research being done with regards to underwater image restoration and color correction is definitely not at a similar level. Applying the standard methodologies might not serve our task, since the context and environment of the ground truth are quite different. Thus we need separate research in this field.

A multitude of approaches and techniques exist, ranging from using straightforward mathematical approaches, to using artificial intelligence and computer vision based approaches.

Photogrammetry is a methodology where a large amount of images of an object or location are taken in order to construct a three-dimensional model. Underwater photogrammetry is often used to map hard to reach objects or locations.

In "Underwater Single Image Color Restoration Using Haze-Lines and a New Quantitative Dataset"(Berman et al., 2020) proposed a dataset called Stereo Quantitative Underwater Image Dataset (SQUID). In "An Underwater Image Enhancement Benchmark Dataset and Beyond"(Li, Chunle, Ren, Cong, Hou, Kwong and Tao, 2019) proposed a dataset and benchmark called the Underwater Image Enhancement Benchmark (UIEB). Also, in "Sea-Thru : A Method for Removing Water From Underwater Images"(Akkaynak and Treibitz, 2019) made available a dataset of underwater images and their enhanced versions, while also providing a mathematical model for these enhancements based on the distance of the objects taken in the photo.

We present an original dataset, formed of a number of underwater images of a particular location, a sunken ship, which have been used for 3-D reconstruction. These images were taken for use in underwater photogrammetry specifically.

We also present a small ResNet based architecture specifically designed fro the task in order to prove that reaching higher performance for specialised tasks is achievable with cheaper and smaller models

We also apply a number of underwater image enhancement methods to this dataset to start benchmarking, and see which methods in the future might be of use in the context of photogrammetry.

The rest of the paper is organized as follows : a section on related works, discussing the different methods for underwater photogrammetry and then underwater image restoration, followed by a section detailing our dataset and the presented model, ending with a methodology of the experiment.

# 2 LITERATURE REVIEW

**Underwater image enhancement (UIE)** has become an important area of research, fueled by the growing need for clear and reliable visuals in fields like marine exploration, underwater robotics, environmental monitoring, and archaeology. Taking good-quality images underwater is notoriously difficult—water absorbs and scatters light in ways that distort colors, reduce contrast, blur details, and introduce noise. These issues make it tough for traditional cameras and image-processing techniques to produce useful results.

Over time, researchers have developed many different strategies to tackle these problems. Early methods were mostly based on classic image processing techniques like histogram equalization and white balancing. While sometimes effective, these approaches often struggled to adapt to the wide variety of underwater conditions.

The game changed with the rise of machine learning and especially deep learning. These newer, data-driven methods have made it possible to train models that can learn how to transform poor-quality images into clearer, more accurate ones. Modern techniques using convolutional neural networks (CNNs), generative adversarial networks (GANs), transformers, and hybrid models have led to impressive improvements in image clarity and color accuracy.

In the following section, we explore the most influential and up-to-date methods for underwater image enhancement. We organize them based on the type of model or learning strategy they use and take a closer look at the datasets, evaluation methods, and specific application needs that shape how these models are built and how well they perform.

## 2.1 Underwater Photogrammetry

Underwater photogrammetry provides an efficient nondestructive means for measurement in environments with limited accessibility. With the growing use of consumer cameras, its application is becoming easier, thus benefiting a wide variety of disciplines. Using cameras for underwater photogrammetry presents significant modeling challenges due to the refraction effect and the combination of the camera and its protective housing into a single unit.

In the work by (Telem and Filin, 2010), they examine the impact of the underwater environment on the photogrammetric process and introduces a model to describe the geometric distortions and estimate the additional parameters involved. The proposed model addresses the multimedia effect and inaccuracies associated with the setup of the camera and housing device.

The paper demonstrates that only a few additional parameters are necessary to model both elements while maintaining the collinearity relation. The findings reveal that a specific setup is not required for estimating these parameters and that the estimation process is robust against noise and initial approximations. Their experiments indicate that high accuracy levels can be achieved with the proposed methods, marking a significant advancement in underwater measurement techniques

In their article (Rende et al., 2022), use a photogrammetry-based micro-bathymetry approach to monitor Posidonia oceanica restoration efforts, incorporating soundings from high-resolution multibeam bathymetry. It emphasizes the integration of underwater photogrammetry with high-resolution multibeam bathymetry data, enhancing the accuracy of georeferencing point clouds and orthomosaics. This integration leads to improved estimations of surface, height, and volume of seagrass canopies, which are crucial for evaluating restoration success.

The study also discusses the potential of using visual Simultaneous Localization and Mapping (vSLAM) techniques for real-time 3D measurements, moving beyond traditional post-processing methods, while highlighting the use of Object-Based Image Analysis (OBIA) for classifying benthic habitats, which incorporates various data types to improve thematic mapping. It finally provides a structured workflow for underwater photogrammetric surveys, which can be applied to other marine studies, including geological and environmental monitoring

In (Urbina-Barreto et al., 2022) paper, the authors integrated underwater photogrammetry with fish surveys to investigate how reefscape characteristics influence the functional structure of reef fish assemblages across three environmentally diverse Indo-Pacific islands : Europa Island, Reunion Island, and New Caledonia. The paper utilizes underwater photogrammetry through Structure from Motion (SfM) to quantify habitat descriptors, enhancing the assessment of coral reef traits beyond traditional methods. The research highlights the efficiency of underwater photogrammetry in generating quantitative habitat data, which can be standardized and free from observer bias, thus improving monitoring programs for coral reefs.

## 2.2 Underwater Image Datasets

Underwater image enhancement (UIE) has become a critical research area due to its wide range of applications in marine biology, underwater exploration, and autonomous underwater vehicles. Enhancing underwater images is particularly challenging because of issues like color distortion, low contrast, and reduced visibility caused by the absorption and scattering of light in water. To address these challenges, researchers have developed various datasets and models aimed at training and evaluating enhancement algorithms. The availability of high-quality datasets is essential for advancing this field, as they provide the necessary ground truth for supervised learning and allow consistent benchmarking of different techniques. The following sections review several significant datasets that have been developed to support research in underwater image and video enhancement.

(Peng et al., 2023) introduce the Large-Scale Underwater Image (LSUI) dataset along with a novel U-shape transformer-based model designed specifically for underwater image enhancement. The LSUI dataset comprises a total of 5004 paired images. Each pair includes a degraded underwater image and its corresponding high-fidelity reference image. These reference images are crucial for both the training and evaluation of enhancement algorithms, providing a clear target for supervised learning. The dataset is meticulously curated to represent a broad range of underwater scenes, thereby ensuring its utility in developing models that are robust and capable of generalizing to various environmental conditions. One of the standout features of the LSUI dataset is its consideration of the uneven attenuation of different color channels in underwater environments. This attention to detail enhances the dataset's ability to support more accurate modeling of underwater light propagation and color correction, which are key aspects of effective underwater image enhancement.

In (Li, Guo, Ren, Cong, Hou, Kwong and Tao, 2019) present the Underwater Image Enhancement Benchmark (UIEB), a comprehensive dataset composed of 950 real-world underwater images. Among these, 890 images are accompanied by high-quality reference images, enabling objective evaluation of enhancement algorithms. The real-world nature of the UIEB makes it particularly valuable for practical applications, as it reflects the complexities and variability encountered in actual underwater scenarios. This realism distinguishes it from synthetic datasets and provides a more relevant foundation for assessing algorithm performance in real-life conditions, such as those found in marine engineering, underwater robotics, and environmental monitoring. The diversity and authenticity of the UIEB dataset make it a cornerstone resource for UIE research. (Kaneko et al., 2024) propose the Physics-Inspired Synthesized Underwater Image Dataset (PHISWID), which brings a unique approach to underwater image synthesis. This dataset includes paired atmospheric (ground-truth) images and their synthetically degraded underwater counterparts. Such paired data are extremely valuable for supervised learning, especially since acquiring real-world underwater-reference pairs is often infeasible due to environmental constraints. What sets PHISWID apart is its inclusion of marine snow—suspended particles such as organic debris and sand—that drastically reduce image clarity underwater. By incorporating these visual artifacts, PHISWID offers a more realistic and challenging dataset for enhancing image processing models. This dataset is instrumental in pushing the boundaries of underwater image enhancement, especially in cases where clarity is compromised by particulate matter.

(Du et al., 2024) present the Synthetic Underwater Video Enhancement (SUVE) dataset, which is specifically tailored for underwater video enhancement tasks. The SUVE dataset consists of 840 videos that simulate a wide array of underwater conditions. Each video is paired with a ground-truth reference video, making it highly suitable for supervised deep learning models. The dataset's video-centric nature is particularly important given the growing demand for real-time enhancement in underwater navigation, surveillance, and documentation. By providing consistent, frame-aligned ground-truth data, SUVE enables researchers to develop and benchmark algorithms that can process and improve video quality over time, accounting for dynamic elements such as motion blur, lighting fluctuations, and water current effects.

## 2.3 Underwater Image Enhancement

### 2.3.1 AI Methods

The field of underwater image enhancement (UIE) has witnessed rapid progress in recent years, particularly with the advent of deep learning and generative models. These techniques have enabled researchers to overcome long-standing challenges posed by underwater environments, such as color distortion, blurring, low contrast, and the presence of noise. In this context, several state-of-the-art models and novel algorithms have been proposed, many of which leverage new architectures, attention mechanisms, and domain-specific datasets to achieve superior enhancement results. The following section provides an overview of significant contributions in underwater image enhancement, highlighting the diverse range of approaches—from conditional generative adversarial networks to diffusion transformers—that have advanced the field.

(Islam et al., 2020) present a real-time underwater image enhancement model based on a conditional Generative Adversarial Network (cGAN). To guide the adversarial training process, the authors formulate a sophisticated objective function that evaluates perceptual image quality using a combination of global content, color consistency, local texture, and style attributes. In addition to the model, the authors introduce the EUVP dataset, a large-scale collection of both paired and unpaired underwater images. These images, categorized as either 'poor' or 'good' quality, were captured using seven different cameras under varying visibility conditions during real-world oceanic expeditions and collaborative human-robot experiments. One of the key contributions of this work is the demonstration that enhanced images significantly improve the performance of downstream computer vision tasks such as underwater object detection, human pose estimation, and saliency prediction. These results highlight the potential of the model for use in real-time preprocessing within the autonomy pipelines of visually guided underwater robots.

(Yadav et al., 2021) explore the application of deep learning techniques, specifically Convolutional Neural Networks (CNNs), for enhancing the visual quality of underwater images. Their approach focuses on removing common underwater image distortions, including color loss, blurriness, and noise. By leveraging the feature extraction capabilities of CNNs, the method effectively restores image clarity and vibrancy, laying a foundation for practical improvements in underwater image processing.

(Liu et al., 2024) propose an innovative underwater image enhancement algorithm that utilizes a parallel fusion architecture combining Transformer and CNN components. This hybrid approach is designed to simultaneously address both local and global feature extraction challenges. A unique transformer model is used to capture fine-grained local features with the aid of peak signal-to-noise ratio (PSNR) attention and linear operations, while the CNN module incorporates both temporal and frequency domain features to capture broader contextual information. The method effectively integrates high and low-frequency information to achieve comprehensive image enhancement. Qualitative and quantitative results show that the proposed method outperforms existing state-of-the-art approaches, achieving higher PSNR and Structural Similarity Index Measure (SSIM) scores. Additionally, the algorithm boosts detection performance by over 10%, making it highly effective for downstream applications.

(Zhao et al., 2024) introduce MSFE-UIENet, a novel neural network architecture tailored for underwater image enhancement. The model is specifically designed to mitigate common underwater imaging issues such as noise, blurring, low contrast, and severe color distortion. At the core of the network is a multi-stage feature extraction module that enhances the richness and detail of the captured features. MSFE-UIENet includes components such as an encoder-decoder structure and pyramid downsampling modules to progressively refine image quality. Experimental validation using various underwater image datasets confirms the network's robustness and effectiveness across diverse underwater conditions.

(Nie et al., 2024) propose an Image-Conditional Diffusion Transformer (ICDT), which introduces a novel approach to enhancing degraded underwater images. This model combines the power of diffusion-based generative modeling with transformer architectures, leveraging a hybrid loss function that incorporates variance to improve the log-likelihood and accelerate the sampling process. The approach yields impressive results, particularly on the challenging Underwater ImageNet dataset, where the model demonstrates substantial improvements in visual quality and enhancement efficiency.

(Zhang et al., 2024) present an advanced underwater image enhancement algorithm based on adversarial learning. Their approach emphasizes the correction of color casting—a common problem in underwater photography that impairs feature recognition and style transfer tasks. By incorporating adversarial objectives into the training process, the model is able to significantly improve the visual realism and fidelity of enhanced images. Evaluations show strong performance in both subjective (human evaluation) and objective (quantitative metrics) assessments, establishing the model as a practical solution for real-world underwater imaging applications.

(Li et al., 2021) propose a novel encoder-decoder architecture designed to enhance feature representation by integrating multiple color spaces within a unified framework. This method addresses the limitations of traditional RGB-only models by increasing the diversity of features extracted from underwater images, thereby improving color correction and contrast enhancement. An attention mechanism is incorporated to highlight the most salient features across different color spaces, ensuring that the network focuses on the most informative regions. The decoder component is guided by medium transmission values, which reflect the degree to which light from the scene reaches the camera sensor. This design allows the model to prioritize enhancement in areas most affected by degradation, resulting in significant improvements in overall image quality.

### 2.3.2 Mathematical Methods

(Shu et al., 2017) presents a novel method for improving the quality of underwater images using an extended multi-scale Retinex approach. This technique effectively addresses the challenges posed by underwater environments, such as color distortion and low visibility. The authors introduce an extended version of the multi-scale Retinex algorithm, which combines multiple scales to better handle varying lighting conditions and enhance color fidelity in underwater images. This approach allows for more accurate color restoration compared to traditional methods. The paper includes a comprehensive evaluation of the proposed method against existing techniques. The results demonstrate significant improvements in image clarity and color accuracy, showcasing the effectiveness of the extended multi-scale Retinex in real-world underwater scenarios.

In their work, (Muniraj and Dhandapani, 2023) introduce a novel approach to color correction that operates on individual LAB color space components. This method enhances the luminance (L) component through a series of steps, including Contrast Limited Adaptive Histogram Equalization (CLAHE), sharpening, gamma correction, and further color adjustments for the chrominance (A and B) components. This comprehensive color correction process significantly improves the overall color fidelity of underwater images. The proposed method also employs an adaptive LUT based on a probability threshold to enhance image contrast. This technique effectively increases the contrast by managing pixel intensity values across each RGB channel. Additionally, the probability threshold helps mitigate halo effects and artifacts, which are common issues in image enhancement processes. The paper incorporates a fast local Laplacian filter (FLLF) to enhance the underwater images while preserving edge details. This filter not only smooths textures but also ensures that important features in the images remain sharp and clear, which is crucial for underwater photography where detail is often lost.

In (Lei et al., 2022), the authors propose a new method for enhancing underwater images that addresses common issues such as color cast, poor contrast, and detail loss. This method combines color correction with dual image multi-scale fusion, which is a unique approach in this domain. The first step in the proposed method involves a color correction technique that effectively mitigates color cast. This is achieved by compensating the two color channels with the channel that has the highest mean value, followed by dynamic stretching of all color channels. This step is crucial for restoring the natural colors of underwater images. The paper introduces a complementary dual image multi-scale fusion method. This involves using pairs of images that have undergone adaptive gamma correction with weighted distribution. These images serve as inputs for the multi-scale fusion process, which enhances the overall image quality by improving contrast and detail. The authors emphasize the importance of selecting appropriate weight maps during the fusion process. This selection is critical for achieving optimal results in the enhancement of underwater images, ensuring that the most relevant features are highlighted. The proposed method includes a multi-scale detail-sharpening technique that enhances the details in the images. This step is essential for improving the visibility of fine details that are often lost in underwater photography due to scattering and absorption of light.

(Akkaynak and Treibitz, 2019) challenge the traditional atmospheric image formation model used for underwater images. They propose a new model that accurately accounts for the non-uniformity of the attenuation coefficient, which varies based on object range and reflectance. This is a crucial advancement as it addresses a fundamental flaw in previous methodologies. The authors introduce the Sea-Thru method, which utilizes RGBD images to recover lost colors in underwater images. This method first calculates back-scatter using the darkest pixels in the image, leveraging their known range information. This innovative approach enhances the recovery process by providing a more reliable estimation of the underwater environment. The method incorporates an estimate of the spatially varying illuminant to determine the range-dependent attenuation coefficient. This step is crucial for accurately modeling the light conditions in underwater settings, leading to better color recovery.

In (Levy et al., 2023), the authors develop a new rendering model for neural radiance fields in scattering media, which is based on the Sea-Thru image formation model, and suggest a suitable architecture for learning both scene information and medium parameters. The architecture enables the simultaneous learning of both scene information and the parameters of the scattering medium. This dual learning approach is crucial for accurately rendering scenes where visibility is significantly affected by the medium. The authors then showcase the effectiveness of their method through both simulated and real-world scenes. They successfully render novel photo-realistic views underwater, demonstrating the capability of their model to produce high-quality images even in challenging conditions

(Sethuraman et al., 2023) introduces *WaterNeRF*, another method based on neural radiance fields. WaterNeRF incorporates a physics-based model for underwater image formation. By estimating parameters from this model, the method enables effective image restoration, allowing for the synthesis of novel views from both degraded and corrected underwater images. This is crucial for improving the quality of underwater imagery. The method not only focuses on image restoration but also facilitates dense depth estimation. This capability is essential for various applications, including 3D reconstruction and depth perception in underwater environments, which are often challenging due to the varying conditions of the water. The authors conduct both qualitative and quantitative evaluations of WaterNeRF on a real underwater dataset. This thorough evaluation demonstrates the effectiveness of their approach in real-world scenarios, providing evidence of its practical applicability.

For (Zhang et al., 2021), the authors propose a simple yet effective color correction technique that utilizes sub-interval linear transformation. This method specifically addresses the common issue of color distortion in underwater images, which is caused by the scattering and absorption of light in water. This innovative approach helps restore the natural colors of underwater scenes. The paper introduces a Bi-interval histogram-based method for contrast enhancement. This technique employs an optimal equalization threshold strategy combined with an S-shaped function to enhance image contrast and highlight details. This dual approach allows for better visibility and detail in underwater images, which are often compromised due to poor lighting conditions. The authors apply a Gaussian low-pass filter to the L channel of the image to separate low- and high-frequency components. This decomposition is crucial for enhancing different aspects of the image, allowing for targeted improvements in both fine details and overall visibility. Inspired by multi-scale fusion methods, the paper employs a simple linear fusion technique to integrate the enhanced high- and low-frequency components. This integration ensures that the final output image retains both clarity and detail, resulting in a high-quality underwater image.

# 3 THE STUDY

## 3.1 Dataset

Most existing underwater image datasets are intentionally diverse, designed to train models that work across a wide range of scenarios. This is the case for two of the most famous underwater image datasets, such as LSUI in (Peng et al., 2023), which has a comparable number of images intentionally kept varied ; or UIEB in (Li, Chunle, Ren, Cong, Hou, Kwong and Tao, 2019), which has around 950 images, again, collected to be as diverse as possible. However, this diversity makes it difficult to achieve consistent results, as it complicates performance comparisons between different methods. These inconsistencies hinder the reliable evaluation of algorithms consistently.
Our dataset addresses this issue by offering a more controlled and stable set of images that enables accurate performance assessments while also incorporating enough variation to prevent overfitting, ensuring that the dataset remains applicable to real-world conditions.

### 3.1.1 Collected Images

This dataset is specifically designed for underwater photogrammetry and consists of a large collection of 3841 high-resolution images taken in submerged environments.
Photogrammetry, a well-established method, is commonly used to create detailed three-dimensional models by capturing numerous images of an object or scene. The quality of photogrammetric reconstruction depends on the following certain technical guidelines, such as maintaining a steady distance between the camera and the object and ensuring enough overlap between consecutive images. Typically, at least 70% overlap is needed to ensure a reliable reconstruction.
As the name suggests, underwater photogrammetry adapts this technique to underwater environments, where data collection is much more challenging. Unlike on land, divers have limited mobility, and the quality of images often suffers due to poor water clarity and changing lighting conditions. These difficulties highlight the need for spe-

cialized datasets that account for the limitations of underwater environments.
A key advantage of our dataset lies in the superior quality of the images compared to previously published datasets. Each image is captured at a high resolution of 3264 x 2448 pixels in RGB format, ensuring finer detail and clearer visual information. Unlike other datasets where image quality is often compromised by poor resolution or inconsistent capture conditions, 640x360 for LSUI and completely random for UIEB, the images in this collection maintain a high level of clarity, even under the challenging conditions of underwater environments. This improvement in quality makes the dataset particularly valuable for advanced photogrammetric applications, such as 3D modeling, where high-resolution data is essential for precision.
This dataset provides researchers with a reliable resource for evaluating the performance of different photogrammetry techniques, particularly in underwater environments where controlled data acquisition is difficult. The combination of controlled and variable elements ensures that the dataset can be used to test algorithms in a way that balances rigor with real-world generalization, fostering the development of more flexible methodologies that can be applied to underwater imaging and beyond.

#### 3.1.1.1 Collection Process

The dataset consists of 3841 images of a submerged ship, taken by a professional diver who maintained a consistent distance from the object and followed a set acquisition rate. Although efforts were made to standardize the capture process, some images were intentionally taken under non-ideal conditions to make the dataset more realistic, reflecting the challenges of actual underwater environments. The high resolution and structured nature of the dataset make it suitable for a wide range of photogrammetric applications, from basic image processing to advanced 3D reconstruction studies.
The end result was collected under relatively ideal conditions for photography. The visibility of the water was clean and the lighting levels are adequate. The main goal of the dataset would be to use in training color correction models.

### 3.1.2 Ground Truth

In most underwater imaging datasets, generating accurate ground truth is particularly difficult because there is no straightforward way to recover the original, undistorted scene. Many datasets attempt to approximate ground truth using human annotators or subjective enhancement techniques, which can introduce bias or inconsistencies. In contrast, our dataset is uniquely positioned to enable a more objective and reproducible approach to ground truth generation.

This advantage arises from the fact that the image acquisition process in our dataset was conducted under carefully controlled conditions. Specifically, the diver who captured the images maintained a known and consistent distance from the object in most frames. This distance information allows us to employ physics-based models of underwater light propagation to perform mathematically grounded color correction and image restoration.

Using these models, we can reverse the known effects of light attenuation and scattering based on the water type, camera parameters, and depth at which each image was taken. These corrections can yield images that closely approximate the true colors and contrasts of the submerged object as if it were imaged in air. In this way, we are able to construct a synthetic but physically meaningful set of ground truth images that serve as reliable targets for supervised learning tasks.

In order to color-correct our dataset, we applied the previously discussed Seathru method from (Akkaynak and Treibitz, 2019). This methodology requires high-quality images and their depth maps. Since the official Sea-thru paper has no released source code, we have found other implementations that also calculate depth maps with given images.

#### 3.1.2.1 Preprocess

To prepare our dataset for use in supervised learning tasks and to generate ground truth approximations, we applied a well-established underwater image correction technique known as Sea-thru, originally proposed in (Akkaynak and Treibitz, 2019). This method was selected due to its strong theoretical foundation and demonstrated success in restoring natural color and contrast in underwater scenes. The Sea-thru methodology leverages depth information along with the original image content to perform physically grounded color correction. Specifically, it models the light attenuation process

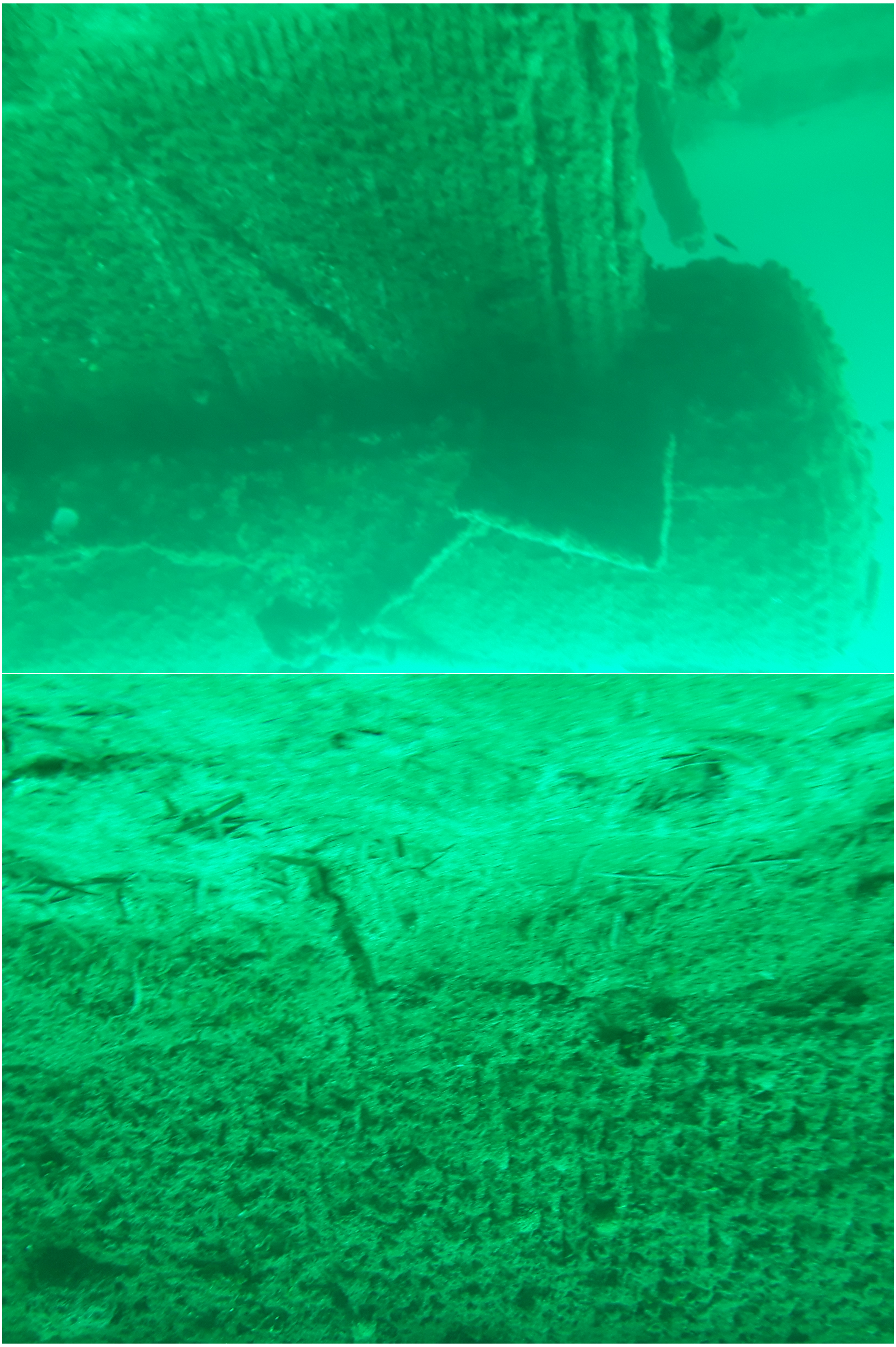


FIGURE 3.1 – Examples from the dataset

that occurs as light travels through water and uses this model to reconstruct the scene as it would appear under ideal, non-underwater conditions.

However, implementing the original Sea-thru pipeline presents certain challenges. Notably, the authors of the original paper did not release an official implementation or source code. As a result, we explored and evaluated alternative open-source implementations that aim to replicate the functionality of the Sea-thru method. These community-developed versions often incorporate additional tools or rely on separate depth estimation modules, and while not identical to the original, they capture the core principles of the technique.

Fortunately, our dataset includes images captured at relatively fixed distances by a trained diver, providing us with a stable foundation for estimating depth. This consistency is a significant advantage, as depth estimation is one of the key inputs for the Sea-thru algorithm. Even though we do not have direct sensor-based depth maps, the regularity and known conditions of the data acquisition process allow us to generate reliable proxy depth information.

To estimate depth maps for each image, we employed a pre-trained model from the MiDaS project, introduced in (Ranftl et al., 2020). MiDaS is a robust and widely used framework for monocular depth estimation, which has demonstrated strong generalization capabilities across diverse imaging domains. The model predicts relative depth maps from single RGB images without requiring stereo pairs or LiDAR inputs, making it well-suited for our underwater dataset where only monocular images are available.

By integrating MiDaS-generated depth maps with the Sea-thru correction model, we were able to reconstruct enhanced versions of the original underwater images. These corrected images approximate how the scenes would appear in air, thereby serving as effective pseudo-ground truth references. This preprocessing step is essential for training supervised models to learn the mapping from degraded to enhanced imagery in a consistent and physically interpretable manner.

#### 3.1.2.2 Color Correction

Sea-thru uses the previously calculated depth maps in order to color correct each pixel accordingly, since the distance between the object and the camera affects the attenuation of colors. Unlike traditional methods that often rely on atmospheric light models, Sea-thru employs a revised image formation model tailored specifically for

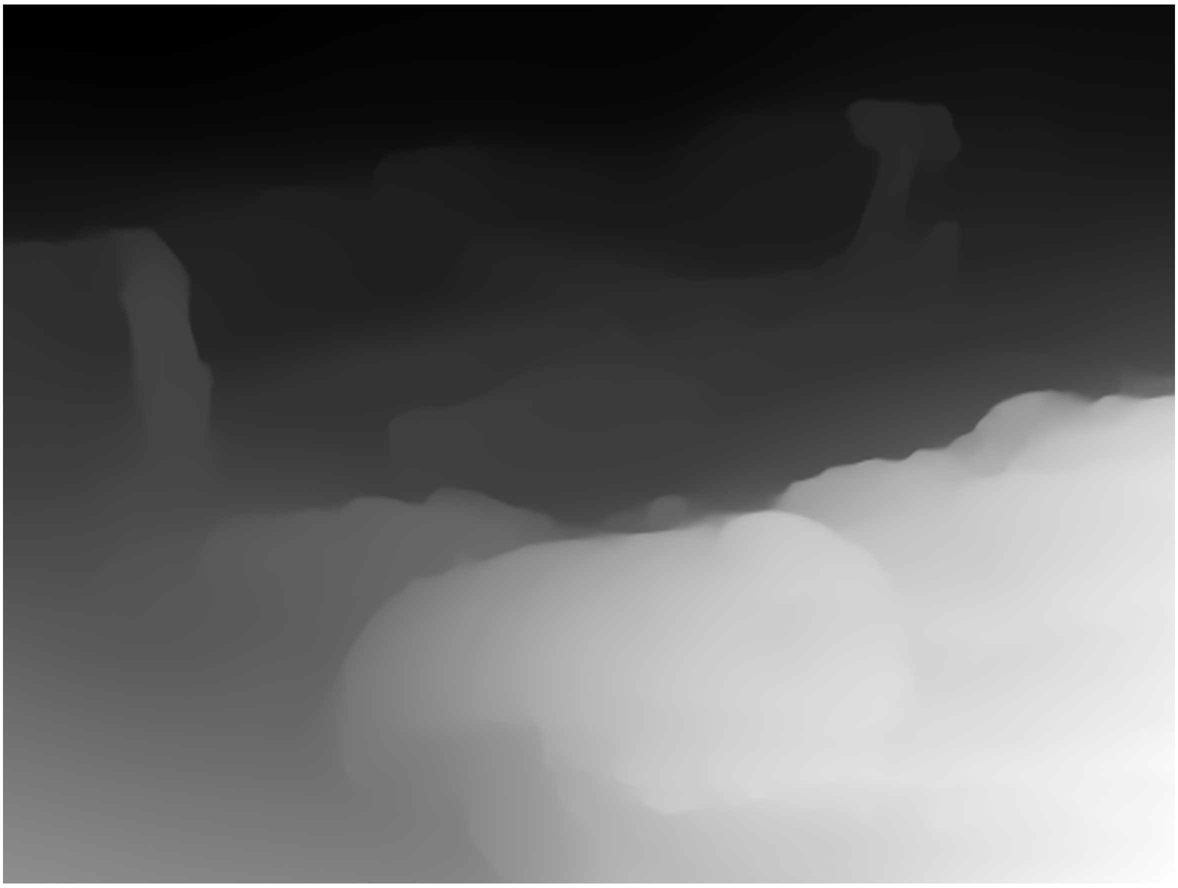


Figure 3.2 – Depth Map of an image from the dataset


FIGURE 3.3 – Color correction using Sea-thru

underwater environments. This model accounts for how light is absorbed and scattered in water, enabling the algorithm to reconstruct images with their true colors, as if taken in air. By understanding the depth information, the algorithm can effectively reverse the effects of light attenuation and backscatter, restoring the image's original colors.

## 3.2 Existing Image Processing and Deep Learning Methods

Image enhancement techniques are essential tools used to improve the overall quality and visual clarity of images. These methods serve a variety of purposes, particularly in making images more suitable for subsequent analysis, interpretation, or advanced processing. In many fields, such as medical imaging, environmental monitoring, or underwater exploration, improving image clarity can significantly enhance the ability to extract meaningful information. Broadly speaking, image enhancement methodologies can be classified into two main categories : those grounded in traditional mathematical models and those that rely on more contemporary artificial intelligence (AI)-based approaches. Mathematical models typically involve the application of algorithms that manipulate pixel values according to pre-defined formulas, such as contrast adjustment, histogram equalization, or filtering techniques. In contrast, AI-driven methods, especially those utilizing deep learning frameworks, adaptively learn from vast datasets to enhance images in ways that may not be easily captured by conventional mathematical techniques.

In our research, we have made the deliberate decision to use and compare four distinct methods in order to better understand which algorithms or models are most suitable for the specific needs of underwater image enhancement. The rationale behind this selection process is multi-faceted. First and foremost, the chosen models have demonstrated strong performance in the domain of underwater image enhancement. This prior success in similar contexts provides a solid foundation for their inclusion. Secondly, the availability of public codes and model weights for these methods ensures that they can be readily implemented and verified within our study, allowing for greater transparency and reproducibility of results. This is a crucial consideration in academic research, as it facilitates comparison and replication by other scholars. Finally, another significant factor in our selection was the diversity of the models themselves. Although all four models employ deep learning techniques, they are based on fundamentally different architectures and strategies. By comparing these dissimilar approaches, we aim to gain a more comprehensive understanding of the strengths and limitations of each, thereby identifying the model or models that are best aligned with the unique requirements of our underwater image enhancement tasks.

### 3.2.1 U-shape Transformer

In (Peng et al., 2023), we are presented with a novel U-shape Transformer network specifically for underwater image enhancement. This is the first time a transformer model has been applied to the UIE task. The transformer architecture, first introduced in (Vaswani et al., 2017), are a key part of LLMs with the "self-attention" mechanic that considers each element in a sequence with regards to the next element, meaning they are great at capturing context. They are currently used in all areas of AI. In the context of computer vision, "attention" considers all of the surrounding pixels of a pixel when evaluating it's information, thus informing a general context.

The U-shape architectures are widely used in Neural Networks where the input and output are expected to be of the same size and there is need for pixel-level precision (like segmentation or denoising tasks). Large inputs are diminished using convolutional and max-pooling layers to capture high-level features and compress the information. The reverse is then applied using up-sampling and transposed convolution layers resulting in an image with precise high-level features. The network architecture includes two specialized modules based on the transformer :

- Channel-wise Multi-scale Feature Fusion Transformer (CMSFFT) : This module focuses on enhancing the attention of the network towards different color channels, addressing the inconsistent attenuation of colors in underwater images.
- Spatial-wise Global Feature Modeling Transformer (SGFMT) : This module is

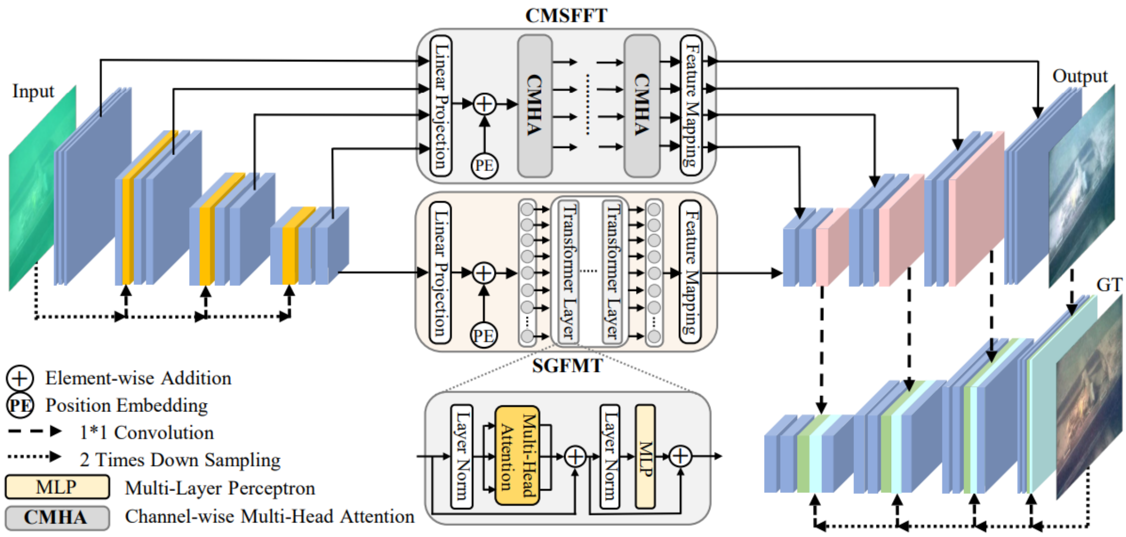



FIGURE 3.4 – Architecture of the U-shape Transformer (Peng et al., 2023)

designed to model global information and enhance the network's focus on severely degraded areas of the image.

These modules are implemented in a general U-shaped architecture, the CMSFFT receiving context information from each convolution and the SGFMT receiving the final convoluted layer. The SGFMT output is then up-sampled using the output of the CMSFFT to return to an enhanced image.

The paper also presents a novel loss function that combines RGB, LAB, and LCH color spaces. This design is based on principles of human vision, aiming to improve the contrast and saturation of the enhanced images. The methodology also incorporates an attention mechanism that operates along the channel-axis rather than the classical patch-axis. This allows the network to focus on channels that exhibit more severe degradation, improving the overall enhancement quality.

### 3.2.2 WaterNet

Water-Net is a novel underwater image enhancement network introduced in the paper (Li, Chunle, Ren, Cong, Hou, Kwong and Tao, 2019). Water-Net is designed specifically for enhancing underwater images. It aims to improve the visual quality of these images, which often suffer from issues like color distortion and reduced visibility due to the underwater environment. The network leverages the insights gained from the

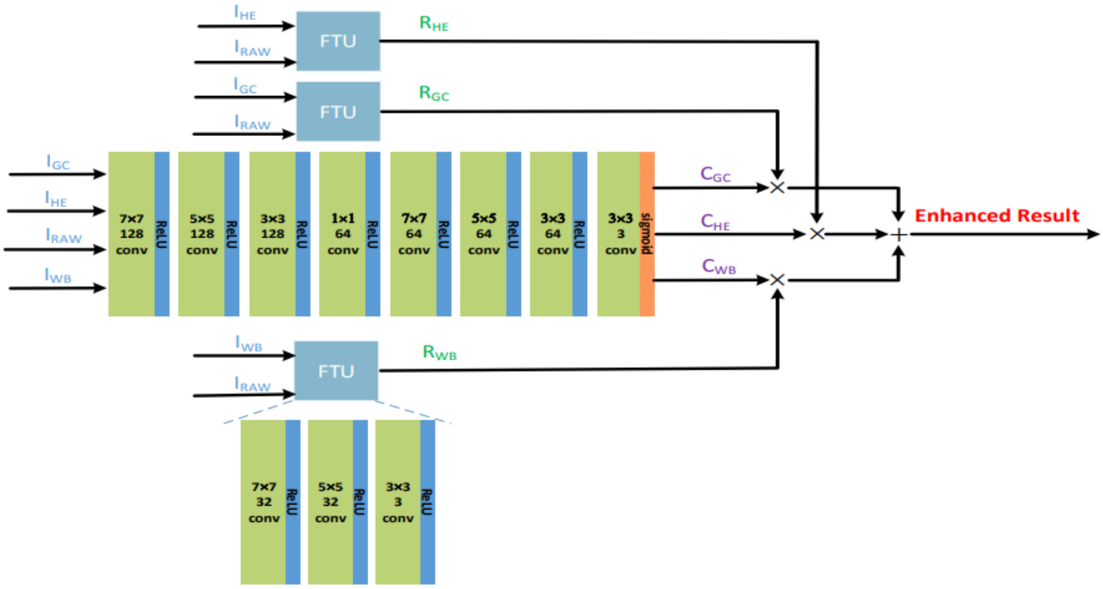



FIGURE 3.5 – Architecture of WaterNet (Li, Chunle, Ren, Cong, Hou, Kwong and Tao, 2019)

Underwater Image Enhancement Benchmark (UIEB) to achieve its goals.
Water-Net is trained using the comprehensive dataset provided by the UIEB, which consists of 950 real-world underwater images. This training allows the network to learn effective enhancement techniques that are applicable to a wide range of underwater images, making it a robust solution for this specific domain.
The paper highlights that Water-Net demonstrates good generalization capabilities when trained on the UIEB. This means that the network can effectively enhance various underwater images, even those that it has not encountered during training. This characteristic is crucial for practical applications in real-world scenarios where the conditions can vary significantly.
The authors conduct evaluations of Water-Net alongside other state-of-the-art algorithms, showcasing its performance and identifying its strengths and weaknesses. This evaluation is essential for understanding how well Water-Net performs in comparison to existing methods and for guiding future improvements in underwater image enhancement techniques.
Water-Net is likely built on a deep learning framework, specifically utilizing Convolutional Neural Networks (CNNs). CNNs are commonly used in image processing tasks due to their ability to learn spatial hierarchies of features, which is essential for enhancing image quality in complex environments like underwater settings.

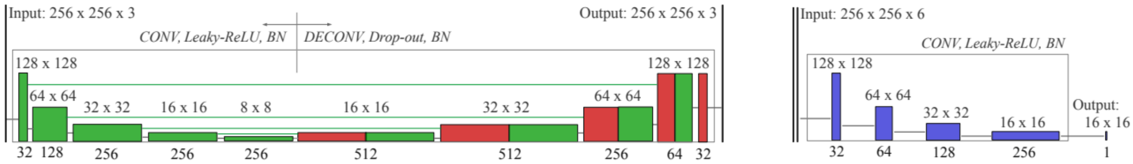



FIGURE 3.6 – Architecture of FUnIE-GAN generator (left) and discriminator (right) (Islam et al., 2020)

### 3.2.3 FUnIE-GAN

The model presented in (Islam et al., 2020) is built on the principles of generative adversarial networks, which consist of two main components : a generator and a discriminator. The generator is responsible for creating enhanced images from the input underwater images, while the discriminator evaluates the quality of the generated images against real images to guide the training process.
The model operates conditionally, meaning it takes specific input conditions into account during the image enhancement process. This allows the model to tailor the enhancement based on the characteristics of the input image, leading to more effective results.
The architectures of the generator and discriminator are relatively simplistic. The generator once again follows a U-shaped architecture, since it is generating a pixel-wise enhancement of the original image. While the discriminator is using convolutional layers to gather information from the generated and input image.
The model is trained using the EUVP dataset, which is also presented in this paper, which includes both paired and unpaired collections of underwater images. This diverse dataset allows the model to learn from various visibility conditions and camera characteristics, enhancing its robustness and effectiveness in real-world applications.

### 3.2.4 NU2NET

(Guo et al., 2022) presents a novel underwater image enhancement model that is closely associated with the URanker method. The authors highlight that a simple U-shape UIE network can achieve promising performance when it is used in conjunction with the pre-trained URanker. This indicates that the U-shape architecture is effective for enhancing underwater images, likely due to its ability to capture both local and global features through its encoder-decoder structure.

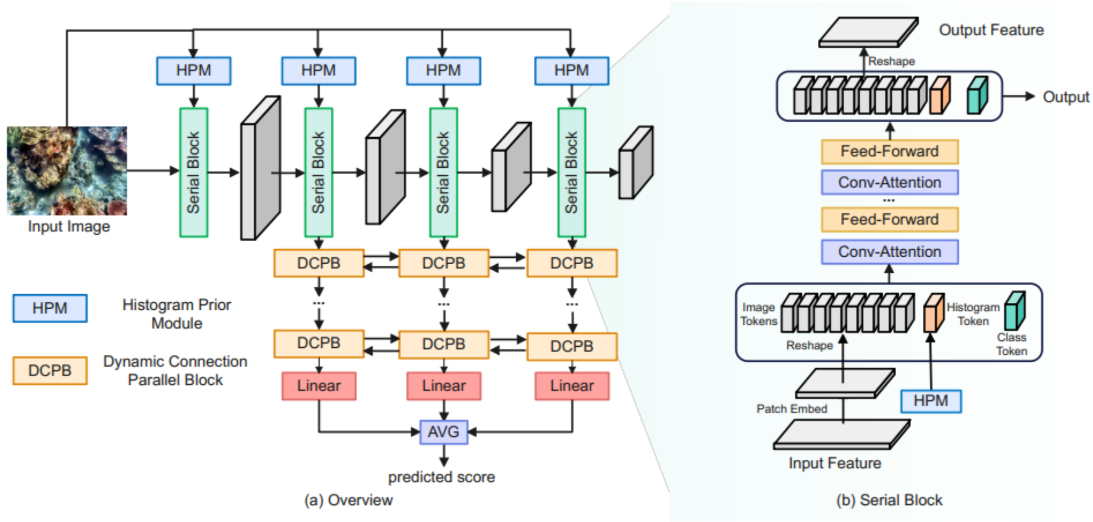



FIGURE 3.7 – Architecture of URanker (Guo et al., 2022)

URanker is a method developed for underwater image quality assessment (UIQA). İt is designed to rank the visual quality of underwater images that have been enhanced by different underwater image enhancement (UIE) algorithms. It effectively determines which enhanced image is better in terms of visual quality, making it a valuable tool for evaluating enhancement techniques.

The method is built on an efficient conv-attentional image Transformer. This architecture allows URanker to leverage attention mechanisms to focus on important features within underwater images, enhancing its ability to assess quality accurately.

- Histogram Prior : URanker incorporates a histogram prior that captures the color distribution of underwater images. This prior is represented as a histogram token, which helps the model attend to global degradation in image quality.
- This feature models local degradation by establishing correspondences across different scales of the image. It allows URanker to consider multi-scale dependencies, which is crucial for accurately assessing the quality of underwater images.

To train URanker, the authors created a dataset called URankerSet. This dataset includes a variety of underwater images enhanced by different UIE algorithms, along with their corresponding perceptual rankings. This comprehensive dataset is essential for training the model to perform effectively in real-world scenarios.

The integration of the U-shape UIE network with URanker serves as additional supervision. This coupling allows the UIE model to benefit from the ranking capabilities of URanker, which assesses the quality of enhanced images. By leveraging URanker's

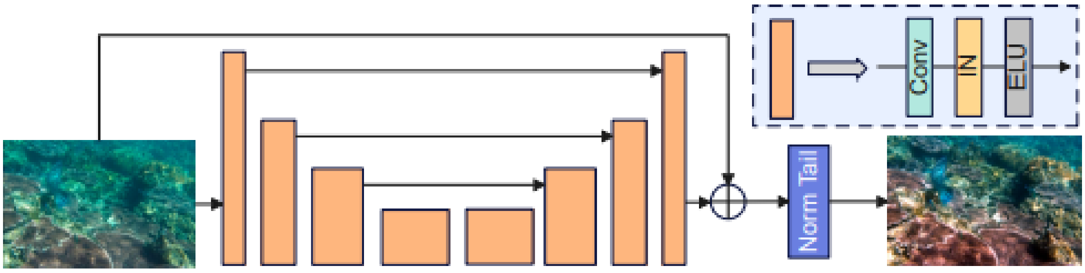



FIGURE 3.8 – Architecture of NU2NET (Guo et al., 2022)

insights, the UIE model can improve its output quality significantly.

The paper also introduces a normalization tail as part of the UIE model. This component is designed to enhance the performance of UIE networks, suggesting that it plays a crucial role in stabilizing and improving the quality of the enhanced images produced by the model.

The extensive experiments conducted in the study demonstrate that the proposed UIE model, especially when coupled with URanker and enhanced by the normalization tail, achieves state-of-the-art performance in underwater image enhancement tasks. This indicates that the model is not only innovative but also effective in practical applications.

## 3.3 Our Model

### 3.3.1 Model Architecture

Our approach to underwater image enhancement is guided by the hypothesis that a lightweight, well-designed neural network can outperform larger, more computationally intensive models when tailored to a specific downstream task—namely, photogrammetry. As shown by (Awasthi et al., 2022), smaller more task-driven models can and do usually outperform larger generalistic models. Unlike general-purpose enhancement models that prioritize visual appeal or perceptual quality across varied scenarios, our goal is to optimize image clarity and structural fidelity in a way that directly benefits 3D reconstruction pipelines. As a result, model interpretability, detail preservation, and computational efficiency are prioritized. In addition to better photogrammetric outcomes, such models are cheaper to train and more adaptable to different deployment environments, especially where hardware resources are limited.

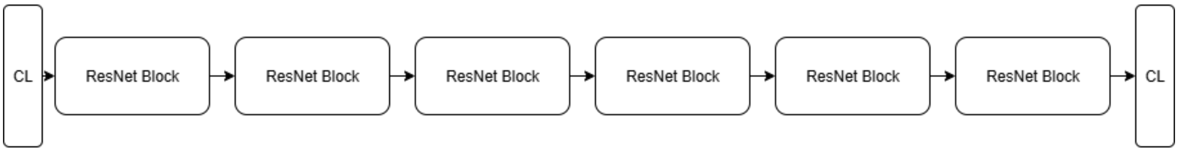



FIGURE 3.9 – The Architecture of the presented model

To achieve this, we base our architecture on residual learning, specifically utilizing ResNet-style blocks introduced in (He et al., 2015). Residual networks have become a cornerstone of modern computer vision due to their effectiveness in training deep architectures without encountering the vanishing or exploding gradient problem. By learning residual functions—i.e., the difference between the input and the desired output—ResNet blocks allow networks to go deeper while maintaining stability and convergence during training.

In our context, ResNet blocks offer several key advantages. First, they are inherently capable of preserving low-level features across layers. This is critical for underwater photogrammetry, where the geometric and textural integrity of the original images must be maintained to ensure accurate 3D reconstructions. Unlike classification tasks that benefit from increasing abstraction, our task demands precise retention of edges, contours, and spatial structure.

Second, the residual learning formulation aligns naturally with our enhancement objective. Since the degraded input image and the ideal output image share much of the same structure, learning the difference between them (i.e., the residual) is more efficient and effective than learning the full mapping directly. This allows the network to focus on correcting the effects of color distortion, contrast loss, and blurring—without unnecessarily altering the underlying scene content.

Finally, the simplicity and modularity of ResNet blocks make them well-suited for resource-constrained settings. Their efficient training behavior enables us to build and fine-tune smaller models that require fewer parameters and less GPU time. This opens the door for future applications where models could be deployed directly on embedded systems or dive computers, enabling real-time or near-real-time image enhancement in underwater field operations.

In summary, our model leverages the strengths of residual learning to build a compact, task-specific architecture that supports high-fidelity image enhancement in underwater

environments—particularly for photogrammetric applications where visual consistency and geometric accuracy are paramount.
In our case, we have decided to use 6 ResNet blocks, with an entry and exit convolutional layer to regulate input and output. Each of our ResNet blocks has an entry and exit size of 64 channels. With the entry and exit layers, the model contains approximately 446000 parameters. These blocks were connected by simply taking the output of one block as the entry parameters for the next.

### 3.3.2 Training Process

Given the nature of our dataset—comprised of a large number of high-resolution images that are varied yet collected under consistent conditions—it is essential to carefully tune our training procedure to balance generalization and specificity. While the dataset is sufficiently diverse to avoid trivial overfitting, the uniformity in image structure and capture protocol makes it possible for a model to overfit subtle features if the training process is not properly regularized.

As a first step, the images in the dataset - original and ground truth - are randomly cropped and resized to fit into our model using bilinear interpolation, contained within the Resize function of the 'torchvision' library. This is standard practice in data augmentation and has been shown to positively affect model performance (Krizhevsky et al., 2012). There hasn't been any extra steps taken to regularize the data, since it has been shown to negatively affect the performance of models in image enhancement tasks by (Zhang et al., 2016), who argue that regularization steps like Dropout disrupt pixel-level mapping.

To guide the learning process, we selected the L1 loss function (mean absolute error) as our primary objective. This choice is motivated by several factors. First, L1 loss is particularly well-suited for pixel-wise image regression tasks, as it encourages the preservation of fine details and avoids the blurring artifacts commonly associated with L2 (mean squared error) loss. Second, L1 loss is computationally lightweight and less sensitive to outliers, making it ideal for high-resolution data where occasional distortions or lighting anomalies may occur. For our image enhancement task, where structural fidelity and color correction are both critical, L1 provides a strong balance between performance and simplicity.

We employed the Adam optimizer for parameter updates during training, with an initial learning rate set to $1 \times 10^{-4}$. Adam is a widely used optimization algorithm that combines the benefits of momentum and adaptive learning rates, making it effective for fast convergence and stable training in a wide range of deep learning applications, including image restoration.

The model was trained for a total of 20 epochs. This relatively modest number of epochs was deliberately chosen to mitigate the risk of overfitting while still allowing the model to learn meaningful features. Throughout training, we monitored both training and validation loss to ensure that performance improvements were consistent and not the result of memorization. Additionally, we empirically verified that this number of epochs was sufficient to achieve acceptable levels of color correction and image enhancement, with no significant signs of performance degradation.

In summary, our training setup emphasizes simplicity and efficiency. By using an L1 loss function, the Adam optimizer, and a conservative number of training epochs, we were able to develop a model that generalizes well to new underwater images while preserving key visual features necessary for photogrammetric applications.

## 3.4 Peak Signal-to-Noise Ratio

The primary metric employed for comparing the effectiveness of these methodologies will be Peak Signal-to-Noise Ratio (PSNR). PSNR is a widely recognized measure in image processing and video compression that quantifies the quality of reconstructed images relative to their original counterparts. Specifically, PSNR assesses the ratio between the maximum possible power of the signal—the original image—and the power of the noise that distorts its fidelity, which arises during the enhancement process.
A higher PSNR value typically indicates superior quality in the reconstructed image, suggesting reduced distortion and noise levels. In practice, PSNR values generally range from 20 dB to 40 dB, with values exceeding 30 dB often categorized as indicative of good quality. However, while PSNR is a valuable metric, it is important to note that it does not always align with human perceptual quality assessments. Factors such as contrast, color fidelity, and structural information can significantly influence the perceived quality of an image, aspects that PSNR alone does not adequately capture.
In 3.10, the $MAX_I$ represents the maximum value that can be attained in the context.

$$\begin{aligned} PSNR &= 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right) \\ &= 20 \cdot \log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right) \\ &= 20 \cdot \log_{10}(MAX_I) - 10 \cdot \log_{10}(MSE). \end{aligned}$$

FIGURE 3.10 – The PSNR Formula

In our case it will correspond to 1, since pixel values will be normalized from 0-255 to 0-1. MSE stands for Mean Squared Error, which will be calculated pixel-wise between the original image and the enhanced image.

### 3.5 Structural Similarity Index Measure

SSIM is a full-reference image quality metric : it compares an image under evaluation to a reference (ideal or ground-truth) image. Its core idea is that the human visual system is very sensitive to structural information (edges, textures, patterns) rather than just raw pixel differences. To that end, SSIM assesses image similarity via three components computed over local windows : luminance (mean intensity), contrast (standard deviation), and structure (covariance normalized by contrast). These three are combined into a multiplicative formula like 3.11 where $\mu$, $\sigma$ refer to means and variances in local patches, $\sigma_x y$ is covariance, and C1, C2 are small constants to stabilize divisions.
SSIM tends to correlate much better with human judgments of image quality than sim-

$$\mathrm{SSIM}(\mathbf{x},\mathbf{y}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)}.$$

FIGURE 3.11 – The PSNR Formula

pler pixel-based metrics such as Mean Squared Error (MSE) or Peak Signal to Noise Ratio (PSNR), especially when distortions affect contrast, blur, or texture. However, SSIM is less useful when there's no reference image available ; also, it's sensitive to misalignment, color shifts, or when distortions are outside its assumed model (non-stationary distortions).

### 3.6 Natural Image Quality Evaluator

NIQE (Mittal et al., 2013) is a no-reference (blind) image quality metric that estimates perceptual quality without needing subjective human scores or a clean reference image. It builds on the concept of Natural Scene Statistics (NSS) : high-quality natural images, when processed through certain transformations (like Mean Subtracted Contrast Normalized coefficients), tend to follow predictable statistical distributions. NIQE is trained on a large corpus of undistorted natural images to model these "quality-aware" statistics. To evaluate a new image, NIQE extracts local statistical features, compares them against the learned distribution of pristine images, and outputs a quality score. One of NIQE's main strengths is that it requires no human opinion data during training, making it a fully unsupervised quality metric. Its scores typically range with lower values indicating better perceptual quality. Because it generalizes across different distortion types, NIQE has been widely used in image enhancement, denoising, and super-resolution research where no reference image is available. However, it can sometimes misjudge images that intentionally deviate from natural statistics (for example, stylized or medical images), since it assumes "naturalness" as a proxy for quality. In practice, NIQE is often used alongside PIQE and BRISQUE to give a more rounded picture of enhancement performance.

### 3.7 Perception based Image Quality Evaluator

PIQE (Venkatanath et al., 2015) is a no-reference (blind) image quality metric, meaning it does not require a pristine or "ground-truth" reference image. Rather, PIQE evaluates distortions directly from the given image. It works by dividing the image into non-overlapping blocks (often 16×16), computing local features in each block—particularly using the MSCN (Mean Subtracted, Contrast Normalized) coefficients—and then determining which blocks are "spatially active" (i.e. showing significant variation). Blocks

with higher local variance or perceptible distortion are identified, and a global quality score is computed based on those.
The resulting PIQE score is inversely correlated with perceptual image quality : a lower score means better quality. The scale is usually from 0 to 100. Besides the global score, PIQE can also produce spatial "masks" : which blocks are active, where noise or artifacts are, etc. This block-wise sensitivity allows it to locate problem areas, which is useful for image enhancement tasks. Although PIQE is "unsupervised" (does not use human opinion scores in training), it has been found to behave similarly to NIQE in many scenarios.

### 3.8 Blind/Referenceless Image Spatial Quality Evaluator

BRISQUE (Mittal et al., 2011) is also a no-reference quality metric, but unlike PIQE it is opinion-aware—that is, it is trained using images with known distortions along with subjective human quality scores. The method is strongly based on Natural Scene Statistics (NSS). In essence, clean, undistorted (or lightly distorted) "natural" images have characteristic statistical properties (e.g. distributions of certain normalized luminance differences) which tend to deviate when distortions, noise, blur, compression artifacts etc. are introduced. BRISQUE extracts features of these statistics (e.g. MSCN coefficients, pairwise products, etc.), then uses a trained regression model (e.g. Support Vector Regression) to map those features to quality scores.
The advantages of BRISQUE include good correlation with human perception across many kinds of distortions, and relatively modest computational cost after training. However, there are limitations : if the algorithm encounters distortions quite different from those in its training set, its predictions may suffer. Also, since it is blind/no-reference, it can't directly tell "how far" from a reference image an enhancement brought you—only that "this looks more or less good" in light of its learned norms. In enhancement work, BRISQUE is often used to compare different enhanced versions when no clean reference exists.

# 4 EXPERIMENT AND RESULTS

## 4.1 Experiment

This dataset presents a valuable opportunity to establish a novel benchmarking framework for underwater image enhancement, a domain that remains comparatively under-explored within the broader field of image processing. The limited availability of specialized datasets, frequently marked by a high variability of image content, hampers the ability to perform rigorous, meaningful comparisons among different enhancement techniques. In response to this gap, our dataset provides a structured collection of underwater images, specifically designed to facilitate a more systematic and robust evaluation of diverse methodologies.

Our investigation encompasses the implementation of a comprehensive range of established image enhancement methods, each offering a unique approach to the enhancement challenge. These methods have previously been benchmarked in the field and are accompanied by publicly accessible model weights and codebases, thereby supporting straightforward and consistent implementation. Such accessibility is fundamental to fostering reproducibility and comparability across research initiatives, enhancing the reliability and impact of comparative analyses.

In the course of this study, each selected method will be applied to our dataset, with performance evaluated quantitatively through Peak Signal-to-Noise Ratio (PSNR) values. These quantitative metrics will enable us to conduct a rigorous comparative assessment of the enhancement techniques. All analyses will be carried out using publicly available models and their corresponding Python implementations, with computational resources provided by a Tesla T4 GPU. This methodologically transparent approach not only strengthens the study's replicability but also situates our research within the expanding body of work dedicated to advancing underwater imaging technologies.

In the end, we will compare the results with the ResNet-based model presented.

### 4.1.1 PSNR

For the calculation of the PSNR values, we have used the following code snippet, following the formula given previously. We will use this for all of the model tests and outputs.

```
def compute_psnr(img1, img2):
    mse = np.mean( (img1/255. - img2/255.) ** 2 )
    if mse < 1.0e-10:
        return 100
    MAX_PIXEL = 1
    return 20 * math.log10(MAX_PIXEL / math.sqrt(mse))
```

### 4.1.2 U-Shape Transformer

Our initial approach involves applying the U-shape transformer model to our dataset. To facilitate this, we have established a computational environment specifically configured with our dataset and are utilizing publicly accessible pre-trained model weights and testing scripts provided by the original authors via GitHub. This publicly available code comprehensively implements the model architecture and facilitates weight loading, enabling streamlined model application. The script sequentially processes images within a designated folder, passing each through the model, and subsequently stores the enhanced outputs in a separate directory for further analysis and evaluation. The loop also contains a calculation for the PSNR value using the previous code snippet given.

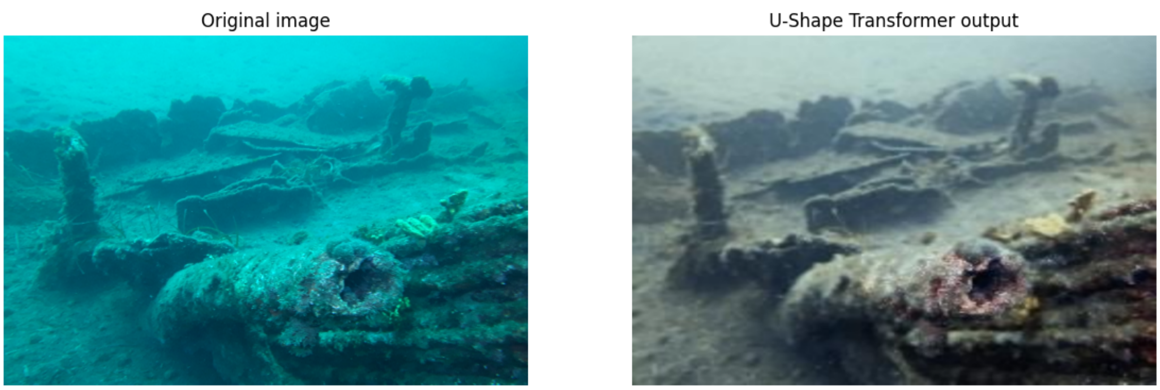



FIGURE 4.1 – U-Shape Transformer Comparisons

### 4.1.3 WaterNet

We will then use the WaterNet model on our dataset. In a similar fashion, we set up our environment with our dataset. The WaterNet model, however, is publicly available in the PyTorch Hub and can easily be called using a single Python command as such :

```
preprocess, postprocess, model = torch.hub.load('tnwei/waternet', 'waternet')
```

We then implemented a similar loop where each image is processed through the model, it's PSNR value is calculated and the output is saved in a separate folder.

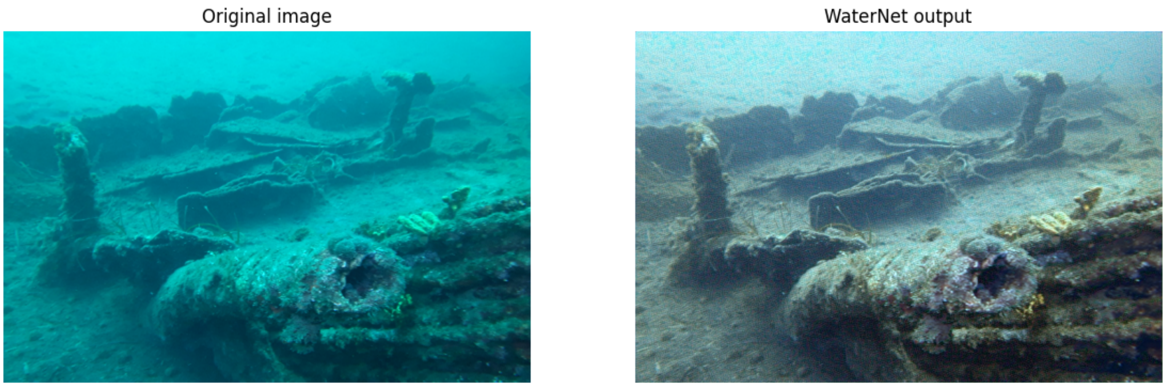



FIGURE 4.2 – WaterNet Comparisons

### 4.1.4 FUnIE-GAN

For this method, we also applied the publicly available code and pre-trained weight provided by the authors. The script sequentially processes images within a designated folder, passing each through the model, and subsequently stores the enhanced outputs in a separate directory for further analysis and evaluation. The PSNR calculation was added to this code snippet using the previously given formula and code.

### 4.1.5 NU2NET

For this model, we also used the publicly available code and pre-trained weight provided by the authors. The script sequentially processes images within a designated folder, passing each through the model, and subsequently stores the enhanced outputs in


Figure 4.3 – FUnIE-GAN Comparisons

a separate directory for further analysis and evaluation. The PSNR values are also calculated within the loop ; this was implemented in another way and was changed to our exact code snippet to preserve consistency.


Figure 4.4 – Nu2NET Comparisons

### 4.1.6 Our ResNet approach

For this model, we implemented the previously discussed model using PyTorch, with a relatively simple approach of using ReLU layers sandwiched by convolutional layers for each ResNet block, as follows :

```
import torch
import torch.nn as nn
import torch.nn.functional as F

class ResidualBlock(nn.Module):
```


FIGURE 4.5 – ResNet Comparisons

```
    def __init__(self, channels):
        super().__init__()
        self.conv1 = nn.Conv2d(channels, channels, 3, padding=1)
        self.relu = nn.ReLU(inplace=True)
        self.conv2 = nn.Conv2d(channels, channels, 3, padding=1)

    def forward(self, x):
        return x + self.conv2(self.relu(self.conv1(x)))

class UnderwaterResNetEnhancer(nn.Module):
    def __init__(self, num_blocks=6):
        super().__init__()
        self.entry = nn.Conv2d(3, 64, kernel_size=3, padding=1)
        self.res_blocks = nn.Sequential(
            *[ResidualBlock(64) for _ in range(num_blocks)]
        )
        self.exit = nn.Conv2d(64, 3, kernel_size=3, padding=1)

    def forward(self, x):
        x = F.relu(self.entry(x))
        x = self.res_blocks(x)
        x = self.exit(x)
        return torch.clamp(x, 0, 1)  # Ensure output is in [0, 1]
```

## 4.2 Results

### 4.2.1 PSNR Comparison

TABLE 4.1 – PSNR Comparisons for different methods

| Method | PSNR average | PSNR Min | PSNR Max |
|---|---|---|---|
| U-shape Transformer | 21.22 | 20.85 | 23.24 |
| WaterNet | 21.92 | 19.98 | 24.54 |
| FUnIE-GAN | 12.28 | 11.56 | 14.5 |
| NU2Net | 22.47 | 21.66 | 23.68 |
| Ours | 25.48 | 22.04 | 28.32 |

The table presents the performance of several models evaluated using the Peak Signal-to-Noise Ratio (PSNR), a common metric for assessing image restoration quality. Looking at the average PSNR values, NU2Net performs the best among the existing methods, with an average of 22.47, followed closely by WaterNet at 21.92 and the U-Shape Transformer at 21.22. FUnIE-GAN, in contrast, has a much lower average PSNR of 12.28, indicating that it struggles to handle the type of data used in this study. Overall, these results suggest that while some state-of-the-art models are reasonably effective, there is a noticeable difference in performance depending on the architecture and approach.

Our model achieves the highest average PSNR of 25.48, significantly outperforming all the other methods. Its minimum PSNR of 22.04 is higher than the averages of most of the other models, showing that it consistently performs better across different examples. The maximum PSNR of 28.32 also demonstrates that our model can reach high-quality reconstructions under favorable conditions. These numbers indicate that a specialized approach tailored to the task can provide meaningful improvements over more general architectures, supporting the idea that task-specific design is valuable for real-world applications.

It is also important to note that even the best-performing models, including ours, do not fully reach the ideal PSNR of around 30, which is often considered a benchmark for high-quality restoration. This gap highlights the challenges of real-world data, where noise, lighting, and other factors can make accurate reconstruction difficult. Nevertheless, the performance of our model demonstrates clear progress compared to existing approaches. In addition to higher PSNR values, our model offers practical advantages such as easier training, lower computational cost, and adaptability, making it a strong candidate for real-world deployment while still leaving room for future improvements.

### 4.2.2 SSIM Comparison

Table 4.2 – SSIM Comparisons for different methods

| Method | SSIM average | SSIM Min | SSIM Max |
|---|---|---|---|
| U-shape Transformer | 0.67 | 0.65 | 0.71 |
| WaterNet | 0.76 | 0.73 | 0.78 |
| FUnIE-GAN | 0.79 | 0.77 | 0.86 |
| NU2Net | 0.69 | 0.68 | 0.76 |
| Ours | 0.85 | 0.84 | 0.88 |

The table shows the performance of different models using the Structural Similarity Index (SSIM), which measures perceived image quality by comparing structural information between images. Among the previously established methods, FUnIE-GAN and WaterNet achieve relatively high average SSIM values of 0.79 and 0.76, respectively, while NU2Net and the U-Shape Transformer have lower averages of 0.69 and 0.67. These results indicate that although some models can preserve structural details reasonably well, there is still variation in their ability to maintain image fidelity consistently across different examples.

Our model achieves the highest average SSIM of 0.85, clearly surpassing all other methods. Its minimum SSIM of 0.84 is higher than the average SSIM of all competing models, suggesting that it maintains strong performance even in more challenging cases. The maximum SSIM of 0.88 shows that under optimal conditions, the model can produce reconstructions that are very close to the original in terms of structural quality. These results highlight the advantage of designing a task-specific model that focuses on preserving key structural details, rather than relying solely on general-purpose architectures.

Despite the strong performance, even the best models do not reach an SSIM of 1.0, which would indicate perfect structural reconstruction. This reflects the inherent difficulties in real-world image restoration, where noise, distortions, and varying conditions make perfect recovery challenging. Nevertheless, the high and consistent SSIM values of our model demonstrate clear improvements over existing methods, supporting the notion that a specialized approach can achieve both high accuracy and reliability. Beyond the quantitative advantage, the model also offers practical benefits such as simpler training, lower computational requirements, and adaptability to different scenarios, making it a strong choice for real-world applications.

### 4.2.3 NIQE Comparison

TABLE 4.3 – NIQE Comparisons for different methods

| Method | NIQE average | NIQE Min | NIQE Max |
|---|---|---|---|
| U-shape Transformer | 4.03 | 3.68 | 4.33 |
| WaterNet | 5.61 | 5.18 | 6.35 |
| FUnIE-GAN | 6.67 | 4.41 | 6.94 |
| NU2Net | 4.43 | 4.25 | 4.46 |
| Ours | 5.91 | 5.06 | 7.08 |

The table presents the performance of several models using the Naturalness Image Quality Evaluator (NIQE), a no-reference metric where lower scores indicate better perceived image quality. According to the results, the U-Shape Transformer and NU2Net achieve the lowest average NIQE values at 4.03 and 4.43, respectively, suggesting that these models produce images that are closer to natural-looking quality compared to the other methods. WaterNet and FUnIE-GAN have higher average NIQE scores of 5.61 and 6.67, indicating a drop in naturalness despite other performance advantages they might have in terms of structural similarity or PSNR.

Our model shows an average NIQE of 5.91, which is higher than U-Shape Transformer and NU2Net but comparable to WaterNet. Its minimum NIQE of 5.06 indicates that in some cases, the generated images are closer to natural quality, while the maximum NIQE of 7.08 shows that there are instances where the output is less natural-looking. These results suggest that while our model excels in PSNR and SSIM, there is a trade-off with perceived naturalness as measured by NIQE. This highlights an important point : different metrics capture different aspects of image quality, and a model optimized for structural fidelity or signal quality may not always produce the most natural-looking images.

Overall, the NIQE results underline the complexity of evaluating image restoration models comprehensively. While our approach clearly improves structural similarity and signal-to-noise quality, there is still room to optimize the natural appearance of the output. This emphasizes the importance of balancing multiple objectives in image restoration tasks. Despite these challenges, the performance of our model across various metrics demonstrates its overall robustness and suggests it is a strong candidate for practical applications, even if some trade-offs in naturalness remain.

Table 4.4 – PIQE Comparisons for different methods

| Method | PIQE average | PIQE Min | PIQE Max |
|---|---|---|---|
| U-shape Transformer | 25.94 | 23.02 | 26.08 |
| WaterNet | 33.10 | 31.39 | 34.95 |
| FUnIE-GAN | 34.01 | 30.58 | 34.34 |
| NU2Net | 36.45 | 35.30 | 36.94 |
| Ours | 53.12 | 46.51 | 63.07 |

### 4.2.4 PIQE Comparison

The table shows the performance of different models using the Perception-based Image Quality Evaluator (PIQE), a no-reference metric where lower scores indicate better perceived quality. Among the previously established methods, the U-Shape Transformer achieves the lowest average PIQE of 25.94, suggesting that it produces images that appear more natural and less degraded according to this metric. WaterNet and FUnIE-GAN show higher average PIQE values of 33.10 and 34.01, while NU2Net performs even worse with an average of 36.45. These results indicate that, in terms of perceptual quality as measured by PIQE, many state-of-the-art models do not consistently produce images that align with human visual preferences.

Our model, in contrast, has the highest average PIQE of 53.12, with a minimum of 46.51 and a maximum of 63.07. This indicates that while our model excels in metrics like PSNR and SSIM, it produces images that are perceived as less natural according to PIQE. The higher PIQE scores highlight a trade-off between optimizing for structural accuracy and signal fidelity versus perceptual naturalness. In other words, our model focuses on improving reconstruction quality and structural similarity, which can sometimes come at the expense of visual smoothness or natural texture.

These findings emphasize the complexity of evaluating image restoration models using multiple metrics. While our approach clearly outperforms other methods in PSNR and SSIM, and performs moderately in NIQE, the PIQE results suggest that further work is needed to enhance the perceptual realism of the outputs. This underscores that no single metric can fully capture image quality and that a holistic evaluation should consider both signal fidelity and human perception. Overall, the results demonstrate the strengths and limitations of our model, showing it is highly effective in reconstruction but may require additional refinement for perceptual quality.

### 4.2.5 BRISQUE Comparison

Table 4.5 – BRISQUE Comparisons for different methods

| Method | BRISQUE average | BRISQUE Min | BRISQUE Max |
|---|---|---|---|
| U-shape Transformer | 69.37 | 63.33 | 70.24 |
| WaterNet | 73.91 | 69.97 | 75.54 |
| FUnIE-GAN | 79.54 | 77.41 | 83.12 |
| NU2Net | 59.98 | 51.82 | 68.89 |
| Ours | 55.73 | 48.67 | 67.77 |

The table presents the performance of different models evaluated using BRISQUE, a no-reference image quality metric where lower scores indicate better perceptual quality. Among the baseline methods, NU2Net achieves the lowest average BRISQUE score of 59.98, followed by the U-Shape Transformer at 69.37. WaterNet and FUnIE-GAN perform worse, with average scores of 73.91 and 79.54, indicating that their outputs are generally less natural-looking according to this metric. These results show that not all state-of-the-art models maintain high perceptual quality, even if they perform well on metrics like PSNR or SSIM.

Our model achieves the lowest average BRISQUE score at 55.73, with a minimum of 48.67 and a maximum of 67.77. This indicates that, compared to other methods, our model produces images that are perceived as more natural and visually appealing. The results suggest that the model's design successfully balances reconstruction accuracy with perceptual quality, at least as measured by BRISQUE, which aligns more closely with human visual perception than some other metrics.

Overall, these BRISQUE results highlight the importance of including perceptual metrics in model evaluation. While metrics like PSNR and SSIM measure structural fidelity and signal quality, BRISQUE provides insight into how realistic the output images appear to human observers. The strong performance of our model across BRISQUE, combined with its high PSNR and SSIM, demonstrates that it not only restores images accurately but also maintains a high level of perceptual quality, making it suitable for real-world applications where visual realism is important.

### 4.2.6 Visual Comparison

For Visual comparison, we conducted a questionnaire using 10 reference images and their enhanced forms with each method blindly and asked experts and professionals to select the one they prefer the best visually. This was done with a small sample size and

might not reflect a general population.
Five reference images were selected at random, and a questionnaire was sent showing the original image and the outputs of the five models. The given images can be seen in 4.6, 4.7, 4.8, 4.9 and 4.10. There were 31 total participants in the questionnaire

and the model names were not given to prevent any kind of bias between questions. The table 4.6 shows us the percentage of selection of each model in each question.

Table 4.6 – Percentage of selection of each model per question

| | U-Shape Transformer | WaterNet | FUnIE-GAN | NU2NET | Our Model |
|---|---|---|---|---|---|
| Q1 | 25.8 | 9.7 | 16.1 | 19.4 | **29** |
| Q2 | 6.5 | **61.3** | 9.7 | 3.2 | 19.4 |
| Q3 | 3.2 | **61.3** | 3.2 | 6.5 | 25.8 |
| Q4 | 0 | **77.4** | 9.7 | 0 | 12.9 |
| Q5 | 25.8 | 9.7 | 25.8 | 3.2 | **35.5** |

The performance comparison across the five metrics highlights clear differences in the strengths of the models. WaterNet stands out in metrics Q2 through Q4, with scores of 61.3, 61.3, and 77.4, respectively, showing it performs exceptionally well in these areas. However, its performance in Q1 and Q5 is much weaker, scoring only 9.7 in Q1 and 9.7 in Q5, suggesting that its strengths are concentrated in specific tasks. In contrast, Our model demonstrates a more balanced profile, achieving the highest scores in Q1 (29) and Q5 (35.5) and maintaining moderate performance in Q2–Q4, where it lags behind WaterNet but still outperforms most of the other models.

The remaining models, U-Shape Transformer, FUnIE-GAN and NU2NET, generally show lower performance overall. FUnIE-GAN and U-Shape Transformer do reasonably well on Q1 and Q5 but fall short on the middle metrics, while NU2NET consistently scores near the bottom across all five metrics. Overall, this comparison suggests that while WaterNet is highly effective in certain focused scenarios, our model offers more consistent and versatile performance across a wider range of evaluation criteria.

Considering the fact that generalistic models are more costly to train and require more hardware to run, we can say that the unstable performance and moderate performance is reason for reconsideration.

| | |
|---|---|
| Original Image |  |
| U-Shape Transformer |  |
| WaterNet |  |
| FUnIE-GAN |  |
| NU2Net |  |
| Our Model |  |

FIGURE 4.6 – 1st image

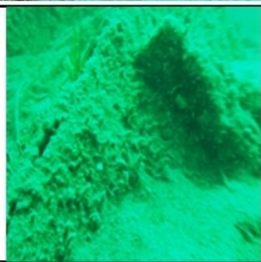
| | |
|---|---|
| Original Image |  |
| U-Shape Transformer |  |
| WaterNet |  |
| FUnIE-GAN |  |
| NU2Net |  |
| Our Model |  |

FIGURE 4.7 – 2nd image

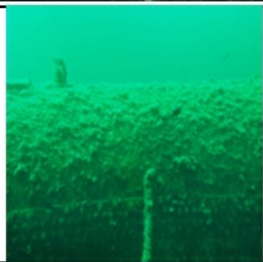
| | |
|---|---|
| Original Image |  |
| U-Shape Transformer |  |
| WaterNet |  |
| FUnIE-GAN |  |
| NU2Net |  |
| Our Model |  |

FIGURE 4.8 – 3rd image

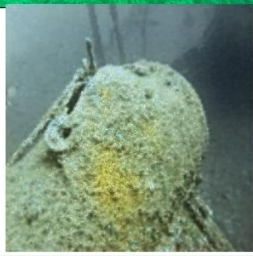
| | |
|---|---|
| Original Image |  |
| U-Shape Transformer |  |
| WaterNet |  |
| FUnIE-GAN |  |
| NU2Net |  |
| Our Model |  |

FIGURE 4.9 – 4th image

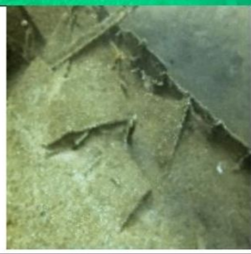
| | |
|---|---|
| Original Image |  |
| U-Shape Transformer |  |
| WaterNet |  |
| FUnIE-GAN |  |
| NU2Net |  |
| Our Model |  |

Figure 4.10 – 5th image

# 5 CONCLUSION AND FUTURE WORKS

## 5.1 Discussion

Underwater photogrammetry has quickly become a vital tool for marine ecologists because it allows them to study the seafloor in a way that is accurate, detailed, and non-destructive. Its popularity has grown not only because of the quality of data it provides but also thanks to the falling costs and wider availability of underwater imaging technologies. This accessibility has opened the door for both scientists and conservationists to use photogrammetry in a wide range of applications.

That said, achieving reliable results goes far beyond simply taking pictures underwater. High-quality photogrammetry requires careful planning and precision. Researchers must carry out independent quality checks, calibrate their imaging systems, design effective camera setups, and thoroughly model the entire imaging process. Without this kind of structured framework—one that accounts for measurement errors and uncertainties—it becomes very difficult to draw scientifically meaningful or statistically reliable conclusions. In other words, the strength of the analysis depends directly on the rigor of the setup.

When it comes to image enhancement and reconstruction, many existing underwater datasets and models take a one-size-fits-all approach. Often, these models are developed without a clear end use in mind. This has led to the rise of increasingly complex architectures, such as U-shaped Transformers and multi-branch convolutional networks. While these models look impressive, they demand massive and diverse datasets for training—something that is particularly hard to gather in underwater environments due to practical, environmental, and technical challenges.

This mismatch between model complexity and the limited availability of suitable data creates a significant gap. Large models may perform well on carefully curated benchmarks, but in real-world conditions they often fail to adapt. In our own experiments, we found that these models frequently struggled to generalize when applied to structured, purpose-driven datasets like ours. Even the gains they achieve tend to be modest. For example, improvements in Peak Signal-to-Noise Ratio (PSNR) rarely surpass 23, despite the enormous computational and time costs of training.

This trend highlights a larger issue in the field of machine learning : chasing ever-bigger, generalized models doesn't always lead to better results in practice. Instead, it suggests a shift in focus may be needed—toward smaller, more efficient models designed for specific tasks, where performance improvements are both meaningful and practical.
There were however shortcomings of using a more task-specific model when using metrics that measure without ground truth, and highlight more "natural" or "appealing" results. This is however very subjective, and some results that show some model is better than another could not align with what a human observer might consider.
Our dataset was developed with a clear and singular focus : underwater photogrammetry. Despite this narrow scope, it includes over 3,800 high-resolution images captured in realistic underwater conditions. This makes it not only one of the largest available underwater image datasets for this task, but also one of the most consistent and structured. This unique combination allows us to rigorously test and evaluate state-of-the-art models in a setting that closely resembles real operational use-cases.
Through extensive experimentation and comparison using widely accepted metrics, we found that while some advanced models do outperform others under ideal conditions, their performance significantly deteriorates when faced with the kinds of realistic distortions and variability found in our dataset. This underscores the importance of aligning model design and evaluation with the specific challenges of the intended application, rather than relying solely on generalized performance benchmarks.
Ultimately, our findings advocate for a more grounded approach to model development—one that prioritizes domain-specific effectiveness, interpretability, and efficiency over sheer complexity. This perspective, we believe, is especially important in fields like marine research, where resource constraints and environmental unpredictability make practical deployment a key consideration. The model that we present, with further fine-tuning, would be better suited for this specific instance of underwater photogrammetry.
We have also shown that the presented model gives a more consistent visual that is acceptable in some cases, and adequate in others.

## 5.2 Future Works

In a continuation to this work, we could more rigorously adjust the model structure, fine-tune the model and adjust the hyper-parameters. This could lead to getting better or more accurate results.

We also did not test the model in other environments and datasets (like the UIEB or the LSUI), as much as they do might not represent real use-cases (as seen by the generalistic models' poor performance), the community does consider these benchmarks as a way to compare certain models and rank them.

We also did not explore the fact that this dataset and approach could be a blueprint for future products. One could imagine an automated system where users could input their own images with depth info in order to get a localized enhancement model, since water conditions vary drastically between locations.

# REFERENCES


Akkaynak, D. and Treibitz, T. (2019). Sea-thru : A method for removing water from underwater images, *Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition* .

Awasthi, P., Recio-Mitter, D. and Sugi, Y. K. (2022). Maximizing use-case specificity through precision model tuning.
**URL:** *https ://arxiv.org/abs/2212.14206*

Berman, D., Levy, D., Avidan, S. and Treibitz, T. (2020). Underwater single image color restoration using haze-lines and a new quantitative dataset, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PP** : 1–1.

Du, D., Li, E., Si, L., Xu, F. and Niu, J. (2024). End-to-end underwater video enhancement : Dataset and model, *arXiv.org* **abs/2403.11506**.

Guo, C., Wu, R., Jin, X., Han, L., Chai, Z., Zhang, W. and Li, C. (2022). Underwater ranker : Learn which is better and how to be better.
**URL:** *https ://arxiv.org/abs/2208.06857*

He, K., Zhang, X., Ren, S. and Sun, J. (2015). Deep residual learning for image recognition.
**URL:** *https ://arxiv.org/abs/1512.03385*

Islam, M. J., Xia, Y. and Sattar, J. (2020). Fast underwater image enhancement for improved visual perception.

Kaneko, R., Higashi, H. and Tanaka, Y. (2024). Physics-inspired synthesized underwater image dataset, *arXiv.org* **abs/2404.03998**.

Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks, *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'12, Curran Associates Inc., Red Hook, NY, USA, p. 1097–1105.

Lei, X., Wang, H., Shen, J. and Liu, H. (2022). Underwater image enhancement based on color correction and complementary dual image multi-scale fusion., *Applied optics* **61 17**(17) : 5304–5314.

Levy, D., Peleg, A., Pearl, N., Rosenbaum, D., Akkaynak, D., Korman, S. and Treibitz, T. (2023). Seathru-nerf : Neural radiance fields in scattering media.

Li, C., Anwar, S., Hou, J., Cong, R., Guo, C. and Ren, W. (2021). Underwater image enhancement via medium transmission-guided multi-color space embedding, *IEEE Transactions on Image Processing* **30** : 4985–5000.

Li, C., Chunle, G., Ren, W., Cong, R., Hou, J., Kwong, S. and Tao, D. (2019). An underwater image enhancement benchmark dataset and beyond, *IEEE Transactions on Image Processing* **PP** : 1–1.

Li, C., Guo, C., Ren, W., Cong, R., Hou, J., Kwong, S. and Tao, D. (2019). An underwater image enhancement benchmark dataset and beyond.

Liu, X., Chen, Z., Xu, Z., Zheng, Z., Ma, F. and Wang, Y. (2024). Enhancement of underwater images through parallel fusion of transformer and cnn, *Journal of Marine Science and Engineering* **12**(9) : 1467–1467.

Mittal, A., Moorthy, A. K. and Bovik, A. C. (2011). Blind/referenceless image spatial quality evaluator, *2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, pp. 723–727.

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

Muniraj, M. and Dhandapani, V. (2023). Underwater image enhancement by modified color correction and adaptive look-up-table with edge-preserving filter, *Signal Processing-image Communication* **113** : 116939–116939.

Nie, X., Pan, S., Zhai, X., Tao, S., Qu, F., Wang, B., Ge, H. and Xiao, G. (2024). Image-conditional diffusion transformer for underwater image enhancement.

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

Ranftl, R., Lasinger, K., Hafner, D., Schindler, K. and Koltun, V. (2020). Towards robust monocular depth estimation : Mixing datasets for zero-shot cross-dataset

transfer.
**URL:** *https ://arxiv.org/abs/1907.01341*

Rende, S. F., Menna, F., Lagudi, A., Bruno, F., Severino, U., Montefalcone, M., Irving, A., Raimondi, V., Calvo, S., Pergent, G., Pergent-Martini, C. and Tomasello, A. (2022). Assessing seagrass restoration actions through a micro-bathymetry survey approach (italy, mediterranean sea), *Water* **14** : 1285.

Sethuraman, A. V., Ramanagopal, M. S. and Skinner, K. A. (2023). Waternerf : Neural radiance fields for underwater scenes.

Shu, Z., Wang, T., Dong, J. and Yu, H. (2017). Underwater image enhancement via extended multi-scale retinex, *Neurocomputing* **245**.

Telem, G. and Filin, S. (2010). Photogrammetric modeling of underwater environments, *ISPRS Journal of Photogrammetry and Remote Sensing* **65**(5) : 433–444.

Urbina-Barreto, I., Elise, S., Guilhaumon, F., Bruggemann, H., Pinel, R., Kulbicki, M., Vigliola, L., Mou-tham, G., Mahamadaly, V., Facon, M., Bureau, S., Peignon, C., Dutrieux, E., Garnier, R., Penin, L. and Adjeroud, M. (2022). Underwater photogrammetry reveals new links between coral reefscape traits and fishes that ensure key functions, *Ecosphere* **13** : e3934.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. and Polosukhin, I. (2017). Attention is all you need, *CoRR* **abs/1706.03762**.
**URL:** *http ://arxiv.org/abs/1706.03762*

Venkatanath, F., Praneeth, Bh., M. C., Channappayya, S. S. and Medasani, S. S. (2015). Blind image quality evaluation using perception based features, *2015 Twenty First National Conference on Communications (NCC)* pp. 1–6.
**URL:** *https ://api.semanticscholar.org/CorpusID :6917137*

Yadav, A., Upadhyay, M. and Singh, G. (2021). Underwater image enhancement using convolutional neural network.

Zhang, K., Zuo, W., Chen, Y., Meng, D. and Zhang, L. (2016). Beyond a gaussian denoiser : Residual learning of deep CNN for image denoising, *CoRR* **abs/1608.03981**.
**URL:** *http ://arxiv.org/abs/1608.03981*

Zhang, M., Li, Y. and Yu, W. (2024). Underwater image enhancement algorithm based on adversarial training, *Electronics* **13**(11) : 2184–2184.

Zhang, W., Dong, L., Zhang, T. and Xu, W. (2021). Enhancing underwater image via color correction and bi-interval contrast enhancement, *Signal Processing-image Communication* **90** : 116030–.

Zhao, S., Mei, X., Ye, X. and Guo, S. (2024). Msfe-uienet : A multi-scale feature extraction network for marine underwater image enhancement, *Journal of Marine Science and Engineering* **12**(9) : 1472–1472.