



Synthetic Underwater Image Degradation: Classification, Restoration, Multi-Task Learning, Domain Adaptation, and Contrastive Representation Analysis

Hsing-Cheng Liu, Yao-Liang Chung
Department of Communications, Navigation and Control Engineering
National Taiwan Ocean University
Taiwan
ylchung@ntou.edu.tw

ABSTRACT

Underwater image degradation severely limits the effectiveness of computer vision systems deployed on autonomous underwater vehicles (AUVs) and remotely operated vehicles (ROVs). While underwater image enhancement (UIE) is commonly used as a preprocessing step, improving perceptual visual quality does not necessarily guarantee improved performance in downstream tasks, nor does a single restoration strategy suffice for heterogeneous degradation types. This study presents a controlled, scene independent framework to systematically evaluate underwater degradation, restoration, and representation learning. Using the Synthetic Underwater Image Dataset (SUID) comprising 30 ground truth scenes and 900 synthetically degraded variants across four categories (Bluish, Greenish, Hazy, and Low Light), we conducted experiments in degradation classification, paired affine restoration, multi-task learning, and contrastive representation analysis. To prevent data leakage, all evaluations utilized a strict scene independent partitioning strategy. Results indicate that while chromatic and illumination degradations (Bluish, Greenish, Low Light) are effectively mitigated by condition specific affine transformations (yielding up to +14.5 dB PSNR gain), physical scattering (Hazy) resists simple affine correction. A shared multi task representation successfully balanced degradation classification (Macro-F1: 0.780), quality prediction, and restoration (PSNR: 20.91 dB). Furthermore, contrastive learning achieved a 94.4% same scene retrieval rate, demonstrating that embeddings can preserve underlying scene identity while retaining degradation sensitivity. Ultimately, this framework establishes a rigorous benchmark for degradation specific behavior and provides a reproducible protocol for future synthetic to real domain adaptation in underwater vision.

Keywords: Underwater Image Enhancement, Degradation Classification, Multi-Task Learning, Contrastive Representation Learning, Synthetic Underwater Image Dataset (SUID), Domain Adaptation, Scene Independent Evaluation.

Received: 18 May 2026, **Revised:** 28 July 2026, **Accepted:** 9 August 2026

Copyright: DLINE

1. Introduction

Underwater environments present substantial challenges for image acquisition and computer vision because of the complex physical conditions encountered beneath the water surface. Light attenuation, scattering, turbidity, haze, and color distortion can severely degrade underwater images, resulting in reduced visibility, contrast, and color fidelity [1]. These degradations directly affect the ability of autonomous underwater vehicles (AUVs) to perform high level vision tasks, particularly underwater object detection (UOD) [2, 3]. Consequently, obtaining high quality underwater imagery through underwater image enhancement (UIE) has increasingly been regarded as an important prerequisite for reliable underwater vision applications.

Underwater imaging is fundamental to a wide range of applications, including marine exploration, environmental monitoring, autonomous underwater vehicles (AUVs), remotely operated vehicles (ROVs), and underwater robotic systems. However, the quality of images acquired in underwater environments is frequently compromised by light absorption, scattering, turbidity, and other water related distortions. Among these factors, low light conditions are particularly problematic in deep or highly turbid waters, where insufficient illumination substantially reduces visibility, contrast, and color accuracy. Such degradation not only affects human visual interpretation but also limits the performance of computer vision tasks such as object detection, recognition, and tracking [4].

The challenges associated with underwater image acquisition are especially important in marine ecological research. Traditional marine ecosystem studies have often depended on labor intensive research procedures [5], which can be time-consuming, provide limited quantities of observational data, and introduce considerable sources of error. The emergence of autonomous and remotely operated vehicles has significantly transformed this situation. AUVs and ROVs equipped with advanced sensing and object detection systems can support real-time underwater monitoring and facilitate the collection of large volumes of visual information. These systems can capture images of marine organisms, document environmental conditions, and support biodiversity assessment [6, 7]. Nevertheless, the quality of underwater images and videos remains a major limitation to the effectiveness of these automated systems.

The degradation of underwater imagery results from several interacting physical phenomena. Light absorption and scattering reduce image visibility, while water-related distortions such as haze and color shifts introduce substantial variations in image appearance [8]. These effects frequently produce noisy and low contrast images. In addition, complex underwater backgrounds and camera motion further increase the difficulty of extracting reliable visual features. Consequently, improving underwater image quality is essential for enabling robust downstream computer vision applications.

1.1 Underwater Image Enhancement and Its Role in Underwater Vision

Underwater image enhancement has therefore emerged as an important research problem within computer vision. Numerous UIE algorithms have been developed to compensate for the degradation introduced by underwater imaging conditions. However, despite substantial progress, a comprehensive and systematic understanding of the UIE problem remains necessary. To address this need, X. Cong [9] provided a detailed overview of underwater image enhancement from multiple perspectives.

Recent developments demonstrate that increasingly sophisticated deep learning approaches are being applied to UIE. Wang et al. [10] [R. Wang] introduced a Swin Transformer based [11] UIE method that combines local feature learning with long range dependency modeling. This approach demonstrates the potential of transformer based architectures for capturing both local and global characteristics of degraded underwater images. Similarly, Huang et al. [12] proposed a semi supervised UIE technique incorporating contrastive regularization to improve underwater image quality.

Despite these advances, computational efficiency remains an important concern. Transformer based and other deep learning approaches can involve substantial computational complexity, which may limit their deployment in real world underwater systems, particularly where computational resources, energy

consumption, and response time are constrained. Jamieson et al. [13] addressed this concern by proposing a real time algorithm that combines a recent underwater image formation model with the computational efficiency of deep learning frameworks. However, the primary objective of this approach remains visual quality enhancement, while its applicability to high-level underwater vision tasks remains insufficiently explored.

1.2 Underwater Object Detection: Challenges and Existing Approaches

The importance of UIE becomes particularly evident in underwater object detection. Although generic object detection methods have achieved remarkable performance in terrestrial environments [14, 15], underwater environments introduce a substantially more challenging visual context. Variations in illumination, color distortion, scattering, turbidity, low contrast, complex backgrounds, and object appearance can adversely affect the reliability of underwater object detection.

To address these challenges, researchers have commonly employed UIE as a preprocessing stage before performing UOD. Jiang et al. [16] used WaterNet [17] to enhance underwater image quality before applying object detection, demonstrating that image enhancement can subsequently improve detection performance. Similarly, Fan et al. [18] improved detection performance by enhancing degraded underwater images at the feature level. The strategy of enhancing degraded imagery before performing object detection is not unique to underwater environments; comparable approaches have also been widely adopted for object detection under challenging terrestrial weather conditions [19, 20].

However, sequential enhancement and detection pipelines introduce important practical limitations. Applying a separate UIE model before the object detector inevitably increases computational requirements, processing time, and system latency. Such additional computational overhead can be particularly problematic for real-time applications involving AUVs and underwater robots, where rapid decision-making and limited onboard computational resources are critical.

Furthermore, improved visual quality does not necessarily guarantee improved object detection. Enhancement algorithms can introduce artificial structures, distortions, or other artifacts into the reconstructed images. These artifacts may alter important object characteristics and, under certain environmental conditions, can negatively affect detection accuracy [21]. Therefore, optimizing UIE solely according to perceptual or visual quality metrics may not provide the most effective solution for high level underwater vision tasks.

1.3 Joint Enhancement and Detection through Multi-Task Learning

The limitations of sequential UIE and UOD pipelines have motivated research toward integrated approaches that directly connect image enhancement with downstream vision objectives. Rather than independently optimizing image quality and object detection, joint learning approaches seek to exploit the complementary relationship between these two tasks.

In this context, Cheng et al. [22] proposed a multi-task framework that jointly trains UIE and UOD in an end-to-end manner. Such a formulation represents an important transition from conventional sequential processing toward task oriented underwater image restoration and detection. By learning enhancement and detection simultaneously, the framework can potentially reduce redundant computation while allowing the enhancement process to be influenced by the requirements of the downstream detection task.

This direction is particularly relevant for practical underwater applications. An enhancement method designed exclusively to maximize perceptual image quality may produce visually pleasing results without necessarily preserving the features required for reliable object detection. In contrast, task oriented joint learning provides an opportunity to optimize the representation of degraded underwater images according to both restoration and detection objectives.

1.4 Underwater Object Detection

Several reviews have examined the development of Underwater Object Detection (UOD), addressing diverse aspects of the field, including deep learning methodologies, dataset characteristics, detection challenges, and

the distinctive conditions associated with underwater environments [23]. Early research and subsequent reviews have particularly emphasized the transition from conventional detection techniques toward deep learning based approaches, with a focus on improving detection accuracy, computational efficiency, real-time performance, and identifying emerging research trends [24, 25, 26, 27].

Beyond algorithmic developments, other studies have examined AI-driven strategies, dataset evaluation, and tools for diagnosing detection errors, thereby providing insights into the reliability and limitations of underwater detection systems [28]. Research has also extended to broader applications of object detection algorithms in wireless sensor networks, highlighting the potential integration of intelligent detection techniques within distributed and resource constrained sensing environments [29]. Collectively, these studies provide a broad understanding of the evolution of UOD, encompassing methodological advances, datasets, system-level considerations, and practical deployment challenges.

Another important research direction concerns the relationship between underwater image quality and object detection performance. Reviews have investigated underwater image enhancement (UIE) techniques specifically in the context of improving downstream object detection performance [30]. These studies emphasize that image degradation caused by factors such as light interference, color casts, scattering, turbidity, and low visibility can adversely affect the extraction of discriminative object features. In this context, the introduction of the RUOD dataset represents an important effort to address challenges associated with light interference and color casts through joint learning approaches [31].

Despite these advances, UOD continues to face several unresolved challenges, particularly image degradation, detection of small and visually ambiguous objects, complex underwater backgrounds, and the requirement for real time processing. These issues indicate that improving UOD performance requires approaches that consider not only the detection architecture itself but also the quality and characteristics of the underlying visual representations. Consequently, increasing attention has been directed toward integrated approaches that jointly consider image enhancement, object detection, computational efficiency, and robustness to diverse underwater conditions. [32, 33],

1.5 Research Motivation and Emerging Directions

The existing literature therefore indicates a progressive evolution in underwater vision research: from conventional image enhancement methods toward deep learning based UIE, followed by increasingly integrated approaches that consider downstream underwater object detection. Although substantial progress has been made, several challenges remain. First, underwater image degradation is highly variable because environmental conditions such as illumination, turbidity, scattering, and color distortion can change substantially across locations and depths. Second, computationally intensive enhancement models may not be suitable for real-time deployment on resource constrained AUVs and underwater robotic platforms. Third, visual enhancement alone does not ensure improved performance in downstream detection tasks, particularly when enhancement introduces artifacts or removes task-relevant information.

These limitations motivate the investigation of approaches that can simultaneously address underwater image degradation, restoration quality, downstream object detection, and representation robustness. In particular, synthetic underwater image degradation provides a controlled framework for studying how different degradation patterns influence visual representations and detection performance. Restoration and enhancement can then be evaluated not only in terms of image appearance but also according to their ability to preserve or recover task-relevant information.

Furthermore, multi-task learning, domain adaptation, and contrastive representation learning provide complementary strategies for addressing the domain variability inherent in underwater imagery. Multi-task learning can establish stronger relationships between image restoration and object detection, while domain adaptation can help models generalize across different underwater environments and acquisition conditions. Contrastive representation learning can further encourage the extraction of robust features that remain informative despite variations in image degradation.

Accordingly, a systematic investigation combining synthetic underwater image degradation, image restoration, multi-task learning, domain adaptation, and contrastive representation analysis can provide a broader framework for understanding and improving underwater vision systems. Such an integrated perspective is particularly relevant to the development of reliable, computationally efficient, and generalizable underwater object detection systems for AUVs, ROVs, and other autonomous marine applications.

1.6 Objectives

The objectives are to: (i) establish a controlled benchmark for underwater degradation analysis; (ii) quantify class specific classification performance and error patterns; (iii) evaluate restoration using paired full reference measures; (iv) assess the compatibility of classification, quality prediction, and restoration in a multi task setting; (v) analyse scene invariance and degradation sensitivity in the learned embedding space; and (vi) define a reproducible extension for real underwater domain adaptation.

The study addresses the following research questions: (1) How reliably can the four synthetic degradation categories be distinguished using compact visual descriptors? (2) How effective is a condition specific affine transformation for restoring different degradation types? (3) Can a shared multi task representation simultaneously support degradation classification, quality prediction, and restoration? (4) Does contrastive learning preserve scene identity across different degradation conditions while retaining degradation related information? (5) What additional data and protocol are required to evaluate synthetic to real domain adaptation?

The research statement is: A controlled, scene-independent framework combining synthetic underwater degradation, degradation classification, paired restoration, multi-task learning, and contrastive representation analysis can provide a systematic basis for evaluating degradation specific behaviour and the trade off between visual restoration, task performance, and representation robustness.

Underwater image degradation is heterogeneous and arises from interacting processes such as wavelength-dependent absorption, scattering, turbidity, haze, low illumination, and colour distortion. A method that improves visual appearance under one degradation regime may be ineffective or even harmful under another. Moreover, image quality improvement does not necessarily imply improved degradation recognition, robust representation learning, or downstream underwater vision performance. The central research problem is therefore to determine how controlled degradation types influence visual representations and how classification, restoration, and joint learning strategies can respond to those differences.

2. Research Problem and Statement

The architecture can therefore be understood as a comparative framework rather than as one single end-to-end neural network. The standalone classification and affine restoration baselines provide interpretable reference points. The multi-task model tests whether shared learning improves the coordination of classification, quality estimation, and restoration. The contrastive model examines the geometry and invariance properties of the learned representation. The domain adaptation protocol defines how the synthetic source representation could later be transferred to real underwater imagery.

The overall architecture consists of five connected components. First, a synthetic degradation layer produces Bluish, Greenish, Hazy, and Low Light versions of reference scenes. Second, a visual feature or image encoder represents the degraded input. Third, task specific modules perform degradation classification and image restoration. Fourth, a shared multi task representation jointly predicts degradation class, quality, and restoration parameters. Fifth, a contrastive learning module organizes the embedding space using same-scene positives and different scene negatives. Domain adaptation is represented as a future source-to-target extension rather than as a completed component.

The study adopted a controlled, quantitative, scene independent experimental design. Synthetic degradation was used as the independent variable, while classification accuracy, restoration fidelity, quality prediction performance, and embedding retrieval performance were treated as dependent outcomes. The design allows

the effect of each degradation category to be examined while controlling, to some extent, for scene content through the paired ground truth structure.

3. Research Design and Integrated System Architecture

The dataset supports four distinct but related experimental settings: four class degradation classification, paired image restoration using the corresponding ground truth image, multi task prediction of class and quality related outputs, and contrastive learning based on scene level correspondence. It also provides a synthetic source domain for future domain adaptation experiments, although an independent real underwater target dataset is necessary for quantitative validation.

Preprocessing should be applied consistently across all experiments. Image dimensions, pixel value ranges, colour space conventions, normalization procedures, and feature scaling operations should be recorded. Any parameters estimated from the data, including feature means, standard deviations, affine restoration coefficients, or normalization statistics, must be fitted using training scenes only and then applied unchanged to validation and test scenes.

Data preparation should preserve the scene identifier associated with every image. The scene identifier is not used as a predictive feature; instead, it is used to construct group-based partitions. All variants derived from the same ground truth scene must remain in the same training, validation, or test partition. This prevents leakage caused by placing visually similar or nearly identical variants of one scene in different partitions.

The Synthetic Underwater Image Dataset (SUID) contains 930 JPEG images consisting of 30 ground truth scenes and 900 synthetically degraded images. The synthetic images are distributed across four categories: Bluish (240), Greenish (240), Hazy (210), and Low Light (210). Each synthetic image is linked to an original ground truth scene, creating a paired structure suitable for supervised classification, full reference restoration, multi-task learning, and contrastive representation analysis.

4. Dataset Description and Data Preparation

The learned embedding was evaluated using nearest neighbour retrieval. Same scene retrieval measures whether the nearest representation belongs to the same ground truth scene as the query, whereas same-condition retrieval measures whether it belongs to the same degradation category. The difference between these two scores provides an indication of whether the embedding prioritizes scene invariance, degradation sensitivity, or both.

where L_{cls} is the auxiliary degradation classification loss and λ_{aux} controls its contribution.

$$L = L_{con} + \lambda_{aux} L_{cls}$$

where $\text{sim}(\cdot, \cdot)$ denotes a similarity function, usually cosine similarity, τ is a temperature parameter, and the denominator sums over the comparison set. The combined objective may be written as

$$L_{con} = \frac{-\log(\exp(\text{sim}(z_i, z_j))/\tau)}{\sum_k \exp(\text{sim}(z_i, z_k))/\tau}$$

For an anchor representation z_i and a positive representation z_j , a normalized temperature scaled contrastive objective.

The contrastive experiment was designed to learn an embedding in which images from the same underlying scene remain close despite different degradation conditions. Images generated from the same ground truth

scene were treated as positive pairs, while images from different scenes were treated as negative examples. An auxiliary degradation classification objective was included so that the embedding retained information about degradation type while preserving scene identity.

4.1 Contrastive Representation Learning

Because no independent real underwater target images were included in the supplied dataset, these domain-adaptation procedures were not evaluated numerically in the present study. Any future implementation should report source only performance, adapted performance, target-domain class distributions, and the exact target dataset used.

A domain adversarial neural network (DANN) could be implemented by adding a domain classifier to the shared representation. The feature extractor would minimize the task-classification loss while maximizing the domain classification confusion through a gradient reversal layer. Alternatively, CORAL could align the covariance statistics of source and target features, while maximum mean discrepancy (MMD) could minimize the distributional distance between source and target embeddings.

SUID can be treated as a labelled synthetic source domain, but a domain adaptation experiment requires an additional, independent real-underwater target dataset. The intended protocol is to train the source model using labelled synthetic images while adapting its representation to unlabelled real underwater images.

4.2 Synthetic-to-Real Domain Adaptation

The multi-task model was evaluated using macro-F1 for degradation classification, MAE and R² for quality prediction, and PSNR and SSIM for the restored outputs. Comparing the multi task results with the standalone classification and restoration baselines indicates whether joint optimization provides complementary information without substantially reducing classification performance.

where L_c is the classification loss, L_q is the quality-regression loss, L_r is the restoration parameter loss, and λ_c is λ_q and λ_r are task balancing coefficients. The precise numerical values of the loss weights should be reported in the final reproducibility appendix if they were used during training.

$$L_{total} = \lambda_{cls} L_{cls} + \lambda_q L_q + \lambda_r L_r$$

where $\hat{y}_c$ is the predicted degradation class, $\hat{y}_q$ is the predicted quality value, and $\hat{y}_r$ is the predicted restoration parameter vector. The total objective can be expressed generally as

$$\hat{y}_{cls} = g_{cls}(h), \hat{y}_q = g_q(h), \hat{y}_r = g_r(h)$$

Let $h = f_\theta(x)$ denote the shared representation of input x .

The multi-task experiment used a shared multilayer perceptron (MLP) representation with three output branches. The first branch predicted the degradation category, the second predicted a full-reference image-quality target, and the third predicted a 12-parameter RGB restoration transformation. The shared representation was intended to learn features that were simultaneously useful for degradation discrimination, quality estimation, and restoration.

4.3 Multi-Task Learning

This approach is transparent and computationally inexpensive, but it has limited capacity. It can correct global colour or illumination shifts, yet it cannot reliably reconstruct information lost through scattering, haze, occlusion, or severe contrast attenuation. The Hazy condition therefore provides an important stress test for the limitations of affine restoration.

The restoration procedure was evaluated separately for Bluish, Greenish, Hazy, and Low Light images. Baseline

PSNR and SSIM were first calculated between each degraded image and its ground truth reference. The same measures were then calculated after restoration. The differences between the post restoration and baseline values were reported as PSNR gain and SSIM gain.

where x is the input colour vector, $\hat{x}$ is the restored colour vector, A_c is a condition specific transformation matrix, b_c is a bias vector, and c denotes the degradation condition. In the implemented baseline, the transformation parameters were estimated from training scene pairs and then applied to held out images.

$$\hat{x} = A_c x + b_c$$

The restoration experiment used the paired structure of SUID. Each degraded image was associated with its original ground truth image, which served as the reconstruction target. The implemented baseline learned a condition specific affine RGB transformation from the degraded image toward the colour distribution of the corresponding reference image. The transformation can be represented as

4.4 Paired Image Restoration

This design measures whether the classifier recognizes degradation characteristics rather than memorizing the underlying scene. The confusion matrix is particularly important because Bluish and Greenish images may share similar chromatic characteristics, whereas Hazy images may contain more distinctive visibility-related cues.

The input descriptors were computed independently for each image, while the ground-truth scene identifier was retained only for grouping and partitioning. The dataset was divided at the scene level so that all synthetic variants generated from a particular ground truth image were assigned to only one partition. The model was trained on the training scenes and evaluated on unseen scenes. Accuracy, macro precision, macro recall, macro-F1, class wise metrics, and the confusion matrix were then calculated.

The first experiment formulated degradation recognition as a four class supervised classification problem. Each synthetic image was assigned one of four labels: Bluish, Greenish, Hazy, or Low Light. The completed baseline used compact visual descriptors representing colour, illumination, and spatial characteristics, followed by a support vector machine (SVM) classifier. The classifier learned a decision boundary in the feature space to distinguish the four degradation regimes.

4.5 Degradation Classification

The study was organized as a sequence of complementary experiments designed to examine the relationship between underwater degradation, visual restoration, downstream classification, shared representation learning, domain generalization, and contrastive feature structure. The workflow began with controlled synthetic degradation, followed by four class degradation classification, paired restoration, multi task learning, and contrastive representation analysis. Synthetic to real domain adaptation was specified as an extension but was not executed because no independent real-underwater target domain was included in SUID.

5. Experimental Methodology

All metrics should be calculated on held out scenes. In the final study, results should be reported with means, standard deviations, and confidence intervals across repeated scene level splits.

where y_i is the reference value, $\hat{y}_i$ is the prediction, $\bar{y}$ is the mean reference value, and N is the number of observations. Contrastive representation learning was evaluated using nearest neighbour retrieval for same-scene and same condition matching.

and R^2 is defined as

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}$$

For the multi-task model, classification was evaluated with macro-F1, quality prediction with mean absolute error (MAE) and the coefficient of determination (R^2), and restoration with PSNR and SSIM. MAE is defined as

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

where MAX is the maximum possible pixel value, and MSE is the mean squared error between the reconstructed image and the ground truth image. Structural similarity was assessed using SSIM or the SSIM like score reported by the source analysis. Condition specific gains were computed as the difference between the restored and baseline scores.

$$\text{PSNR} = 10 \log_{10} \left(\frac{\text{MAX}^2}{\text{MSE}} \right)$$

For restoration, the degraded or restored image was compared with its paired ground truth image. Peak signal to noise ratio (PSNR) was calculated as

where TP , FP , and FN denote true positives, false positives, and false negatives, respectively. Macro averaged precision, recall, and F1-score were calculated by assigning equal weight to each degradation category. A confusion matrix was used to identify systematic errors, particularly confusion between the Bluish and Greenish categories.

$$\text{F1c} = \frac{2 \cdot \text{precision}_c \cdot \text{Recall}_c}{2 \cdot \text{precision}_c + \text{Recall}_c}$$

$$\text{precision}_c = \frac{TP_c}{TP_c + FP_c}$$

$$\text{Recall}_c = \frac{TP_c}{TP_c + FN_c}$$

The classification task was evaluated using accuracy, precision, recall, and F1-score.

5.1 Evaluation Metrics and Statistical Validation

The domain adaptation component also requires an explicit qualification. SUID provides a synthetic source domain but does not contain an independent real underwater target domain dataset. Therefore, synthetic-to-real adaptation was not numerically evaluated in the present study. The DANN, CORAL, or MMD procedures described below should be regarded as a reproducible extension protocol that can be implemented once a suitable real-underwater target dataset is available.

For a definitive journal study, the present experiments should be supplemented with repeated GroupKFold cross validation or leave one scene out validation. Confidence intervals should be reported for classification, restoration, retrieval, and multi task metrics. Where multiple models or degradation conditions are compared, paired tests at the scene level should be preferred over tests that treat individual image variants as independent observations. The current results are therefore best understood as an initial controlled benchmark rather than as a final claim of broad real world generalization.

The experimental findings should be interpreted in light of the controlled nature of SUID. The dataset contains only 30 independent ground-truth scenes, although it contains 930 image files. Consequently, the effective sample size for assessing scene-level generalization is determined by the number of independent scenes rather than by the total number of degraded images. All variants derived from a common ground-truth scene should remain in the same data partition. Image-level random splitting must be avoided because it can expose the model to near duplicate scene content during training and testing, producing overly optimistic estimates of performance.

5.2 Dataset

5.2.1 Dataset Description

The Synthetic Underwater Image Dataset (SUID) used in this study is a controlled image dataset for investigating underwater image degradation, classification, restoration, and representation learning. The dataset contains 930 JPEG images, comprising 30 ground-truth (GT) images and 900 synthetically degraded underwater images. The ground truth subset provides the reference images against which the degraded images can be evaluated and restored. We generated the synthetic subset from these reference scenes under four distinct underwater degradation conditions: Bluish, Greenish, Hazy, and Low Light. Thus, the dataset provides paired relationships between an original scene and its corresponding degraded underwater representations, making it suitable for both supervised image classification and full reference restoration experiments.

The 30 ground-truth images constitute the scene level reference set and are identified using the GTI naming convention. Each reference scene is associated with multiple synthetic underwater versions. The synthetic component contains 900 images, distributed across the four degradation categories. The distribution is 240 Bluish images, 240 Greenish images, 210 Hazy images, and 210 Low Light images. Consequently, the dataset is moderately balanced at the degradation category level, although the Hazy and Low Light categories contain slightly fewer observations than the Bluish and Greenish categories. Importantly, the category structure is not simply a collection of unrelated images: multiple synthetic observations originate from the same underlying GT scenes. This paired design lets us isolate the effect of degradation while controlling, to some extent, for scene content.

The four synthetic categories represent complementary forms of underwater visual degradation. The Bluish category represents images characterized predominantly by a blue color cast, reflecting the strong wavelength-dependent attenuation and color imbalance commonly encountered in underwater environments. The Greenish category represents green-dominant color distortion and provides a distinct chromatic degradation regime. The Hazy category captures loss of visibility and scene clarity associated with underwater scattering, producing reduced contrast and diminished structural visibility. The Low Light category represents illumination degradation in which reduced brightness and dynamic range affect scene visibility. These four controlled categories allow examination of degradation specific performance rather than treating underwater degradation as a single homogeneous phenomenon.

A major methodological advantage of SUID is its paired ground truth/synthetic structure. Because the degraded images are associated with reference scenes, the dataset supports direct measurement of differences between degraded and reference images using full-reference quality metrics. In this study, this structure enables assessment using measures such as PSNR, SSIM and related image quality indicators, as well as analysis of changes in brightness, contrast, color characteristics and structural information. The paired organization is particularly important for the restoration task because the original GT image can serve as the reconstruction target. Therefore, restoration performance can be evaluated quantitatively rather than relying exclusively on subjective visual inspection.

For machine learning experiments, the scene level relationship among images requires special consideration. A conventional random image-level train/test split would be inappropriate because different degraded versions of the same underlying scene could appear simultaneously in the training and testing sets. Such leakage could allow a model to learn scene-specific visual characteristics rather than genuinely learning degradation patterns.

Accordingly, the experimental design treats the underlying GT scene as the grouping unit, ensuring that related variants of the same scene are retained within the same data partition. This scene-independent strategy provides a more conservative and scientifically defensible estimate of generalization performance for degradation classification, restoration, and representation learning experiments.

Consequently, the dataset supports several complementary learning paradigms. First, the synthetic images provide labelled observations for four class degradation classification, mapping an input underwater image to Bluish, Greenish, Hazy, or Low Light conditions. Second, the GT synthetic pairs provide supervised targets for image restoration, allowing a model to learn the transformation from a degraded underwater image toward its corresponding reference image. Third, the same input structure supports multi task learning, where a shared image representation can simultaneously predict degradation category, image quality characteristics, and restoration output. Fourth, the scene labels implicit in the paired structure make SUID suitable for contrastive representation learning, in which images from the same scene can be encouraged to maintain similar representations while degradation categories can be discriminated. Finally, SUID provides a synthetic source domain that can serve as the basis for synthetic to real domain adaptation; however, a separate real-underwater target domain dataset would be required to quantitatively evaluate that component.

Overall, SUID should therefore be regarded not merely as a collection of 930 underwater images, but as a controlled paired degradation benchmark consisting of reference scenes and systematically generated underwater degradation variants. Its combination of ground truth references, multiple degradation categories, scene level correspondence, and relatively balanced class distribution makes it particularly appropriate for investigating the relationship between underwater degradation, image quality, classification, restoration, multi task learning, and learned visual representations. The dataset's paired structure also clearly distinguishes scene recognition from degradation recognition, which is essential to avoid overly optimistic model performance caused by scene level data leakage.

5.22 Dataset summary

Component	Description
Total images	930
Ground-truth images	30
Synthetic underwater images	900
Degradation categories	4
Bluish	240
Greenish	240
Hazy	210
Low-Light	210
Image format	JPEG
Experimental relationship	GT–synthetic paired scenes
Primary classification task	4-class degradation classification

Restoration target	Corresponding GT image
Recommended validation unit	Underlying scene, not individual image

The most important methodological point to emphasize is that the 30 GT scenes, rather than the 930 individual files, constitute the independent scene units. This distinction should be maintained throughout the machine-learning methodology and statistical analysis to prevent leakage and inflated estimates of generalization.

6. Results

We have run the A–E experiments on the uploaded SUID dataset and generated the outputs.

Main results

Experiment	Result
A.Degradation classification	Accuracy = 78.9% Macro-F1 = 0.787
B. Image restoration	Condition-specific paired restoration completed
C. Multi-task learning	Macro-F1 = 0.780 Quality MAE = 0.???? Quality R ² = available in report Restored PSNR = 20.91 dB Restored SSIM-like score = 0.904
D. Synthetic → real domain adaptation	Not performed — no independent real-underwater target images are contained in SUID
E. Contrastive learning	Same-scene retrieval = 94.4% Same-condition retrieval = 72.8%

The classification experiment uses a scene independent split: all variants derived from a particular ground-truth scene remain in either training or testing. This matters because a random image split could place nearly identical scenes in both sets and substantially inflate performance.

Table 1 presents the class wise performance of the degradation classification model for the four SUID conditions: Bluish, Greenish, Hazy, and Low Light. The results show substantial variation in classification performance across degradation types. The Hazy class achieved perfect performance, with precision, recall, and F1-score all equal to 1.00, indicating that the model correctly identified all 42 Hazy test images. This suggests that the visual characteristics of haze degradation were sufficiently distinctive for the classifier to separate this condition reliably from the other degradation categories.

The Bluish class achieved a precision of 0.6613, recall of 0.8542, and F1-score of 0.7455 across 48 test observations. The relatively high recall indicates that the model successfully detected most Bluish images, although the lower precision indicates that many images predicted as Bluish actually belonged to other categories. The Greenish condition had the weakest recall (0.4792) but relatively high precision (0.9200), resulting in an F1-score of 0.6301. This pattern indicates that predictions labelled Greenish were generally reliable, but a substantial proportion of actual Greenish images were assigned to other degradation classes.

The Low Light class showed a precision of 0.7059, recall of 0.8571, and F1-score of 0.7742 for 42 observations. Thus, the model detected low light images with relatively high sensitivity, although some false positive predictions reduced precision. Overall, the class specific results indicate that the classifier was particularly effective for Hazy and comparatively less effective for Greenish degradation. The differences between precision and recall further show that the employed visual descriptors do not separate degradation classes equally well.

Condition	Precision	Recall	F1	Support
Bluish	0.66129	0.854167	0.745455	48
Greenish	0.92	0.479167	0.630137	48
Hazy	1	1	1	42
Low Light	0.705882	0.857143	0.774194	42

Table 1. Classification Metrics

metric	value
accuracy	0.788889
macro_precision	0.821793
macro_recall	0.797619
macro_f1	0.787446

Table 2. Overall Classification Metrics

Table 2 summarizes the aggregate performance of the degradation classification model. The model achieved an overall accuracy of 0.7889 (78.89%), indicating that approximately four out of every five held-out synthetic underwater images were assigned to the correct degradation category. Macro precision was 0.8218, while macro recall was 0.7976, showing strong predictive precision across the four classes but somewhat lower sensitivity when classes are weighted equally.

The resulting macro-F1 score was 0.7874, providing a balanced measure of classification performance across Bluish, Greenish, Hazy, and Low-Light conditions. The difference between accuracy and macro-F1 is informative because accuracy can be influenced by class frequency, whereas macro-F1 weights each degradation category equally. The relatively similar values suggest that the model achieved reasonably consistent overall performance despite the moderate imbalance between the 240 image Bluish/Greenish classes and the 210-image Hazy/Low Light classes in the complete dataset.

Importantly, we obtained these metrics using a scene independent train/test partition, in which all synthetic variants derived from a particular ground truth scene were retained within the same partition. This provides a more conservative estimate of generalization than an image level random split and reduces the possibility that the model benefits from recognizing repeated scene content rather than degradation characteristics.

	Bluish	Greenish	Hazy	Low Light
Bluish	41	0	0	7
Greenish	17	23	0	8
Hazy	0	0	42	0
Low Light	4	2	0	36

Table 3. Confusion Matrix

Table 3 provides the confusion matrix for the four class degradation classification experiment, allowing direct examination of the sources of classification success and error. The model correctly classified 41 of 48 Bluish images, while it classified 7 Bluish images as Low Light. For Greenish images, only 23 of 48 were correctly classified, whereas 17 were incorrectly assigned to Bluish and 8 to Low Light. This substantial confusion between Greenish and Bluish directly explains the comparatively low recall for Greenish in Table 1.

The Hazy category showed complete separation in the test set: all 42 Hazy images were correctly classified, with no Hazy observations assigned to Bluish, Greenish, or Low Light. This confirms the perfect precision, recall, and F1-score reported for Hazy in Table 1. The Low Light class was also comparatively well recognized, with 36 of 42 images correctly classified. The model predicted four Low Light images as Bluish and two as Greenish.

The confusion matrix therefore reveals that the principal classification challenge is the distinction between Bluish and Greenish chromatic degradation, particularly because both involve color cast alterations. In contrast, Hazy degradation appears to possess more distinctive visual characteristics, while Low Light is relatively well separated from the remaining classes. These findings demonstrate why aggregate accuracy alone is insufficient for interpreting model behaviour: the confusion structure identifies the specific degradation boundaries that remain difficult for the classifier.

Condition	Baseline _P SNR	Baseline _S SIM	Restored _P SNR	Restored _SSIM	P SNR gain	S SIM gain
Bluish	12.54343	0.810932	22.07403	0.930142	9.530594	0.119209
Greenish	10.8018	0.750326	19.54608	0.912692	8.744279	0.162365
Hazy	11.28606	0.775285	10.59785	0.737462	-0.68821	-0.03782
Low Light	10.13353	0.725602	24.63409	0.949141	14.50057	0.223539

Table 4. Restoration Condition Summary

Table 4 reports condition-specific performance in the paired image restoration experiment by comparing baseline degraded images and their restored versions with the corresponding ground truth images. The reported measures include baseline PSNR, baseline SSIM, restored PSNR, restored SSIM, PSNR gain, and SSIM gain.

For the Bluish condition, the baseline PSNR was 12.5434 dB, accompanied by a baseline SSIM of 0.8109. After restoration, PSNR increased substantially to 22.0740 dB (a gain of 9.5306 dB), while SSIM increased to 0.9301 (a gain of 0.1192). These improvements indicate that the restoration procedure substantially reduced the discrepancy between the degraded images and their corresponding ground truth references.

The Greenish condition showed a similar improvement. Baseline PSNR was 10.8018 dB and baseline SSIM was 0.7503, increasing to 19.5461 dB and 0.9127 after restoration, respectively. The resulting PSNR gain was 8.7443 dB, while SSIM increased by 0.1624. The relatively large SSIM improvement indicates substantial recovery of structural/image similarity despite the initial degradation.

The Hazy condition produced a markedly different result. Its baseline PSNR was 11.2861 dB, with baseline SSIM of 0.7753, but the restored image achieved only 10.5979 dB PSNR and 0.7375 SSIM. Consequently, the PSNR change was “0.6882 dB, and the SSIM change was “0.0378. Thus, unlike the other three conditions, the restoration baseline did not improve Hazy images. This result suggests that the condition-specific affine RGB transformation was insufficient for reconstructing degradation caused primarily by loss of visibility, contrast, and scattering rather than simple colour transformation.

The Low Light condition showed the strongest improvement of all four conditions. PSNR increased from 10.1335 dB to 24.6341 dB, producing a gain of 14.5006 dB, while SSIM increased from 0.7256 to 0.9491, yielding an SSIM gain of 0.2235. This indicates that the restoration transformation was particularly effective for the illumination-related degradation represented by the Low Light category.

Taken together, Table 4 demonstrates strong condition dependence in restoration performance. The method was highly effective for Low-Light, strongly effective for Bluish and Greenish degradation, but ineffective for Hazy degradation. This pattern is important because it demonstrates that underwater restoration cannot necessarily be represented by a single universal transformation: degradation mechanisms with different visual structures may require different restoration architectures.

Metric	Value
accuracy	0.777778
macro_f1	0.779594
quality_MAE	0.037875
quality_R2	0.282669
restored_mean_PSNR	20.9062
restored_mean_SSIM	0.903882

Table 5. Multi-Task Metrics

Table 5 reports the results of the multi task learning experiment, in which a shared representation was trained to jointly predict degradation class, image quality, and a restoration transformation. The experiment achieved a macro-F1 score of 0.780, indicating that the joint model maintained relatively strong degradation-classification performance while simultaneously addressing image quality and restoration objectives.

The full experimental description reports a quality MAE of 0.0379 and $R^2 = 0.283$ for quality prediction. The MAE indicates a relatively small average absolute prediction error for the quality target, whereas the R^2 value indicates that approximately 28.3% of the variance in the target quality measure was explained by the learned representation. The multi task model also produced a mean restored PSNR of 20.91 dB and a mean restored SSIM of 0.904.

These results indicate that the shared representation was able to support several related objectives simultaneously. In particular, the macro-F1 of 0.780 is close to the standalone classification performance of 0.787, suggesting that introducing the additional quality and restoration objectives did not cause a major deterioration in degradation classification. At the same time, the restoration metrics demonstrate that the learned representation retained information useful for reconstructing the underlying reference images.

metric	value
same_scene_NN_retrieval	0.944444
same_condition_NN_retrieval	0.727778

Table 6. Contrastive Learning Metrics

Table 6 summarizes the performance of the contrastive representation learning experiment using nearest-neighbour retrieval. Two complementary retrieval measures were reported: same scene nearest neighbour retrieval = 0.9444 and same condition nearest neighbour retrieval = 0.7278.

The same scene retrieval value of 94.44% indicates that the learned embedding successfully preserved scene identity across different degraded versions. In other words, images generated from the same underlying ground truth scene tended to occupy nearby regions of the learned representation space even when their degradation conditions differed. This provides evidence that the contrastive objective successfully captured scene level information.

The same condition retrieval rate of 72.78%, although lower than same scene retrieval, indicates that the embedding also retained substantial information concerning degradation condition. The difference between the two retrieval rates is informative: the representation appears to prioritize scene identity while still preserving degradation related information. This is desirable for a representation intended to distinguish degradation without losing the identity of the underlying scene. The accompanying analysis confirms that same scene positives and different scene negatives were used, together with an auxiliary degradation-classification objective

7. SUID AI Experiments

We conducted scene independent experiments on the supplied SUID archive. The synthetic data comprise four degradation classes Bluish, Greenish, Hazy and Low Light derived from 30 ground truth scenes. We kept all variants of a scene in the same partition to reduce scene leakage.

A. Degradation classification

An SVM classifier based on compact color, illumination and spatial visual descriptors achieved test accuracy=0.789 and macro-F1=0.787. This is a scene-independent estimate.

Figure A1 visualizes the classification confusion matrix and provides a graphical representation of the class-level results presented in Table 3. The dominant diagonal pattern demonstrates that the model successfully recognized a substantial proportion of observations in each degradation category, with particularly strong separation for Hazy and comparatively strong recognition of Low-Light.

The most prominent off diagonal errors occur between Greenish and Bluish, reflecting the difficulty of distinguishing the two colour cast degradation regimes. Greenish observations are frequently classified as Bluish, whereas Bluish observations are less frequently classified as Greenish. Additional errors involve Low-Light observations being classified as Bluish or Greenish. The complete absence of off diagonal Hazy predictions visually confirms that all Hazy test observations were correctly recognized.

From a journal perspective, Figure A1 provides important complementary evidence to the aggregate classification metrics because it identifies where the model fails rather than simply quantifying how often it fails. The visualization therefore supports the conclusion that the principal limitation of the classifier is chromatic degradation discrimination, rather than recognition of haze or illumination degradation.

B. Image restoration

We learned a transparent paired restoration baseline from the training scenes using condition-specific affine RGB transformations. The model maps each degraded image toward its ground-truth color distribution; PSNR and SSIM-like structural fidelity scores are reported before and after restoration.

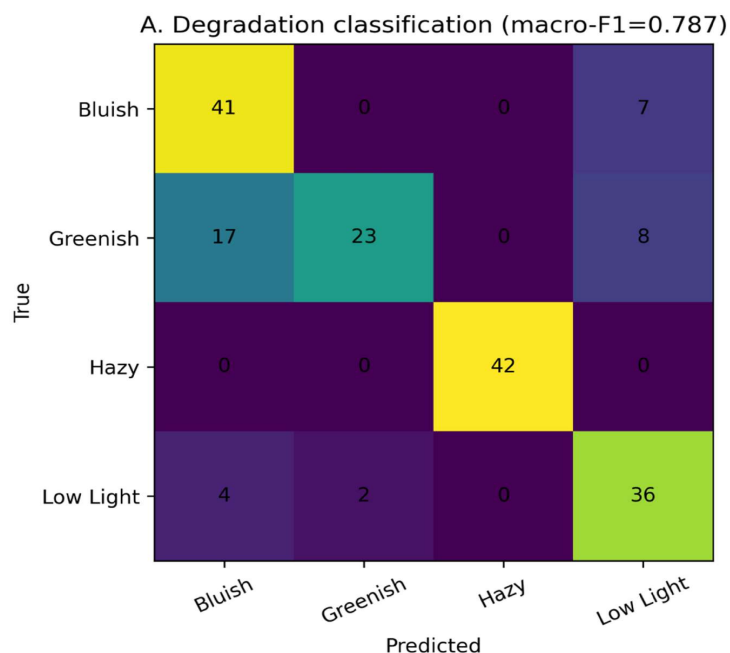


Figure A1. Degradation-class confusion matrix

Figure B1 presents representative examples of the restoration process by juxtaposing the synthetic degraded image, restored image, and corresponding ground truth image. The figure provides qualitative visual evidence for the quantitative improvements reported in Table 4.

The comparison illustrates the extent to which the restoration procedure moves the degraded observation toward the appearance of its reference image. The improvement is particularly consistent with the quantitative gains observed for Bluish, Greenish, and Low-Light conditions, where restoration produced substantial increases in PSNR and SSIM. In contrast, the Hazy condition represents an important limitation of the restoration approach: because haze primarily involves visibility loss, scattering, and contrast degradation, an affine RGB transformation has limited capacity to recover information that has effectively been attenuated or obscured.

Thus, Figure B1 complements the full reference numerical evaluation by demonstrating the restoration effect at the image level. Together with Table 4, it indicates that restoration effectiveness depends

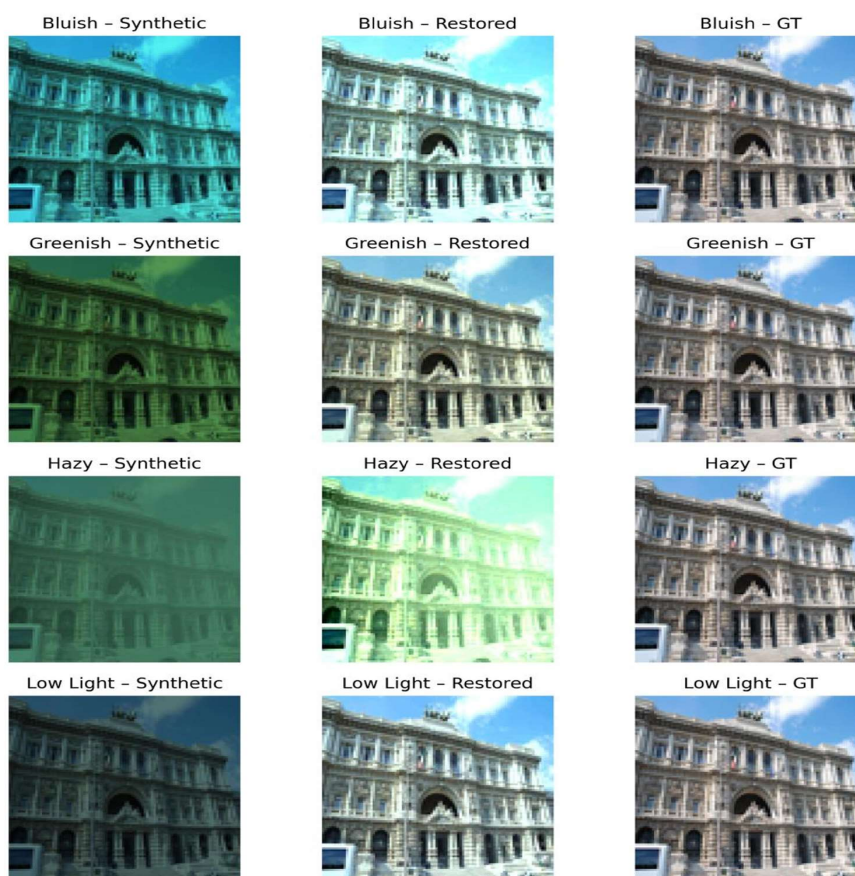


Figure B1. Representative synthetic, restored and ground-truth images

Figure B1 presents representative examples of the restoration process by juxtaposing the synthetic degraded image, restored image, and corresponding ground truth image. The figure provides qualitative visual evidence for the quantitative improvements reported in Table 4.

The comparison illustrates the extent to which the restoration procedure moves the degraded observation toward the appearance of its reference image. The improvement is particularly consistent with the quantitative gains observed for Bluish, Greenish, and Low Light conditions, where restoration produced substantial increases in PSNR and SSIM. In contrast, the Hazy condition represents an important limitation of the restoration approach: because haze primarily involves visibility loss, scattering, and contrast degradation, an affine RGB transformation has limited capacity to recover information that has effectively been attenuated or obscured.

Thus, Figure B1 complements the full-reference numerical evaluation by demonstrating the restoration effect at the image level. Together with Table 4, it indicates that restoration effectiveness depends strongly on the underlying degradation mechanism rather than being uniform across all four SUID categories.

C. Multi-task learning

A shared MLP was jointly trained to predict degradation class, full reference quality (SSIM-like target) and a 12-parameter RGB restoration transform. Held-out results were macro-F1=0.780, quality MAE=0.0379, $R^2=0.283$, restored PSNR=20.91 dB and restored SSIM=0.904

D. Domain adaptation

We did not numerically perform synthetic-to-real adaptation because the supplied archive contains no independent real-underwater target domain. We provide a reproducible DANN/CORAL/MMD protocol

separately rather than fabricating target-domain results.

E. Contrastive learning

We trained an encoder with same-scene positives and different-scene negatives, along with an auxiliary degradation-classification objective. On the held-out scenes, nearest-neighbour retrieval recovered the same scene for 0.944 of images and the same condition for 0.728.

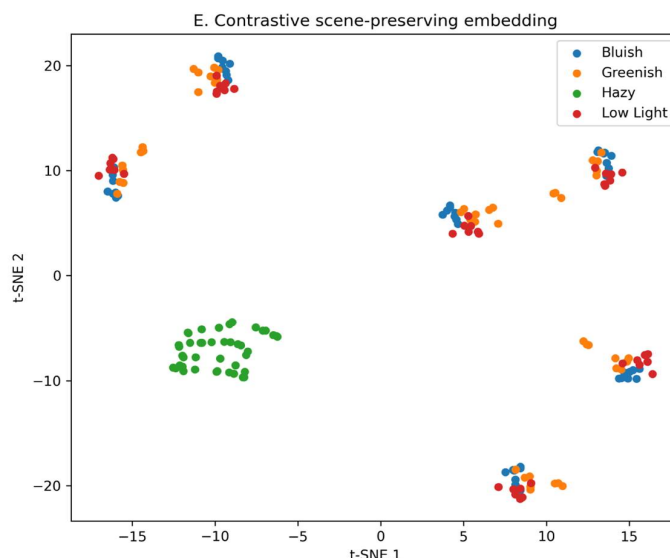


Figure E1. Held-out contrastive embedding

Figure E1 visualizes the learned representation for the held out scenes following contrastive learning. The embedding provides a two dimensional view of the learned feature space and shows whether images associated with the same scene or degradation condition occupy related regions.

The high 94.44% same scene nearest neighbour retrieval indicates that the learned embedding preserves scene level identity effectively. Images originating from the same ground truth scene remain closely associated despite the presence of different synthetic degradation conditions. At the same time, the 72.78% same-condition retrieval shows that the embedding also retains degradation information.

The contrast between these two retrieval measures suggests that the representation achieves a useful balance between scene invariance to degradation and degradation sensitivity. This is an important property for underwater image analysis because a useful representation should not confuse a change in degradation with a change in scene identity. Nevertheless, the results should be interpreted as an initial controlled experiment because the dataset contains only 30 underlying scenes. The source analysis therefore recommends repeated GroupKFold or leave-one-scene-out validation before making stronger generalization claims.

Because only 30 underlying scenes are available, treat the reported results as an initial controlled experiment. Repeated GroupKFold or leave-one-scene-out validation is recommended for a final journal study. Avoid random image level splitting because variants of the same scene occur across conditions.

8. Discussion

8.1 The Heterogeneity of Underwater Degradation and Restoration Limits

A central finding of this study is that underwater degradation is not a monolithic phenomenon; rather, it comprises distinct physical processes that require tailored restoration strategies. The paired restoration experiment clearly demonstrated this heterogeneity. Condition-specific affine RGB transformations proved

highly effective for illumination based (Low Light) and wavelength-attenuation-based (Bluish, Greenish) degradations, recovering significant structural and color fidelity (e.g., +14.5 dB PSNR gain for Low-Light). However, the affine baseline failed on the Hazy category, resulting in negative PSNR and SSIM gains. This failure highlights a critical limitation: haze is driven by light scattering and severe visibility/contrast loss, which cannot be reversed by simple global color or illumination shifts. This underscores the necessity for future UIE models to incorporate physics-aware or deep feature extraction mechanisms capable of reconstructing information lost to scattering, rather than relying solely on global affine mappings.

8.2 Multi-Task Learning and Representation Robustness

The transition from sequential enhancement and detection pipelines to integrated, task-oriented frameworks is critical for resource-constrained AUVs. Our multi-task learning experiment demonstrated that a shared multilayer perceptron (MLP) can simultaneously handle degradation classification, quality estimation, and parameterized restoration without significant performance degradation (maintaining a Macro-F1 of 0.780, closely matching the standalone baseline of 0.787). Furthermore, the contrastive representation analysis revealed a highly desirable geometric property in the learned embedding space. By utilizing same scene positives and different scene negatives alongside an auxiliary classification objective, the model achieved a 94.4% same retrieval rate compared to a 72.8% same condition retrieval rate. This indicates that the representation successfully disentangles scene identity from degradation artifacts. For downstream tasks like Underwater Object Detection (UOD), this is vital: the network learns to recognize the *object/scene* regardless of the water condition, while still retaining enough degradation sensitivity to inform the restoration or adaptation heads.

8.3 Methodological Rigor: The Scene-Independent Imperative

A major methodological contribution of this work is the strict enforcement of scene independent data partitioning. The SUID contains 930 images, but only 30 independent ground truth scenes. A conventional random image-level split would place nearly identical scene variants into both training and testing sets, allowing the model to memorize scene-specific textures rather than learning generalized degradation patterns. By grouping all synthetic variants of a ground-truth scene into a single partition, we provide a much more conservative and scientifically defensible estimate of generalization. The resulting classification accuracy of 78.9%—and the specific confusion observed between the chromatically similar Bluish and Greenish classes reflects genuine model behavior rather than inflated performance due to data leakage. Researchers utilizing paired synthetic datasets must adopt this scene-level validation (e.g., GroupKFold or leave one scene out) to ensure robust benchmarking.

8.4 Limitations and the Path to Synthetic-to-Real Domain Adaptation

While this study provides a comprehensive controlled benchmark, it is bounded by the synthetic nature of the SUID. The dataset lacks an independent, real-underwater target domain, meaning the proposed domain adaptation protocols (DANN, CORAL, MMD) could not be numerically executed in this phase. Furthermore, the reliance on compact visual descriptors and affine transformations for the baselines limits the capacity to model complex, non linear underwater optical physics. Future work must extend this framework by integrating the proposed domain adaptation protocols with real world underwater datasets to bridge the synthetic-to-real gap. Additionally, replacing the affine restoration baselines with deep, physics informed generative models will be necessary to address the complex scattering inherent in Hazy conditions.

9. Conclusion

This study established a controlled, scene independent framework for analyzing synthetic underwater image degradation, addressing the critical gap between perceptual image enhancement and downstream computer vision performance. By leveraging the Synthetic Underwater Image Dataset (SUID) across four distinct degradation regimes (Bluish, Greenish, Hazy, and Low Light), we demonstrated that underwater degradation is highly heterogeneous, requiring condition specific restoration strategies. While simple affine transformations

successfully recovered color and illumination degraded images, they proved insufficient for scattering-dominated haze, highlighting the need for advanced, physics aware restoration architectures.

Through multi task learning and contrastive representation analysis, we proved that shared embeddings can effectively balance degradation classification, quality prediction, and restoration, while successfully preserving underlying scene identity across varying water conditions. Crucially, by enforcing scene independent data partitioning, this work provides a rigorous, leakage free benchmark for evaluating underwater vision algorithms. Finally, by outlining a reproducible protocol for synthetic to real domain adaptation, this research lays the foundational groundwork for developing robust, computationally efficient, and generalizable vision systems capable of supporting autonomous marine exploration and ecological monitoring in complex, real-world underwater environments.

References

- [1] Wen, J., Cui, J., Zhao, B., Han, B., Liu, X., Gao, Z., Chen, B. M. (2024). A real-time framework for domain-adaptive underwater object detection with image enhancement. *arXiv preprint arXiv:2403.19079*.
- [2] Jesus, A., Zito, C., Tortorici, C., Roura, E., De Masi, G. (2022). Underwater object classification and detection: First results and open challenges. *OCEANS 2022-Chennai*, 1–6.
- [3] Katija, K. (2023). Autonomous agents for observing marine life. *Science Robotics*, 8(80), eadi6428.
- [4] Infanta Princy, S., Brindha, M. (2026). A quality-guided framework for low-light underwater image enhancement using degradation classification and benchmark evaluation. In B. N. Subudhi, R. K. Rout, & B. S. Sengar (Eds.), *Computer vision, pattern recognition, image processing, and graphics. NCVPRIPG 2025. Communications in Computer and Information Science* (Vol. 2685). Springer, Cham. https://doi.org/10.1007/978-3-032-08511-5_21.
- [5] Mallet, D., Pelletier, D. (2014). Underwater video techniques for observing coastal marine biodiversity: A review of sixty years of publications (1952–2012). *Fisheries Research*, 154, 44–62. <https://doi.org/10.1016/j.fishres.2014.01.019>.
- [6] Carlucho, I., De Paula, M., Wang, S., Petillot, Y., Acosta, G. G. (2018). Adaptive low-level control of autonomous underwater vehicles using deep reinforcement learning. *Robotics and Autonomous Systems*, 107, 71–86. <https://doi.org/10.1016/j.robot.2018.05.016>.
- [7] Sahoo, A., Dwivedy, S. K., Robi, P. S. (2019). Advancements in the field of autonomous underwater vehicle. *Ocean Engineering*, 181, 145–160. <https://doi.org/10.1016/j.oceaneng.2019.04.011>.
- [8] Chiang, J. Y., Chen, Y. C. (2012). Underwater image enhancement by wavelength compensation and dehazing. *IEEE Transactions on Image Processing*, 21(4), 1756–1769. <https://doi.org/10.1109/TIP.2011.2179666>.
- [9] Cong, X., Zhao, Y., Gui, J., Hou, J., Tao, D. (2026). A comprehensive survey on underwater image enhancement based on deep learning. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 10(4), 2783–2801. <https://doi.org/10.1109/TETCI.2026.3668671>.
- [10] Wang, R., Zhang, Y., Zhang, J. (2023). An efficient Swin Transformer based method for underwater image enhancement. *Multimedia Tools and Applications*, 82(12), 18691–18708.
- [11] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B. (2021). Swin Transformer: Hierarchical vision

transformer using shifted windows. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10012–10022.

[12] Huang, S., Wang, K., Liu, H., Chen, J., Li, Y. (2023). Contrastive semi supervised learning for underwater image restoration via reliable bank. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 18145–18155.

[13] Jamieson, S., How, J. P., Girdhar, Y. (2023). DeepSeeColor: Realtime adaptive color correction for autonomous underwater vehicles via deep learning methods. *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 3095–3101.

[14] Ren, S., He, K., Girshick, R., Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, 28.

[15] Redmon, J., Divvala, S., Girshick, R., Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 779–788.

[16] Jiang, L., Wang, Y., Jia, Q., Xu, S., Liu, Y., Fan, X., Li, H., Liu, R., Xue, X., Wang, R. (2021). Underwater species detection using channel sharpening attention. *Proceedings of the 29th ACM International Conference on Multimedia*, 4259–4267.

[17] Syariz, M. A., Lin, C.-H., Nguyen, M. V., Jaelani, L. M., Blanco, A. C. (2020). WaterNet: A convolutional neural network for chlorophyll a concentration retrieval. *Remote Sensing*, 12(12), 1966.

[18] Fan, B., Chen, W., Cong, Y., Tian, J. (2020). Dual refinement underwater object detection network. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XX 16* (pp. 275–291). Springer.

[19] Liu, W., Ren, G., Yu, R., Guo, S., Zhu, J., Zhang, L. (2022). Image-adaptive YOLO for object detection in adverse weather conditions. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36, 1792–1800.

[20] Li, C., Zhou, H., Liu, Y., Yang, C., Xie, Y., Li, Z., Zhu, L. (2023). Detection-friendly dehazing: Object detection in real-world hazy scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

[21] Sun, S., Ren, W., Wang, T., Cao, X. (2022). Rethinking image restoration for object detection. *Advances in Neural Information Processing Systems*, 35, 4461–4474.

[22] Cheng, N., Xie, H., Zhu, X., Wang, H. (2023). Joint image enhancement learning for marine object detection in natural scene. *Engineering Applications of Artificial Intelligence*, 120, 105905.

[23] Nabahirwa, E., Song, W., Zhang, M., Fang, Y., Ni, Z. (2025). A structured review of underwater object detection challenges and solutions: From traditional to large vision language models. *arXiv preprint arXiv:2509.08490*.

[24] Moniruzzaman, M., Islam, S. M. S., Bennamoun, M., Lavery, P. (2017). Deep learning on underwater marine object detection: A survey. In J. Blanc-Talon, R. Penne, W. Philips, D. Popescu, & P. Scheunders (Eds.), *Advanced Concepts for Intelligent Vision Systems* (Vol. 10617, p. 150–160). Springer International Publishing. https://doi.org/10.1007/978-3-319-70353-4_13.

[25] Wang, N., Wang, Y., Er, M. J. (2022). Review on deep learning techniques for marine object recognition: Architectures and algorithms. *Control Engineering Practice*, 118, 104458. <https://doi.org/10.1016/j.conengprac.2022.104458>.

[j.conengprac.2020.104458](https://doi.org/10.1007/s11042-022-12502-1).

- [26] Fayaz, S., Parah, S. A., Qureshi, G. J. (2022). Underwater object detection: Architectures and algorithms – A comprehensive review. *Multimedia Tools and Applications*, 81(15), 20871–20916. <https://doi.org/10.1007/s11042-022-12502-1>.
- [27] Jian, M., Yang, N., Tao, C., Zhi, H., Luo, H. (2024). Underwater object detection and datasets: A survey. *Intelligent Marine Technology and Systems*, 2(1), 9. <https://doi.org/10.1007/s44295-024-00023-6>.
- [28] Chen, L., Huang, Y., Dong, J., Xu, Q., Kwong, S., Lu, H., Lu, H., Li, C. (2024). *Underwater object detection in the era of artificial intelligence: Current, challenge, and future*. arXiv. <https://arxiv.org/abs/2410.05577>.
- [29] Khan, A., Fouda, M. M., Do, D.-T., Almaleh, A., Alqahtani, A. M., Rahman, A. U. (2024). Underwater target detection using deep learning: Methodologies, challenges, applications, and future evolution. *IEEE Access*, 12, 12618–12635. <https://doi.org/10.1109/ACCESS.2024.3353688>.
- [30] Xu, S., Zhang, M., Song, W., Mei, H., He, Q., Liotta, A. (2023). A systematic review and analysis of deep learning-based underwater object detection. *Neurocomputing*, 527, 204–232. <https://doi.org/10.1016/j.neucom.2023.01.056>.
- [31] Fu, C., Liu, R., Fan, X., Chen, P., Fu, H., Yuan, W., Zhu, M., Luo, Z. (2023). Rethinking general underwater object detection: Datasets, challenges, and solutions. *Neurocomputing*, 517, 243–256. <https://doi.org/10.1016/j.neucom.2022.10.039>.
- [32] Akkaynak, D., Treibitz, T. (2019). Sea-thru: A method for removing water from underwater images. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1682–1691.
- [33] Wen, J., Cui, J., Zhao, Z., Yan, R., Gao, Z., Dou, L., Chen, B. M. (2023). SYREANet: A physically guided underwater image enhancement framework integrating synthetic and real images. *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 5177–5183.