

# **ENHANCEMENT AND RESTORATION OF IMAGES UNDER ADVERSE VISUAL CONDITIONS USING DEEP LEARNING TECHNIQUES**

**A Thesis Submitted  
In Partial Fulfilment of the Requirements for the  
Degree of**

**MASTER OF TECHNOLOGY**

**in  
Artificial Intelligence  
by**

**Jyotirmaya Tembhurne  
(Roll No. 2K23/AFI/19)**

**Under the Supervision of  
Prof. Rahul Katarya  
(Dept. of Computer Science & Engineering)**



**Department of Computer Science and Engineering**

**DELHI TECHNOLOGICAL UNIVERSITY**

**(Formerly Delhi College of Engineering)**

**Shahbad Daultpur, Main Bawana Road, Delhi-110042. India**

**May, 2025**

## ACKNOWLEDGEMENTS

This research would not have been possible without the guidance, support, and encouragement of many individuals and institutions.

I would like to express my deepest gratitude to Prof. **Rahul Katarya** for his invaluable mentorship, insightful feedback, and unwavering support throughout this project. His expertise and encouragement were instrumental at every stage, from conceptualization to implementation and analysis. I am also sincerely thankful to the **Head of the Department of Computer Science and Engineering at Delhi Technological University** for providing a stimulating academic environment and for their continuous encouragement. My appreciation extends to all faculty and staff members of the department, whose assistance and cooperation greatly facilitated the completion of this work. I acknowledge my colleagues and peers for their collaborative spirit, constructive discussions, and technical assistance, which enriched the research process and contributed to the successful realization of these implementation papers.

Finally, I am grateful to all the people involved for their patience, understanding, and unwavering moral support throughout this journey. Their encouragement provided the foundation that enabled me to persevere and complete this research.

Jyotirmaya Tembhurne  
1 June 2024

**DELHI TECHNOLOGICAL UNIVERSITY**  
(Formerly Delhi College of Engineering)  
Shahbad Daulatpur, Main Bawana Road, Delhi-42

**CANDIDATE’S DECLARATION**

I, **Jyotirmaya Tembhurne**, Roll No. 2K23/AFI/19, student of M. Tech (Artificial Intelligence), hereby certify that the work which is being presented in the thesis entitled “**Enhancement and Restoration of Images Under Adverse Visual Conditions Using Deep Learning Techniques**” in partial fulfilment of the requirements for the award of the Degree of Master of Technology in Artificial Intelligence in the Department of Computer Science and Engineering, Delhi Technological University is an authentic record of my own work carried out during the period from August 2023 to June 2025 under the supervision of Dr. Rahul Katarya, Professor, Department of Computer Science and Engineering. The matter presented in the thesis has not been submitted by me for the award of any other degree of this or any other Institute.

**Candidate’s Signature**

This is to certify that the student has incorporated all the corrections suggested by the examiners in the thesis and the statement made by the candidate is correct to the best of our knowledge.

**Signature of Supervisor**

**Signature of External Examiner**

**DELHI TECHNOLOGICAL UNIVERSITY**  
(Formerly Delhi College of Engineering)  
Shahbad Daulatpur, Main Bawana Road, Delhi-42

**CERTIFICATE**

This is to certify that **Jyotirmaya Tembhurne** (Roll No. 2K23/AFI/19) has carried out the research work presented in the thesis titled **“Enhancement and Restoration of Images Under Adverse Visual Conditions Using Deep Learning Techniques”**, for the award of Degree of Master of Technology from Department of Computer Science and Engineering, Delhi Technological University, Delhi under my supervision. The thesis embodies result of original work and studies are carried out by the student himself and the contents of the thesis do not form the basis for the award of any other degree for the candidate or submit else from the any other University /Institution.

Date:

Prof. Rahul Katarya  
(Supervisor)  
Department of CSE  
Delhi Technological University

## ABSTRACT

The rapid development of deep learning has revolutionized the field of image enhancement and restoration, particularly for images captured under challenging conditions such as low-light and underwater environments. This thesis investigates the efficacy of advanced deep neural network architectures in improving image quality, visibility, and structural fidelity in scenarios where traditional methods often fall short. Leveraging state-of-the-art models-including architectures with edge-aware modules, attention mechanisms, and transformer-based context modeling-this research demonstrates significant improvements in both quantitative metrics (such as PSNR, SSIM, and LPIPS) and qualitative visual outcomes. Experiments conducted on benchmark datasets, including LOLv1, LOLv2, SID, LSUI, EUVP, and UFO-120, reveal that the proposed frameworks achieve high restoration accuracy and efficiency, with real-time processing capabilities suitable for deployment in resource-constrained environments. The results show substantial gains over traditional and contemporary baselines, confirming the models' robustness and adaptability across diverse real-world conditions. This study further addresses practical considerations such as computational demands, generalization, and the integration of these methods into applications ranging from surveillance and autonomous navigation to marine exploration and medical imaging. The findings highlight the transformative potential of deep learning in advancing image enhancement technology, offering scalable and effective solutions that benefit a broad spectrum of scientific, industrial, and societal domains.

## List of Publications

Paper Title	Conference	Publication Series
<b>LEARN: Laplacian Enhanced Attention and Residual Network for Low-Light Image Enhancement</b>	6 <sup>th</sup> International Conference on Data Analysis and Management (ICDAM), June 2025, London Metropolitan University, London, UK (Accepted)	Springer LNNS (Indexing: SCOPUS, INSPEC, WTI Frankfurt eG, zbMATH, SCImago)
<b>SETAU-NET: Sobel Transformer Attention U-Net for Underwater Image Enhancement</b>	6 <sup>th</sup> International Conference on Data Analysis and Management (ICDAM), June 2025, London Metropolitan University, London, UK (Accepted)	Springer LNNS (Indexing: SCOPUS, INSPEC, WTI Frankfurt eG, zbMATH, SCImago)

# TABLE OF CONTENTS

<b>TITLE</b>	<b>PAGE NO.</b>
<b>Acknowledgements</b>	<b>ii</b>
<b>Candidate's Declaration</b>	<b>iii</b>
<b>Certificate</b>	<b>iv</b>
<b>Abstract</b>	<b>v</b>
<b>List of Publications</b>	<b>vi</b>
<b>Table of Contents</b>	<b>vii</b>
<b>List of Tables</b>	<b>viii</b>
<b>List of Figures</b>	<b>ix</b>
<b>List of Abbreviations</b>	<b>xi-xii</b>
<b>CHAPTER 1- INTRODUCTION</b>	<b>1-5</b>
1.1 Motivation	2
1.2 Objective	2-3
1.3 Challenges	3-4
1.4 Thesis Organization	4-5
<b>CHAPTER 2- RELATED WORK</b>	<b>6-13</b>
2.1 Literature Survey	6-11
2.2 Datasets	12-13
2.3 Problem Statement	13
<b>CHAPTER 3- PROPOSED METHODOLOGY</b>	<b>14-28</b>
3.1 LEARN	14-20
3.1.1 Architecture	14-15
3.1.2 Training Configuration	15-16
3.1.3 Model Operation	16-18
3.2 SETAU-Net	20-28
3.2.1 Architecture	20-23
3.2.2 Training Configuration	23-25
3.2.3 Model Operation	25-27
<b>CHAPTER 4- EXPERIMENTAL RESULTS AND ANALYSIS</b>	<b>29-37</b>
4.1 Experimental Setup	28-30
4.2 Datasets	30
4.3 Baselines	31
4.4 Results and Analysis	31-37
<b>CHAPTER 5- CONCLUSION, FUTURE WORK AND SOCIAL IMPACT</b>	<b>38-41</b>

5.1 Conclusion	38-39
5.2 Future work	39-40
5.3 Social Impact	40-41
<b>References</b>	<b>42-46</b>
<b>Acceptance Letters and Registration Payment Receipts</b>	<b>47</b>



## List of Tables

Table Number	Table Name	Page Number
2.1	Overview of Traditional and Machine-Learning Approaches to Underwater Image Enhancement	8-10
2.2	Overview of Traditional and Machine-Learning Approaches to Underwater Image Enhancement	10-11
2.3	Common Datasets used in Enhancement and restoration of Adverse (Low-Light and Underwater) Images	13
3.1	LEARN Model Configuration	19-20
3.2	SETAU-Net Architecture Configuration	26-27
4.1	Dataset Image Split Pairs Used for Testing	30
4.2	Real-time Performance Evaluation of SETAU-Net	31
4.3	Comparative Analysis of SETAU-Net and Existing Underwater Image Enhancement (For each metric/dataset, the best result is highlighted in <b>red</b> , second best is highlighted in <b>blue</b> and third best is highlighted in <b>purple</b> ). ↑ Denotes that a higher value for a particular metric is better while ↓ denotes that a lower value for a particular metric is better	32
4.4	Real-time Performance Evaluation of LEARN Model (in seconds)	33
4.5	Comparative Analysis of SETAU-Net and Existing Underwater Image Enhancement (For each metric/dataset, the best result is highlighted in <b>red</b> and second best is highlighted in <b>blue</b> ). ↑ Denotes that a higher value for a particular metric is better while ↓ denotes that a lower value for a particular metric is better	33
4.6	Visual Comparison of Low Light Image Enhancement Across Multiple Datasets	36
4.7	Visual Comparison of Underwater Enhancement Across Multiple Datasets	36

## List of Figures

<b>Figure Number</b>	<b>Figure Name</b>	<b>Page Number</b>
3.1	LEARN Model Architecture Diagram	18
3.2	SETAU-Net Architecture Diagram	27
4.1	Training Loss Convergence of LEARN	30
4.2	PSNR Comparison of Low Light Image Enhancement Methods Across LOL-v1, LOL-v2 Real and LOL-v2 Synthetic Datasets (Higher is better)	33
4.3	SSIM Comparison of Low Light Image Enhancement Methods Across LOL-v1, LOL-v2 Real and LOL-v2 Synthetic Datasets (Higher is better)	34
4.4	SSIM Comparison of Underwater Image Enhancement Methods Across LSUI, EUVP and UFO-120 Datasets (Higher is better)	34
4.5	LPIPS Comparison of Underwater Image Enhancement Methods Across LSUI, EUVP and UFO-120 Datasets (lower is better)	35
4.6	PSNR Comparison of Underwater Image Enhancement Methods Across LSUI, EUVP and UFO-120 Datasets (Higher is better)	35
4.7	Comparison of Number of Parameters in various Underwater Image Enhancement Methods Across LSUI, EUVP and UFO-120 Datasets (lower is better)	35

## List of Abbreviations

Abbreviation	Full Form
AdamW	Adaptive Moment Estimation Optimizer with Weight Decay
AI	Artificial Intelligence
CA-VAE	Channel Attention Variational Autoencoder
CBAM	Convolutional Block Attention Module
CLUIE-Net	Contrast Limited Underwater Image Enhancement Network
CNN	Convolutional Neural Network
DL	Deep Learning
EUVP	Enhancing Underwater Visual Perception Dataset
FPS	Frames Per Second
GAN	Generative Adversarial Network
GFLOPs	Giga Floating Point Operations per Second
HE	Histogram Equalization
KinD	Kindling the Darkness (Low-Light Enhancement Model)
LEARN	Laplacian Enhanced Attention and Residual Network
LLIE	Low Light Image Enhancement
LOL	Low-Light Dataset
LPIPS	Learned Perceptual Image Patch Similarity
LSUI	Large-Scale Underwater Image Dataset
MIRNet	Multi-Scale Residual Network
ML	Machine Learning
MSE	Mean Squared Error
PSNR	Peak Signal-to-Noise Ratio
ReLU	Rectified Linear Unit
RGB	Red Green Blue (color space)
RGHS	Red Channel and Gradient Histogram Stretching

SETAU-Net	Sobel Enhancement with Transformer Attention U-Net
SID	See-in-the-Dark Dataset
SimAM	Simple, Parameter-Free Attention Module
SSIM	Structural Similarity Index
TWIN	Two-stage Water-type Identification Network
UDCP	Underwater Dark Channel Prior
UFO-120	Underwater Image Dataset with 120 Scenes
UGAN	Underwater Generative Adversarial Network
UIBLA	Underwater Image Blurriness and Light Absorption
UIE	Underwater Image Enhancement
U-Net	Encoder-Decoder Network with Skip Connections (U-shaped Network)
VGG	Visual Geometry Group (refers to VGGNet, a deep CNN architecture)
Zero-DCE	Zero-Reference Deep Curve Estimation

# CHAPTER 1

## INTRODUCTION

Enhancing and restoring images captured under adverse visual conditions is a persistent and critical challenge in computer vision, with significant implications for diverse fields such as photography, surveillance, marine exploration, and autonomous systems. Images acquired in low-light environments or underwater are particularly susceptible to severe degradation, including reduced visibility, color distortion, amplified noise, and loss of structural detail. These degradations not only hinder human interpretation but also impede the effectiveness of automated analysis and decision-making systems.

In the context of low-light image enhancement, poor illumination frequently results in diminished visibility, inaccurate color representation, and increased noise, making it difficult to extract meaningful information from the images. Traditional enhancement techniques, such as histogram equalization [1,2,3,4] and Retinex-based methods [5, 6, 14, 15], have been widely adopted but often introduce artifacts, unnatural color shifts, or fail to generalize across varying lighting conditions. Deep learning approaches-including RetinexNet [6], EnlightenGAN [7], and transformer-based models [15, 17] have propelled the field forward by learning complex mappings between low-light and standard-brightness images, achieving improved performance in visibility and color accuracy. However, these methods typically require extensive computational resources, complex architectures, and large paired datasets, which can limit their practical deployment. To address these challenges, recent research has focused on developing lightweight and efficient architectures that integrate edge enhancement modules, expanded kernel residual blocks, and attention mechanisms. Such designs enable real-time enhancement of low-light images, preserving structural details and natural color while remaining suitable for resource-constrained environments.

On the other hand, underwater image enhancement poses a distinct set of challenges due to the unique optical properties of aquatic environments. Light absorption and scattering underwater cause substantial color casts (often blue or green), reduced contrast, and significant loss of detail, which are exacerbated by the complexity and variability of underwater scenes. These challenges not only affect the visual appeal of underwater photographs but also limit the effectiveness of vision-based marine research, ecological monitoring, and robotic exploration. Traditional underwater enhancement methods-both non- physics-based [30, 31] and physics-based [32, 33, 34] - often struggle to generalize across different water types and lighting conditions, and may require precise knowledge of environmental parameters. Recent advances in deep learning, particularly those leveraging U-Net architectures [62], attention mechanisms [48, 49, 50, 54], and transformer modules [27, 51, 52, 53], have demonstrated improved generalization and restoration quality. However, many of these models remain computationally intensive and dependent on large paired or unpaired datasets, which are difficult to obtain in underwater settings. To overcome these limitations, novel approaches have emerged that combine lightweight encoder-decoder [55, 62] frameworks, parameter-free attention modules [54], and transformer-based global context modelling. These architectures are specifically designed to extract both local and global features, enabling robust color correction, detail recovery, and real-time performance suitable for deployment in autonomous and resource-constrained underwater systems.

This thesis uses advanced deep learning approaches to tackle the pressing difficulties of image improvement and restoration under unfavourable visual conditions. The next chapters will go over the associated research, recommended methodology, experimental analyses, and major

conclusions. Through this research, we aim to contribute robust, efficient solutions with practical relevance across a range of real-world applications.

## 1.1 Motivation

Enhancing and restoring images captured under challenging conditions such as low-light and underwater environments holds significant societal importance across multiple domains. In autonomous driving and surveillance, improved image clarity directly enhances safety by enabling better scene understanding and decision-making in poor visibility conditions. In marine science and environmental conservation, restoring underwater images facilitates accurate monitoring of fragile ecosystems, biodiversity assessment, and sustainable resource management, which are crucial for protecting ocean health and supporting global ecological balance. Furthermore, enhanced imaging plays a vital role in medical diagnostics, disaster response, and industrial automation by providing clearer visual data for timely and accurate interventions. By enabling robust visual perception in these adverse settings, image enhancement technologies contribute to safer transportation, environmental sustainability, improved healthcare, and economic efficiency, thereby positively impacting society at large.

Despite the critical need and wide-ranging applications, image enhancement under adverse conditions remains a complex technical challenge. Low-light images suffer from noise amplification, color distortion, and loss of detail, while underwater images face unique degradations such as color casts, scattering, and uneven illumination. Traditional enhancement methods often fail to generalize across diverse scenarios or introduce artifacts, limiting their practical utility. Although deep learning approaches have advanced the state of the art, many existing models are computationally intensive, require large paired datasets, and lack real-time feasibility, restricting their deployment in resource-constrained or autonomous systems. Motivated by these limitations, this research aims to develop lightweight, efficient deep learning architectures that integrate edge enhancement, attention mechanisms, and global context modelling to achieve high-quality restoration with practical computational demands. This balance is essential to enable real-world applications ranging from autonomous underwater vehicles to mobile low-light photography, ultimately bridging the gap between research and impactful deployment.

## 1.2 Objective

The main objective of this study is to develop deep learning-based models capable of accurately and efficiently enhancing and restoring images captured under adverse visual conditions, with a specific focus on low-light and underwater scenarios. The proposed frameworks are designed to analyze and process complex degradations—such as reduced visibility, color distortion, noise amplification, and loss of detail—by leveraging advanced architectural components including edge enhancement modules, attention mechanisms, and global context modelling. These models aim to deliver high-quality, perceptually natural images while maintaining computational efficiency suitable for real-time and resource-constrained environments. This study's key objectives are summarized below:

- To develop robust and efficient deep learning-based architectures for enhancing and restoring images captured under adverse visual conditions, specifically targeting low-light and underwater scenarios. The goal is to address common degradations such as reduced visibility, color distortion, noise amplification, and loss of detail that hinder both human interpretation and automated analysis.
- To design models that integrate advanced feature extraction and refinement mechanisms, including edge enhancement modules (such as Laplacian and Sobel

operators), attention mechanisms (like CBAM and SimAM [54]), and global context modelling (via transformer blocks [52] or expanded kernel residuals). These components aim to preserve structural details, improve color fidelity, and adaptively focus on relevant features while maintaining computational efficiency.

- To achieve a balance between high-quality enhancement and practical deployability by maintaining lightweight architectures with low computational and memory requirements, enabling real-time or near real-time performance suitable for resource-constrained environments such as autonomous vehicles, embedded systems, and underwater robots.
- To rigorously evaluate the proposed methods on standard and diverse benchmark datasets for both low-light and underwater image enhancement (e.g., LOLv1 [6], LOLv2 [9], SID [10], LSUI [27], EUVP [28], UFO-120 [29]), using quantitative metrics such as PSNR, SSIM, and LPIPS, as well as qualitative visual assessment, to demonstrate superiority or competitiveness over traditional and state-of-the-art approaches.
- To ensure generalization and robustness of the enhancement models across a wide range of real-world conditions by employing extensive data augmentation, multi-dataset training, and comprehensive validation, thereby supporting practical applications in fields like surveillance, marine exploration, medical imaging, and environmental monitoring.

### 1.3 Challenges

Enhancing and restoring images captured under adverse visual conditions-such as low-light and underwater environments-presents a range of complex and interrelated challenges that impact both the quality of the output and the practicality of deploying enhancement models in real-world scenarios.

Images captured in low-light conditions typically suffer from reduced visibility, low contrast, significant noise amplification, and color distortion. Traditional enhancement methods, such as histogram equalization [1, 2, 3, 4] and Retinex-based algorithms [5, 6, 14, 15], often introduce artifacts or unnatural colors and struggle to generalize across varying lighting scenarios. Deep learning-based methods have shown promise in overcoming some of these issues, but they frequently require large, paired datasets for supervised learning, involve complex architectures, and demand substantial computational resources. Furthermore, many existing approaches focus primarily on brightness recovery, often neglecting the simultaneous suppression of noise and preservation of structural details, which can lead to over-smoothed or artifact-laden results. Balancing enhancement quality, noise reduction, and computational efficiency remains a significant challenge, especially for real-time or resource-constrained applications such as surveillance as well as autonomous vehicles.

Underwater images face unique and severe degradations due to the optical properties of water, including wavelength-dependent light absorption and scattering, which result in strong color casts (typically blue or green), reduced contrast, blurring, and loss of fine details. The presence of suspended particles and varying water turbidity further exacerbates these issues, leading to spatially non-uniform degradations that are difficult to model and correct. Traditional physics-based enhancement methods [30, 31] often require precise knowledge of environmental parameters and are limited in their ability to generalize across different underwater conditions. Data-driven and deep learning approaches have improved restoration quality, but they are challenged by the scarcity of diverse, high-quality, and paired underwater datasets and often exhibit high computational complexity. Moreover, most existing methods are designed to

address a single type of degradation (e.g., color correction or deblurring), and struggle to handle the interplay between multiple, simultaneous degradations present in real-world underwater scenes. Ensuring robust color correction, detail recovery, and perceptual quality-while maintaining efficiency and adaptability-remains an open challenge.

General Challenges Across Both Domains:

- **Generalization and Robustness:** Models must perform reliably across a wide range of real-world conditions, including unseen lighting environments or diverse underwater scenes, which is difficult given the variability and complexity of degradations.
- **Computational Efficiency:** Achieving high-quality enhancement with lightweight models suitable for real-time or embedded applications is a persistent challenge, as many state-of-the-art methods are computationally intensive.
- **Data Limitations:** The lack of large-scale, diverse, and well-annotated datasets-especially for underwater scenarios-limits the ability to train and validate robust deep learning models.
- **Trade-off Between Enhancement and Artifacts:** Aggressive enhancement can lead to overexposure, color shifts, or loss of naturalness, while insufficient processing leaves noise and degradations unaddressed.
- **Domain Shift and Environmental Variability:** Deep learning models often struggle to generalize across different water types (saltwater, freshwater), depths, turbidity, and lighting conditions. A model trained in one environment may not perform well in another due to significant domain shifts, necessitating domain adaptation techniques or retraining for new conditions.
- **Lack of Paired and Diverse Data for Complex Scenarios:** While data limitations are a general challenge, this is especially acute for paired datasets in complex scenarios like low-light underwater scenes. Most available datasets focus on either low-light or underwater conditions separately, making it difficult to train models that can handle simultaneous degradations such as scattering and insufficient illumination. The scarcity of high-quality, annotated, and paired data hinders the development and validation of robust models for these compounded scenarios.

Addressing these challenges is essential for the development of practical, efficient, and generalizable image enhancement models that can be reliably deployed in real-world low-light and underwater applications.

## 1.4 Thesis Organization

This thesis is structured across several chapters to ensure a logical and thorough presentation of the research:

- **Chapter 1: Introduction** – This section provides an overview of the thesis, outlining the motivation, significance, and scope of the research on image enhancement and restoration under adverse visual conditions. It introduces the core challenges associated with low-light and underwater imaging, and highlights the necessity for robust, efficient solutions. The chapter also presents the main objectives, societal impact, and structure of the thesis, setting the stage for the detailed discussions in subsequent chapters.
- **Chapter 2: Related Work** – This chapter reviews recent advancements and existing literature on the application of deep learning techniques for image enhancement and



restoration under challenging conditions such as low-light and underwater environments. It highlights key methodologies and findings that provide the foundation for the present research.

- **Chapter 3: Proposed Methodology** – This section offers a detailed account of the research approach, including steps such as data pre-processing, augmentation strategies, and a comprehensive description of the architectures and components of the proposed models.
- **Chapter 4: Experiments and Results** – This chapter presents the experimental setup, including dataset selection, training procedures, and evaluation metrics. It provides an in-depth analysis of the results obtained from testing the developed models, comparing their performance against existing methods.
- **Chapter 5: Conclusion and Future Work** – The final chapter summarizes the main findings of the research and discusses potential directions for future work that could further improve the effectiveness and applicability of deep learning-based image enhancement and restoration techniques.

## CHAPTER 2

### RELATED WORK

#### 2.1 Literature Survey

Traditional low-light image enhancement techniques use histogram adjustments or Retinex. HE methods like CLAHE [4] often increase noise and artificial results while increasing visibility and contrast by ignoring illumination. With multi-scale processing and color restoration, MSRCR expands on Retinex algorithms, which divide images into illuminated and reflective parts. Since these methods assume corruption-free images, they often introduce noise or color distortion. LIME [13] estimates and refines illumination maps using structure priors, but noise in very dark regions is a problem. A robust Retinex-based method to improve low-light images by explicitly adding a noise map to the standard model is described in [14]. This method cuts down on noise while showing structural details by using an optimization function with new regularizing parts, like the gradient integrity term for reflectance and the  $\ell_1$  norm for illumination smoothness. In low-light conditions, logarithmic transformations are ineffective for noise management. The suggested method uses Lagrange multiplier-based optimization. The method reduces noise and improves visibility, but it risks feature loss in low-noise images and is computationally expensive. Not all settings can use it because it requires human parameter adjustment.

Deep learning improves low-light enhancement in many ways. In the absence of paired data, EnlightenGAN [7] uses unpaired training and a global-local discriminator structure to produce visually appealing results. However, extreme conditions can distort colors. With a one-stage Retinex framework, transformer-based methods like Retinexformer [15] improve low-light image enhancement. Initially, the illumination brightens the image, followed by the restoration of noise, artifacts, and color distortions. Its illumination-guided transformer models non-local interactions between differently lit regions, outperforming previous methods across multiple benchmarks. DSLR [16] (Deep Stacked Laplacian Restorer) Employs the Laplacian pyramid in both image and feature domains. DSLR [16] breaks the input image into a three-level Laplacian pyramid for multi-scale processing, recovering global illumination at the coarsest level while preserving local details at finer levels. When applied to LLIE workloads, Swin IR, a Swin Transformer-based backbone, performs well but has overexposure and brightness imbalance artifacts. To deal with these problems, new methods have been developed, such as SNR-aware Swin Transformer networks [17], which use signal-to-noise ratio (SNR) maps to balance local and global feature extraction and drive spatially varying augmentation. On benchmark datasets like LOL-v1 [6] and LOL-v2 [9], these methods and unsupervised learning models like Retinex models perform similarly to paired training data. PSNR and SSIM are also competitive.

In [41, 42], autonomous image processing techniques have been introduced to rectify non-uniform illumination, mitigate noise, augment contrast, and modify colors. Alternative techniques used edge detection activities to facilitate object-edge preservation during filtering processes for color enhancement [43]. In [44], it has been noted that the image channels respond variably to light disruption: red hues diminish after a few meters from the surface, but green and blue exhibit greater persistence. The disparities led to the development of enhancement techniques that operate distinctly on each color channel, prioritizing specialized filters influenced by ambient parameters over generalization [45, 46]. Alternative methodologies assessed the parameters of global background light [45, 47] to implement specific color corrections (i.e., to mitigate the bluish and greenish impacts). These models employ the principles of optics and chromatics to address diverse underwater situations. While

these models are more precise, acquiring all the necessary elements that influence underwater footage constrains their utilization.

Underwater image augmentation machine learning methods typically utilized a U-Net-like architecture in order to improve the source image while keeping the spatial details and object interactions intact. With a focus on attention and pooling layers [48], skip connections are commonly used to send the raw inputs to the final layers in order to preserve spatial links [49, 50]. Other methods explored the emerging use of Transformer architectures [27, 51, 52, 53] and advanced attention mechanisms have become popular for improving underwater images, taking use of their effectiveness in broader computer vision tasks. Originally designed for NLP, transformers [52] have shown to be highly effective at representing global context and long-range dependency in image models. Underwater enhancement approaches like UDAformer [53] and U-shape Transformer [27] use channel-wise and spatial-wise self-attention modules to deal with complex, irregular degradations. In order to improve color correction, contrast, and detail restoration, these approaches combine convolutional and transformer blocks to capture both local properties and global associations. If better feature representations are required without significantly increasing computing complexity, lightweight attention approaches like SimAM [54] and channel attention modules are suitable. Through the incorporation of these enhancements, recent methodologies have demonstrated the efficiency of transformer-based and attention-driven structures for robust and generalizable underwater image enhancement, achieving improved performance on demanding underwater datasets.

SETAU-Net and LEARN exemplify advanced machine learning-based strategies for image enhancement under adverse visual conditions, each tailored to address the unique challenges of underwater and low-light scenarios, respectively. SETAU-Net leverages a deep neural network architecture that integrates transformer-driven global context modeling, attention-based refinement, and explicit edge feature extraction to significantly improve underwater images. At the input stage, fixed Sobel kernels extract horizontal and vertical gradients, providing the network with enriched edge information and improved low-level features, which are then concatenated with the original image channels and further processed by convolutional layers. The U-Net encoder-decoder backbone, enhanced with parameter-free SimAM attention modules, enables adaptive feature weighting at every level, while a transformer bridge at the bottleneck efficiently captures long-range dependencies and global context. Unlike approaches that rely on direct skip connections from the raw input, SETAU-Net propagates rich multi-scale information through its skip connections, facilitating the reconstruction of high-quality, detail-preserving, and color-corrected images suitable for real-time, resource-constrained deployments in autonomous underwater systems. In parallel, LEARN (Laplacian Enhanced Attention and Residual Network) is designed for low-light image enhancement and employs a Laplacian enhancement module to extract and amplify high-frequency edge details, which are then adaptively scaled and fused with the original image features. Its encoder-decoder structure, built with expanded  $5 \times 5$  kernel residual blocks, captures broader contextual information, while Convolutional Block Attention Modules (CBAM) provide sequential channel and spatial attention to selectively emphasize important structures and suppress noise. Skip connections ensure efficient propagation of spatial details, resulting in enhanced images that maintain natural color and structural fidelity. LEARN's lightweight and computationally efficient design enables real-time performance, making it well-suited for applications such as surveillance, autonomous vehicles, and medical imaging. Together, these models demonstrate how the integration of edge-aware modules, attention mechanisms, and efficient architectural principles can deliver robust, high-quality image enhancement across diverse and challenging environments.

**Table 2.1.** Overview of Traditional and Machine-Learning Approaches to Low Light Image Enhancement

Author(s)	Paper Reference	Publication/ Proceeding	Year	Advantages	Limitations
Land, E. H.	[5]	Scientific American	1977	Presented the Retinex hypothesis; a seminal contribution to illumination-reflectance decomposition.	Assumes corruption-free images; prone to color distortion and noise amplification.
Pisano, E. D., et al.	[4]	Journal of Digital imaging	1998	CLAHE enhances visibility and contrast in low-light photographs proficiently.	Enhances noise; generates artificial outcomes in extreme low-light environments.
Abdullah-Al-Wadud, M., et al.	[1]	IEEE transactions on consumer electronics	2007	Dynamic histogram equalization adaptively enhances contrast under fluctuating lighting conditions.	Suboptimal performance in extreme low-light conditions; susceptible to over-enhancement artifacts.
Celik, T., & Tjahjadi, T.	[2]	IEEE Transactions on Image Processing	2011	Provides better local contrast enhancement by incorporating contextual information.	Enhances local contrast more effectively by integrating contextual information.
Cheng, H. D., & Shi, X. J.	[3]	Digital signal processing	2004	Straightforward execution; efficient under moderate illumination conditions.	Limited adaptation to various lighting conditions; exacerbates loudness in less lit areas.
Wei, C., Wang, W., Yang, W., & Liu, J.	[6]	arXiv preprint	2018	Thorough separation of illumination and reflection; enhanced visibility in dim lighting conditions.	Inadequately mitigates noise; resource-intensive.
Li, M., Liu, J., Yang, W., Sun, X., & Guo, Z.	[14]	IEEE transactions on image processing	2018	Integrates a noise map into the Retinex model, substantially diminishing noise while maintaining structural features.	Significant computational expense; potential for feature loss in low-noise photos; necessitates human parameter adjustment.

Zhang, Y., Zhang, J., & Guo, X.	[12]	Proceedings of the 27th ACM international conference on multimedia	2019	Functional and lightweight design; efficient for relatively low-light conditions.	Suboptimal performance in extreme low-light environments; difficulties with noise reduction.
Chen, C., Chen, Q., Do, M. N., & Koltun, V.	[10]	Proceedings of the IEEE/CVF International conference on computer vision	2019	High-quality enhancement of extreme low-light images; effective noise suppression.	Demands specialist gear (e.g., Sony $\alpha$ 7SII camera); incurs substantial computational expenses.
Guo, C., Li, C., Guo, J., Loy, C. C., Hou, J., Kwong, S., & Cong R.	[8]	Proceedings of the IEEE/CVF conference on computer vision and pattern recognition	2020	Zero-reference training removes reliance on paired data; the lightweight design is appropriate for deployment.	Restricted generalization in varied lighting situations; has difficulties with significant noise or artifacts.
Zeng, H., Cai, J., Li, L., Cao Z., & Zhang L.	[20]	IEEE Transactions on Pattern Analysis and Machine Intelligence	2020	Real-time performance utilizing a lightweight architecture; augmentation that adapts to images.	Constrained to particular image formats; encounters difficulties in extreme low-light conditions.
Lim S., & Kim W.	[16]	IEEE Transactions on Multimedia	2020	Multi-scale processing using the Laplacian pyramid facilitates efficient global illumination recovery and the preservation of local details.	Computational complexity arising from multi-scale processing; restricted real-time application.
Jiang Y., Gong X., Liu D., Cheng Y., Fang C., Shen X.	[7]	IEEE transactions on image processing	2021	Unpaired training eliminates the necessity for paired datasets, yielding aesthetically acceptable outcomes in the majority of instances.	Generates color aberrations under high situations; has difficulties in noise reduction.
Liu R., Ma L., Zhang J.	[19]	Proceedings of the IEEE/CVF conference on computer vision and pattern recognition	2021	Efficient unrolling methodology grounded in Retinex theory; enhanced noise management and detail retention.	Resource-intensive; requires meticulous parameter optimization for optimal outcomes.
Cai Y., Bian H.	[15]	Proceedings of the IEEE/CVF	2023	Illumination-guided transformer models	Elevated computational

		international conference on computer vision		efficiently capture non-local interactions and greatly surpass benchmarks.	demands stemming from transformer construction; intricate design constrains deployment efficacy.
--	--	---	--	--	--

**Table 2.2.** Overview of Traditional and Machine-Learning Approaches to Underwater Image Enhancement

Author(s)	Paper	Publication/Proceeding	Year	Key Contributions
Ronneberger, O., Fischer, P., & Brox, T.	[62]	Medical Image Computing and Computer-Assisted Intervention	2015	Proposed U-Net, an encoder-decoder architecture featuring skip connections, designed for accurate biomedical segmentation. Demonstrated effectiveness with constrained data through elastic augmentation, attaining state-of-the-art results in ISBI challenges.
Ghani, A. S. A., & Isa, N. A. M.	[30]	Applied Soft Computing	2015	Proposed UIE combining an integrated color model with histogram modification based on Rayleigh distribution to enhance contrast and visual quality.
Ancuti, C. O., Ancuti, C., De Vleeschouwer, C., & Bekaert, P.	[41]	IEEE Transactions on Image Processing	2017	Proposed a fusion-based UIE approach. Combines a color-corrected version and a contrast-enhanced version using perceptual weight maps (luminance, chromaticity, saliency).
Huang, D., Wang, Y., Song, W., Sequeira, J., & Mavromatis, S.	[56]	MultiMedia Modeling Conference	2018	Proposed Relative Global Histogram Stretching (RGHS) for shallow-water images. Adaptively stretches Green/Blue channel histograms based on distribution & light absorption.
Islam, M. J., Xia, Y., & Sattar, J.	[28]	IEEE Robotics and Automation Letters	2020	Proposed FUnIE-GAN, a real-time conditional GAN for UIE that utilizes a multi-modal perceptual loss. The EUVP dataset was introduced, demonstrating enhanced perception capabilities.

Islam, M. J., Luo, P., & Sattar, J.	[29]	Robotics: Science and Systems	2020	Introduced the simultaneous enhancement & super-resolution (SESR) task. Proposed Deep SESR network utilizing multi-modal loss and saliency guidance. The UFO-120 dataset for SESR has been released.
Yang, L., Zhang, R.-Y., Li, L., & Xie, X.	[54]	International Conference on Machine Learning	2021	Proposed SimAM, a simple, parameter-free 3D attention (channel & spatial) module. Employs a neuroscience-inspired energy function and a closed-form solution to allocate distinct neuron weights.
Tang, Y., Iwaguchi, T., Kawasaki, H., Sagawa, R., & Furukawa, R.	[52]	Asian Conference on Computer Vision	2022	Utilized Neural Architecture Search (NAS) to automatically identify optimal U-Net configurations for UIE. Transformers were included in the search space, demonstrating their suitability for high-level features.
Peng, L., Zhu, C., & Bian, L..	[27]	IEEE Transactions on Image Processing	2023	Introduced the Transformer architecture to UIE, incorporating specialized channel and spatial modules. A novel multi-color space loss is proposed, along with the release of the LSUI dataset.
Khan, R., Mishra, P., Mehta, N., Phutke, S. S., Vipparthi, S. K., Nandi, S., & Murala, S.	[58]	IEEE/CVF Winter Conference on Applications of Computer Vision	2024	Proposed Spectroformer, a multi-domain (spatial/frequency) query cascaded Transformer for UIE. Uses MQCA, Spatio-Spectro Fusion Attention & Hybrid Fourier-Spatial Upsampling blocks.
Pucci, R., & Martinel, N.	[60]	arXiv preprint	2024	Proposed CE-VAE for the simultaneous compression and enhancement of underwater images. Employs an attention-aware encoder and an innovative dual decoder, comprising spatial and capsule-based components, to achieve state-of-the-art results with compression.

## 2.2 Datasets

High-quality datasets are fundamental to the effectiveness of any deep learning-based image enhancement or restoration model. In this thesis, the careful selection and utilization of benchmark datasets play a pivotal role in developing, training, and validating the proposed frameworks for low-light and underwater image enhancement. These datasets not only provide diverse and challenging real-world scenarios but also serve as standardized benchmarks for evaluating and comparing the performance of different models. For low-light image enhancement, datasets such as LOLv1 [6], LOLv2-Real [9], LOLv2-Synthetic [9], and SID [10] offer paired images captured under varying illumination and noise conditions, ensuring that the models are exposed to a broad spectrum of real and synthetic low-light environments. For underwater image enhancement, datasets like LSUI [27], EUVP [28], and UFO-120 [29] encompass a wide range of underwater scenes with varying degrees of color distortion, contrast loss, and detail degradation. The use of these comprehensive datasets is crucial for achieving robust model generalization and for facilitating meaningful comparisons with state-of-the-art approaches in the field. The datasets employed in this research are summarized below, each contributing to the thorough assessment and advancement of deep learning-based image enhancement techniques under challenging visual conditions.

**Low Light Image Enhancement Datasets:** The testing and training of LEARN utilizes LOLv1 [6], LOLv2-Real [9], LOLv2-Synthetic [9], and SID [10] datasets. The LOLv1 [6] dataset comprises 500 paired low/normal-light images, split into 485 pairs for training and 15 pairs for testing, while LOLv2-Real [9] and LOLv2-Synthetic [9] datasets each contain 689 training pairs and 100 test pairs, offering greater diversity. For the SID [10] dataset, we use a subset of the Sony  $\alpha 7SII$  camera data, selecting only the highest exposure time from the available 10 exposure settings for each low-light scene, paired with the corresponding ground truth image. LOLv1 [6] consists of indoor scenes shot under different ISO levels and exposure times. These pictures are a basic benchmark for improvement models since they feature natural noise artifacts common in low-light photography. Using a three-step shooting approach to average several normal-light photographs for high-quality ground truth creation, LOLv2-Real [9] guarantees alignment and lowers artifacts like motion blur or camera shake, therefore offering increased variability. LOLv2-Synthetic [9] creates low-light photographs by artificially darkening normal-light ones, therefore enabling controlled research of degradation effects and illumination matching genuine dark photography. These datasets taken together offer a complete framework for assessing models in both natural and synthetic low-light environments, therefore allowing strong generalization to many illumination conditions.

**Underwater Image Enhancement Datasets:** SETAU-Net utilizes three benchmark datasets common in underwater image enhancement research; EUVP [28], LSUI [27], and UFO-120 [29] for supervised training and assessment. The training phase utilizes a mix of EUVP [28], LSUI [27], and UFO-120 [29]. Training uses the "Underwater Scenes" subset of the EUVP [28] dataset, which contains 2,185 pairs of poor and good perceptual quality images from seven camera types in diverse oceanic locations and visibility conditions to ensure representativeness of real-world robotic deployments. For scene diversity and high-quality references, the LSUI dataset [27] collected real-world underwater images and generated high-quality reference images through automated enhancement, objective filtering, and multiple rounds of human perceptual rating and refinement. It provides 3,423 pairs from compilation to our training split. Our training split includes 1,500 training pairs from the UFO-120 dataset [29], which uses domain transfer techniques to imitate damaged images and is utilized for UIE, super-resolution, and salient object recognition. PyTorch's 'ConcatDataset' [18] combines these three training sets to create a 7,108-image-pair training pool that exposes the model to more degradation



patterns. EUVP [28] (515 pairs), LSUI [27] (400 pairs), and UFO-120 [29] (120 pairs) test partitions are used to evaluate performance of the model.

**Table 2.3.** Common Datasets used in Enhancement and restoration of Adverse (Low-Light and Underwater) Images

Dataset	Number of Images/Image pairs	Year Released
LOL-v1 [6]	500 (485 train, 15 test) pairs	2018
LOL-v2 [9]	789 (689 train, 100 test) pairs	2024
SID [10]	5094 raw images	2018
LSUI [27]	4279 pairs	2023
EUVP [28]	12000 pairs	2020
UFO-120 [29]	120 pairs	2021

All training and testing datasets use a standard input dimension of  $256 \times 256$  pixels to preserve detail while optimizing computing performance. These datasets collectively provide a robust foundation for training and evaluating deep learning models for both low-light and underwater image enhancement, ensuring the models are tested on a wide variety of challenging real-world conditions.

### 2.3 Problem statement

The accurate and timely enhancement and restoration of images captured under adverse visual conditions, such as low-light and underwater environments, is essential for a wide range of real-world applications-including autonomous vehicles, surveillance, medical diagnostics, and marine exploration. Images acquired in these challenging settings are frequently affected by severe degradations, including reduced visibility, color distortion, amplified noise, blurring, and significant loss of structural detail. These issues not only hinder human interpretation but also compromise the performance of automated computer vision systems. Existing traditional and machine learning techniques for adverse (low-light and underwater images) image enhancement and restoration methods are limited by their tendency to introduce artifacts, unnatural color shifts, and their inability to generalize across diverse and complex scenarios. While deep learning has shown promise in overcoming some of these limitations, existing models often demand large, high-quality paired datasets, are computationally intensive, or fail to balance the trade-off between enhancement quality and efficiency. Furthermore, real-world conditions bring additional challenges, such as variations in illumination, scene complexity, and the presence of multiple simultaneous degradations (e.g., both noise and color cast in underwater images), which can severely impact model accuracy and robustness.

There is, therefore, a pressing need for the development of robust, efficient, and generalizable deep learning-based frameworks that can effectively restore visibility, correct color, suppress noise, and preserve fine structural details in images captured under these adverse conditions. Such frameworks should be capable of real-time or near real-time deployment in resource-constrained environments, ensuring practical applicability across a variety of domains. The main objective of this study is to design and validate lightweight deep learning architectures that address these challenges, advancing the state of the art in image enhancement and restoration for low-light and underwater scenarios.

## CHAPTER 3

### PROPOSED METHODOLOGY

This section details the proposed methodology for image enhancement and restoration under challenging visual conditions using deep learning. In this research, we utilize several benchmark datasets, including LOLv1 [6], LOLv2-Real [9], LOLv2-Synthetic [9], SID [10] for low-light images, and LSUI [27], EUVP [28], and UFO-120 [29] for underwater images, to comprehensively train and evaluate our models. Two state-of-the-art architectures are developed: LEARN, which employs a Laplacian enhancement module, expanded kernel residual blocks, and Convolutional Block Attention Modules (CBAM) for low-light image enhancement; and SETAU-Net, which integrates Sobel-based edge extraction, SimAM attention modules, and a transformer bridge within a U-Net backbone for underwater image enhancement. Figures 3.1 and 3.2 illustrate the respective architectures of the proposed frameworks. The following sub-sections provide a detailed discussion of data acquisition, pre-processing, network architecture, as well as model training and validation procedures.

**3.1 LEARN (Laplacian Enhanced Attention and Residual Network):** In this section, we present LEARN (Laplacian Enhanced Attention and Residual Network), our proposed approach for low-light image enhancement. LEARN uses a small yet powerful architecture to solve the basic problems of maintaining fine details while enhancing visibility in low-light photos. Our model achieves competitive performance while requiring significantly fewer computational resources than existing approaches. The methodology combines three key components: a Laplacian enhancement module for edge preservation, residual blocks for contextual information extraction, and convolutional block attention modules (CBAM) for adaptive feature refinement.

#### 3.1.1 Architecture

The LEARN architecture (Fig. 3.1.) follows an encoder-decoder structure optimized for low light image enhancement. The network processes RGB input images of dimensions height  $\times$  width  $\times$  3 channels ( $H \times W \times 3$ ) through three main components. First, our Laplacian enhancement module applies a fixed  $3 \times 3$  Laplacian kernel  $\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$  to extract edge information. This discrete approximation of the Laplacian operator highlights rapid intensity changes in the image. A learnable scaling factor  $\alpha$ , initialized to 0.1, refines this enhancement by amplifying edge details while preserving overall structure.

By use of its specific Laplacian enhancement module in combination with enlarged kernel residual blocks, the LEARN architecture improves low-light images. Following a learnable scaling factor and  $1 \times 1$  convolution to adaptively enhance these features, the Laplacian module detects edges and high-frequency details using a fixed  $3 \times 3$  Laplacian kernel. Larger  $5 \times 5$  kernel residual blocks with more context information complement this. Together, they provide a potent mix in which the Laplacian module especially focuses edge preservation and detail enhancement, producing sharper, more detailed outputs; the residual blocks extract rich features and preserve spatial information through skip connections.

The encoder consists of three convolutional layers: an initial  $3 \times 3$  convolutional layer producing 64 feature maps, subsequent to two downsampling layers with a stride of 2 that produce 128 and 256 feature maps respectively. We use  $5 \times 5$  kernels in our residual blocks instead of the standard  $3 \times 3$  kernels to increase the receptive field by 25%, allowing the network to capture broader contextual information. The middle section contains two residual blocks, each with two  $5 \times 5$  convolutional layers maintaining 256 feature channels. Each residual block follows the structure  $F(x) + x$ , where residual mapping  $F(x)$  represents the

function to be learned by the network. This skip-connection architecture mitigates the problem of vanishing gradients and improves the training efficiency of deeper networks.

Each residual block is integrated with CBAM modules that apply sequential channel and spatial attention. The CBAM initially establishes a one-dimensional channel attention map  $M_c \in \mathbb{R}^{C \times 1 \times 1}$  and then a two-dimensional spatial attention map  $M_s \in \mathbb{R}^{1 \times H \times W}$ . The channel attention module exploits both max-pooled and average-pooled features to compute channel-wise attention, while the spatial attention module uses concatenated average and max pooling features to focus on meaningful regions in the image. This dual attention method enables the network to focus selectively on relevant features while reducing noise.

The Convolutional Block Attention Module (CBAM) significantly enhances feature refinement in low-light image enhancement by implementing a dual attention mechanism. CBAM sequentially applies channel attention and spatial attention to hone in on "what" and "where" is important in the feature maps. This focused attention method lets the network dynamically highlight vital features while suppressing less important ones, hence enhancing low-light areas and maintaining natural look.

The decoder mirrors this structure with two transposed convolutional layers (kernel=4, stride=2) that upsample from  $256 \rightarrow 128 \rightarrow 64$  channels, followed by a final  $3 \times 3$  convolutional layer that produces the enhanced RGB output. Skip connections transfer feature maps from encoder to decoder through addition operations. After decoding, The Laplacian enhancement module sharpens edge details before passing the final output through a sigmoid activation function, ensuring pixel values remain within a valid range.

The LEARN model employs two primary activation functions, ReLU and Sigmoid placed throughout the architecture. ReLU (Rectified Linear Unit) activation function is used in the encoder, decoder, and residual blocks to incorporate non-linearity and enable the network to learn intricate patterns. ReLU is specifically utilized following each convolutional layer in the encoder, after the transposed convolution layers in the decoder, and within the residual blocks to activate intermediate feature mappings. The sigmoid activation function is used in the final output ensuring compatibility with image data formats. These activations are chosen to balance efficient training with stable output generation.

LEARN demonstrates a lightweight architecture that enables excellent performance in image enhancement while highlighting critical trade-offs. To reduce the number of parameters, it employs a simplistic encoder-decoder architecture. To compensate for this, the leftover blocks employ bigger  $5 \times 5$  kernels to acquire more contextual information without adding depth. By allocating processing power to key regions instead of treating the entire image equally, convolutional back-projection with attention mechanisms (CBAM) enables effective feature refining. The dedicated Laplacian improvement module focuses on edge details with minimal parameter increase, and skip connections preserve details without adding parameters. By carefully considering these factors, LEARN was able to keep its computing requirements low enough for real-world applications while still achieving competitive performance on enhancement tasks.

### 3.1.2 Training configuration

- **Loss function-** LEARN is trained using a combined loss function with three components: VGG perceptual loss, L1 loss, and SSIM loss (window size = 11,  $C_1 = 0.01^2$ ,  $C_2 = 0.03^2$ ) with 60%, 20% and 20% weightage respectively. The VGG perceptual loss uses features from a pre-trained VGG19 network to maintain natural visual characteristics, while L1 loss measures pixel-level accuracy and SSIM preserves structural information.

- **Parameters-** LEARN employs the AdamW optimizer for training, with a  $1e-4$  learning rate and  $1e-2$  weight decay. A cosine annealing scheduler gradually reduces the learning rate from the initial value  $1e-4$  to  $1e-6$  over 50 epochs. Training is conducted by leveraging a batch size of 16, where input images are resized to  $256 \times 256$  pixels using LANCZOS resampling to maintain consistency across datasets. For the Laplacian enhancement module, the scaling factor  $\alpha$  is initialized to 0.1. The CBAM module uses a reduction ratio of 8 for the channel attention mechanism and a  $7 \times 7$  kernel for spatial attention.
- **Datasets and data augmentation-** Proposed model is trained on LOLv1 [6], LOLv2-Real [9], LOLv2-Synthetic [9], and SID [10] datasets. The LOLv1 [6] dataset comprises 500 paired low/normal-light images, split into 485 pairs for training and 15 pairs for testing, while LOLv2-Real [9] and LOLv2-Synthetic [9] datasets each contain 689 training pairs and 100 test pairs, offering greater diversity. For the SID [10] dataset, we use a subset of the Sony  $\alpha 7SII$  camera data, selecting only the highest exposure time from the available 10 exposure settings for each low-light scene, paired with the corresponding ground truth image. To improve generalization and expand our training dataset, we apply several data augmentation techniques including horizontal flipping (50% probability), vertical flipping (25% probability), and slight Gaussian noise addition ( $\sigma=0.01$ ). For SID [10] images specifically, we apply additional transformations including random rotation ( $\pm 10^\circ$ ), limited random cropping (90-95% of original size), and occasional light smoothing with Gaussian blur (radius=1, 50% probability). LOLv1 [6] consists of indoor scenes shot under different ISO levels and exposure times. These pictures are a basic benchmark for improvement models since they feature natural noise artifacts common in low-light photography. Using a three-step shooting approach to average several normal-light photographs for high-quality ground truth creation, LOLv2-Real [9] guarantees alignment and lowers artifacts like motion blur or camera shake, therefore offering increased variability. LOLv2-Synthetic [9] creates low-light photographs by artificially darkening normal-light ones, therefore enabling controlled research of degradation effects and illumination matching genuine dark photography. These datasets taken together offer a complete framework for assessing models in both natural and synthetic low-light environments, therefore allowing strong generalization to many illumination conditions.
- **Implementation-** LEARN is programmed and executed in PyTorch [18] and trained on an NVIDIA GeForce RTX 4060 GPU (8GB GDDR6) paired with a 13th Gen Intel Core i7-13700HX 2.10 GHz processor with 16 GB RAM. The training process takes approximately 2-6 minutes per epoch. Mixed precision training is employed using PyTorch [18], which accelerates computation and reduces memory usage by performing certain operations in FP16 while maintaining stability for others in FP32. The model uses batch normalization layers where applicable, ensuring stable training dynamics. Skip connections are implemented as addition operations to preserve spatial information while minimizing memory overhead. The training setup includes a combined dataset loader that merges multiple datasets (LOLv1 [6], LOLv2-Real [9], LOLv2-Synthetic [9], and SID [10]) into a unified pipeline. The best-performing weights are saved based on validation loss during model checkpointing to ensure optimal performance and allow training to be continued if interrupted. All training artifacts, including ideal and final model weights, are kept in a directory for reproducibility.

### 3.1.3 Model Operation

1. LEARN architecture integrates multiple interrelated components to improve visibility and preserve details in low-light images. The encoder captures hierarchical features through three convolutional layers to start the model. The middle phase refines features using residual blocks and Convolutional Block Attention Modules (CBAM). The decoder

reconstructs the augmented image using transposed convolutions and skip connections to restore spatial features lost during downsampling. After reconstructing features, the Laplacian Enhancement module gathers edge information and refines fine details before output.

2. The image enters the encoder path consisting of three convolutional layers. First, a  $3 \times 3$  convolution transforms the input into 64 feature maps:

$$F_1 = \text{Conv}_{3 \times 3}(I_{\text{enhanced}}) \quad (1)$$

3. This operation applies 64 different  $3 \times 3$  filters to the enhanced image, with each filter learning to detect specific patterns. The calculation for each output pixel in feature map 'j' at position (x, y) is:

$$F_1^j(x, y) = \sum_{i=0}^2 \sum_{m=0}^2 \sum_{n=0}^2 W_{m,n,i,j} * I_{\text{enhanced}}(x + m - 1, y + n - 1, i) + b_j \quad (2)$$

where W represents the learnable weights and  $b_j$  is the bias for feature map 'j'.

4. Next, two downsampling convolutions with a stride of 2 reduce spatial dimensions while increasing feature depth:
  - i.  $F_2 = \text{Conv}_{3 \times 3, \text{stride}=2}(F_1)$  producing 128 feature maps at half resolution.
  - ii.  $F_3 = \text{Conv}_{3 \times 3, \text{stride}=2}(F_2)$  producing 256 feature maps at quarter resolution.
5. The encoder's output  $F_3$  enters the middle section containing two residual blocks with CBAM modules. Each residual block first processes features through two consecutive  $5 \times 5$  convolutional layers:

$$F_{\text{mid}} = \text{Conv}_{5 \times 5}(\text{ReLU}(\text{Conv}_{5 \times 5}(F_{\text{in}}))) \quad (3)$$

These larger  $5 \times 5$  kernels expand the receptive field compared to standard  $3 \times 3$  convolutions, allowing the network to incorporate broader spatial context. The residual connection then adds the original input to these processed features:

$$F_{\text{res}} = F_{\text{in}} + F_{\text{mid}} \quad (4)$$

6. After the residual connection, the CBAM module applies sequential channel and spatial attention. For channel attention, it computes:

$$M_c = \sigma(\text{MLP}(\text{AvgPool}(F_{\text{res}})) + \text{MLP}(\text{MaxPool}(F_{\text{res}}))) \quad (5)$$

where average-pooled and max-pooled features are processed through shared MLPs with reduction ratio 8, then combined and activated with sigmoid  $\sigma$ . This produces a channel attention map that scales each feature channel based on importance. Spatial attention follows with:

$$M_s = \sigma(\text{Conv}_{7 \times 7}(\text{AvgPool}(F_{\text{res}} \otimes M_c) \oplus \text{MaxPool}(F_{\text{res}} \otimes M_c))) \quad (6)$$

where channel-refined features are pooled across channels, concatenated (denoted by  $\oplus$ ), processed by a  $7 \times 7$  convolution, and activated with sigmoid to generate a spatial attention map highlighting important regions. The final output of each residual block with CBAM is  $F_{\text{out}} = (F_{\text{res}} \otimes M_c) \otimes M_s$ , where  $\otimes$  represents element-wise multiplication.

7. After the middle section, the decoder path reconstructs the enhanced image by progressively upsampling the refined features and incorporating details via skip

connections. The output of the middle section,  $F_{\text{out}}$ , is passed through a transposed convolution layer to double its spatial dimensions and reduce its channel count from 256 to 128:

$$D_1 = \text{TransConv}_{4 \times 4, \text{stride}=2}(F_{\text{out}}) \quad (7)$$

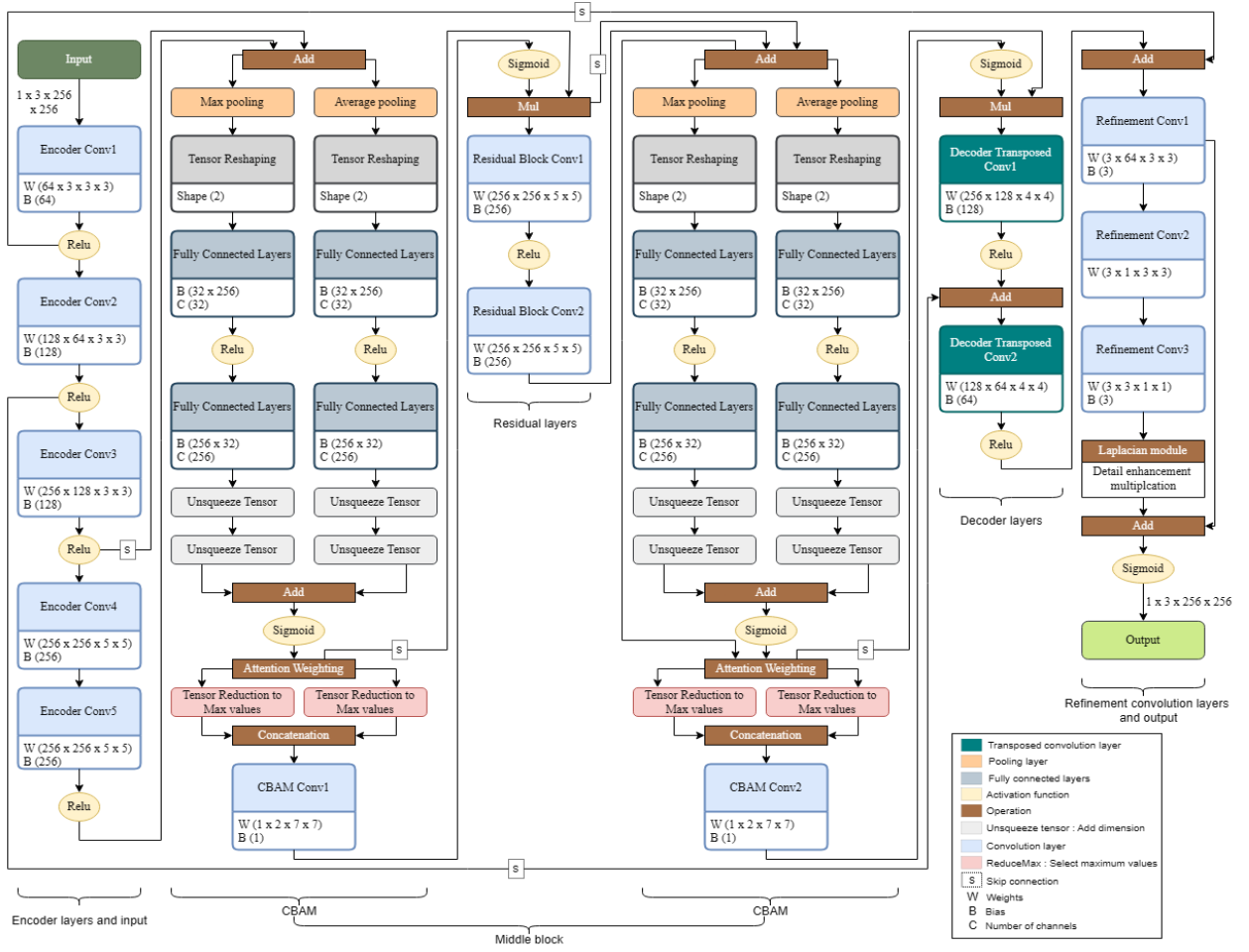
8. A skip connection adds the corresponding encoder feature map  $F_2$  to preserve spatial details:  $D_1 = D_1 + F_2$ . Next,  $D_1$  is passed through another transposed convolution to further double its spatial dimensions and reduce its channel count from 128 to 64:

$$D_2 = \text{TransConv}_{4 \times 4, \text{stride}=2}(D_1) \quad (8)$$

9. Another skip connection adds the feature map  $F_1$  (from the first encoder layer):  $D_2 = D_2 + F_1$ . Finally,  $D_2$  is passed through a  $3 \times 3$  convolutional layer to produce three output channels (RGB)  $D_3 = \text{Conv}_{3 \times 3}(D_2)$ . This completes the decoder path, producing an intermediate enhanced image that retains both global context and local details.
10. After the decoder path reconstructs intermediate features, the final enhanced image is generated. The reconstructed features,  $D_3$ , are passed through the Laplacian enhancement module to refine edge details. The Laplacian Enhancement module is applied at the end of the decoder path to refine edge details before producing the final enhanced output. This module uses a fixed  $3 \times 3$  Laplacian kernel to extract edge information and applies a learnable scaling factor  $\alpha$  for enhancement. The Laplacian operator is defined as:

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

The enhanced features are computed as:  $I_{\text{laplacian}} = \text{Conv}_{1 \times 1}(\text{Laplacian}(D_3))$  where  $D_3$  is the output from the decoder's final convolutional layer. The final enhanced image is then calculated as  $I_{\text{enhanced}} = D_3 + \alpha \cdot I_{\text{laplacian}}$ ; Here,  $\alpha$  scales the effect of the Laplacian augmentation, allowing the network to learn the amount of edge refinement required. Finally, a sigmoid activation function is employed to constrain pixel values within the range of 0 to 1:  $I_{\text{output}} = \sigma(I_{\text{enhanced}})$ . This results in a final enhanced RGB image with increased visibility and preserved edge details.



**Fig. 3.1.** LEARN Model Architecture Diagram

**Table 3.1.** LEARN Model Configuration

Layer	Output channels	Kernel size	Stride	Padding	Activation	Parameters	Description
2D Convolutional Layer (encoder convolution 1)	64	3x3	1	1	ReLU	1,792	Initial feature extraction
2D Convolutional Layer (encoder convolution 2)	128	3x3	2	1	ReLU	73,856	Downsampling and feature expansion
2D Convolutional Layer (encoder convolution 3)	256	3x3	2	1	ReLU	295,168	Further downsampling and feature expansion
Residual Block (residual block 1)	256	5x5	1	2	ReLU	3,277,312	Residual learning with expanded receptive field

Convolutional Block Attention Module (attention module 1)	256	-	-	-	Sigmoid	16,771	Channel and spatial attention refinement
Residual Block (residual block 2)	256	5x5	1	2	ReLU	3,277,312	Second residual block for deep feature extraction
Convolutional Block Attention Module (attention module 2)	256	-	-	-	Sigmoid-	16,771	Additional attention refinement
2D Transposed Convolutional Layer (decoder convolution 1)	128	4x4	2	1	ReLU	524,416	Upsampling and feature reduction
2D Transposed Convolutional Layer (decoder convolution 2)	64	4x4	2	1	ReLU	131,136	Further upsampling and feature reduction
2D Convolutional Layer (decoder convolution 3)	3	3x3	1	1	-	1,731	Final reconstruction to image space
Laplacian Enhancement Module	3	3x3	1	1	-	40	Edge enhancement using Laplacian operator

**3.2 SETAU-Net** - This section presents the proposed SETAU-Net, an innovative hybrid network architecture specifically developed for the complex job of Underwater Image Enhancement (UIE). SETAU-Net integrates the advantages of Convolutional Neural Networks (CNNs) for hierarchical feature extraction and spatial detail retention with the global context modeling skills of Transformers, particularly designed for efficiency and efficacy in the underwater domain. The design has a preliminary edge-enhancement phase, attention (self-attention and SimAM) techniques integrated into its basic framework, and a Transformer-based bottleneck bridge.

### 3.2.1 Architecture

The proposed SETAU-Net (Fig. 3.2) utilizes a U-Net design, with an encoder pathway that captures hierarchical information and a symmetric decoder pathway for precise localization and reconstruction, interconnected by skip connections. This architecture is meticulously crafted for the complexities of underwater image enhancement (UIE) through several key



innovations: an initial Sobel-based edge enhancement stage, parameter-free SimAM attention integrated within the convolutional blocks, and a Transformer-based attention bridge at the network's bottleneck. The network is engineered to process a 3-channel input image  $X \in \mathbb{R}^{H \times W \times 3}$  and generate an enhanced 3-channel output  $I_{out} \in \mathbb{R}^{H \times W \times 3}$ . The network utilizes a base channel width, referred to as  $C_0$ , which dictates the number of feature channels following the first processing and establishes a foundation for the dimensions of succeeding layers; in our design, this base width is established at 40 channels. Figure 1 presents a schematic overview of the whole SETAU-Net architecture.

Processing begins with the Sobel Enhancement Module, designed to explicitly leverage structural information often degraded underwater. Instead of operating directly on the input image  $X$ , this module first computes horizontal ( $G_x$ ) and vertical ( $G_y$ ) gradients using fixed 3x3 Sobel filter kernels ( $K_x, K_y$ ) applied via depthwise convolution (groups =  $C_{in} = 3$ ) to preserve channel-specific edge details:

$$G_x = \text{Conv2d}(X, K_x, \text{padding} = 1, \text{groups} = 3) \quad (9)$$

$$G_y = \text{Conv2d}(X, K_y, \text{padding} = 1, \text{groups} = 3) \quad (10)$$

The original input  $X$  (3 channels) is then concatenated with both gradient maps ( $G_x$ : 3 channels,  $G_y$ : 3 channels) along the channel dimension, creating a 9-channel feature map  $X'_{in} = \text{Concat}(X, G_x, G_y)$ . This combined map, rich in both color and edge information, is subsequently processed by two sequential processing blocks. Each block comprises a 3x3 convolutional layer, succeeded by batch normalization, and culminates with a rectified linear unit (ReLU) activation function ( $\text{ReLU}(x) = \max(0, x)$ ). The resultant feature map  $E_0$  denotes the output where:

$$E_0 = \text{Block2}(\text{Block1}(X'_{in})) \quad (11)$$

$$\text{Block}(Z) = \text{ReLU}(\text{BatchNorm}(\text{Conv3x3}(Z))) \quad (12)$$

The rationale for this initial step is twofold: it injects strong structural priors derived from edges at minimal computational cost (using fixed Sobel filters), and the subsequent learnable convolutional layers allow the network to adaptively integrate this information, guiding feature extraction in the deeper layers. This module outputs an edge-aware feature map  $E_0$  with dimensions  $\mathbb{R}^{H \times W \times C_0}$ , having the base channel width of 40.

The resulting feature map  $E_0$  is then fed into the Encoder Path, which progressively reduces spatial resolution while increasing channel depth across four stages to learn multi-scale contextual representations. Each encoder stage utilizes computationally efficient depthwise-separable convolutions. Specifically, within each stage, a 3x3 depthwise convolution performs spatial filtering independently for each input channel. For downsampling, this depthwise convolution uses a stride of 2 in the second, third, and fourth encoder stages (producing  $E_2, E_3, E_4$ ). A further 1x1 pointwise convolution modifies the features to the designated output channel dimension, thus consolidating information across channels. Batch normalization is succeeded by the implementation of the Gaussian Error Linear Unit (GELU) activation function, expressed as  $\text{GELU}(x) = x\Phi(x)$ , where  $\Phi(x)$  symbolizes the CDF (Cumulative Distribution Function) of the normal distribution in its conventional form. A crucial element incorporated post-activation in each encoder step is the SimAM [54] (Simple, Parameter-free Attention Module). SimAM adaptively enhances features by computing attention weights using an energy function based on neuron data inside each channel, without adding further learnable parameters. The energy function for a target neuron  $t$  with respect to others  $x_i$  in a channel is:

$$e_t(w_t, b_t, y, x_i) = (y_t - \hat{t})^2 + \frac{1}{M-1} \sum_{i=1}^{M-1} (y_o - \hat{x}_i)^2 \quad (13)$$

where  $\hat{t}, \hat{x}_i$  are linear transformations and  $y_t = 1, y_o = -1$ ,  $M = H \times W$ . Minimizing this function yields attention weights  $\text{sigmoid}(E)$ , where  $E$  relates to the neuron's variance, effectively highlighting more informative features. Finally, the output of the SimAM module is combined with the original input to the stage ( $X$ ) via a residual connection; this input  $X$  may first be passed through a  $1 \times 1$  convolutional projection if its dimensions or channel count changed within the stage (e.g., due to striding). This sequential structure – depthwise convolution, pointwise convolution, normalization, activation, SimAM attention [54], and residual addition – promotes efficient feature extraction, incorporates adaptive feature refinement with minimal overhead, and ensures stable training. The encoder produces feature maps  $E_1, E_2, E_3, E_4$  with spatial dimensions reducing from  $H \times W$  to  $H/8 \times W/8$  and channel dimensions increasing progressively as multiples of the base width:  $C_0, 2C_0, 4C_0, 8C_0$  (corresponding to 40, 80, 160, and 320 channels).

At the U-Net's bottleneck, the most compressed feature map  $E_4$  (with dimensions  $\mathbb{R}^{H/8 \times W/8 \times 320}$ ) is processed by the Transformer Attention Bridge. This module substitutes normal convolutions to clearly represent global context and long-range dependencies, essential for tackling spatially variable degradation such as color casts in UIE. The bridge consists of a sequence of two identical Transformer layers. Each layer begins with Instance Normalization (often beneficial in generative tasks for normalizing style/contrast information per instance), followed by a multi-head scaled dot-product self-attention mechanism and a residual connection adding the layer's input to its output. The self-attention mechanism implements attention using four parallel attention heads. Input features  $X$  are linearly projected to Query (Q), Key (K), and Value (V) representations using  $1 \times 1$  convolutions. Attention is computed as:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (14)$$

where  $d_k$  is the dimension per head (80 in this configuration). This allows the module to weigh information globally based on feature similarity. Our implementation enhances this with a parallel path using a  $3 \times 3$  depthwise convolution (potentially aiding positional awareness) and incorporates an interaction term derived from the Query and Key projections, modulating the attention output before combining it with the depthwise path result. A final  $1 \times 1$  projection follows the attention block. Furthermore, the bridge incorporates multi-scale context aggregation: features after the Transformer layers are pooled to global ( $1 \times 1$ ) and medium ( $2 \times 2$ ) sizes using adaptive average pooling, processed with separate  $1 \times 1$  convolutions, bilinearly upsampled, and fused (via concatenation and a  $1 \times 1$  fusion convolution) with the main feature path. This design allows the bottleneck to effectively capture both global dependencies and summarized multi-scale context, outputting a refined bottleneck feature map  $B$  of size  $\mathbb{R}^{H/8 \times W/8 \times 320}$ . The refined bottleneck features  $B$  initiate the decoder path, which symmetrically mirrors the encoder to reconstruct the enhanced image. It comprises three stages. Each decoder stage receives input from the preceding stage (or the bridge  $B$  for the first stage) and the corresponding feature map from the encoder via a skip connection ( $E_3, E_2, E_1$  respectively). The incoming feature map  $X_{in}$  is first upsampled by a factor of 2 using a  $2 \times 2$  Transposed Convolution (ConvTranspose2d) with stride=2, providing a learnable upsampling mechanism. The crucial skip connection operation follows: the upsampled map is concatenated channel-wise with the feature map  $X_{skip}$  from the corresponding encoder level  $\text{Concat}(\text{Up}(X_{in}), X_{skip})$ . Padding is applied if spatial dimensions differ slightly. This concatenation ensures that high-resolution spatial details from the encoder are directly available during reconstruction. A  $1 \times 1$  fusion convolution subsequently integrates the

concatenated information and modifies the channel dimension. Each decoder stage has two successive blocks, each featuring a typical 3x3 convolution, succeeded by batch normalization and GELU activation. A SimAM module is incorporated post-convolutional blocks for parameter-free feature refining, analogous to the encoder. A residual link then adds the fused input feature map (after 1x1 fusion convolution) to the SimAM output, followed by a final GELU activation.

$$\left( \text{Output} = \text{Act}(\text{SimAM}(\text{ConvBlocks}(\text{FusedFeatures})) + \text{FusedFeatures}) \right) \quad (15)$$

The process wherein ConvBlocks denotes the series of Conv-BN-GELU operations and Act signifies the concluding GELU, enables the decoder to incrementally enhance features and augment spatial resolution while assimilating intricate details from the skip connections. The decoder stages produce feature maps  $D_3, D_2, D_1$  with spatial dimensions increasing from  $H/4 \times W/4$  back to  $H \times W$  and channel dimensions decreasing progressively:  $4C_0, 2C_0, C_0$  (corresponding to 160, 80, and 40 channels).

Finally, the output layer maps the high-resolution feature map  $D_1$  (shape  $\mathbb{R}^{H \times W \times 40}$ ) from the last decoder stage to the final enhanced image. This is achieved using a single 1x1 convolution layer that projects the 40 feature channels down to the required 3 output channels (RGB). A Sigmoid activation function,  $\sigma(x) = 1/(1 + e^{-x})$ , is applied element-wise to the output of the convolution. This ensures that the final pixel values of the enhanced image  $I_{\text{out}}$  are normalized to the standard range (0, 1), producing the final result  $I_{\text{out}} \in \mathbb{R}^{H \times W \times 3}$ .

### 3.2.2 Training configuration

- **Loss function-** We employ a composite loss function to direct the SETAU-Net model in producing improved images, trained in an end-to-end fashion. This function integrates many objectives, preserving balance among pixel-level reconstruction accuracy, structural integrity, perceptual realism, and edge clarity, all vital components for effective underwater image enhancement. The total loss ( $L_{\text{total}}$ ) is a weighted combination of five distinct loss components: L1 loss ( $(L_1)$ ), L2 loss (Mean Squared Error,  $(L_2)$ ), Structural Similarity Index Measure loss ( $L_{\text{SSIM}}$ ), VGG-based perceptual loss ( $L_{\text{VGG}}$ ), and Laplacian loss ( $L_{\text{Lap}}$ ). Each term addresses a distinct facet of image quality, with their contributions calibrated by predetermined weights established from the configuration:  $w_{L1} = 0.30, w_{L2} = 0.10, w_{\text{SSIM}} = 0.30, w_{\text{VGG}} = 0.15$  and  $w_{\text{Lap}} = 0.15$ . The entire loss is expressed as:

$$L_{\text{total}} = w_{L1}L_1 + w_{L2}L_2 + w_{\text{SSIM}}L_{\text{SSIM}} + w_{\text{VGG}}L_{\text{VGG}} + w_{\text{Lap}}L_{\text{Lap}} \quad (16)$$

The Mean Absolute Error (L1 loss), executed through 'nn.L1Loss', is incorporated to ensure pixel-level precision. It promotes the network output  $I_{\text{out}}$  to closely align with the ground truth  $I_{\text{gt}}$  on a pixel-by-pixel basis and is recognized for its relative resilience to outliers in comparison to L2. The Mean Squared Error (L2 Loss), executed through 'nn.MSELoss', enhances pixel-level accuracy while imposing greater penalties on bigger discrepancies compared to L1.

$$L_2 = \left\| I_{\text{out}} - I_{\text{gt}} \right\|_2^2 \quad (17)$$

Its incorporation, however with reduced significance, enhances L1 and constitutes a fundamental element in several image restoration endeavors. Solely utilizing pixel-level losses may result in indistinct outcomes. The SSIM loss is essential for maintaining structural information by comparing localized patterns of pixel intensities adjusted for brightness and contrast.

$$L_{SSIM} = 1 - SSIM(I_{out}, I_{gt}) \quad (18)$$

To enhance the output's conformity with human visual perception, we integrate a VGG16-based perceptual loss, executed within the VGG16PerceptualLoss class. By reducing the mean squared error between feature maps ( $\phi_i$ ) obtained from designated intermediate ReLU activation layers (indexed 2, 7, 12, 21) of a pre-trained VGG16 network (models.VGG16\_Weights.DEFAULT).

$$L_{VGG} = \sum_{i \in \{2, 7, 12, 21\}} \left\| \phi_i(I_{out}) - \phi_i(I_{gt}) \right\|_2^2 \quad (19)$$

This loss function promotes the generation of images with authentic textures and details, circumventing the excessively smooth outputs often linked to solely pixel-based or structural losses. To particularly mitigate the blurring frequently observed in underwater images, which may be intensified by enhancement networks, the Laplacian loss is incorporated.

$$L_{Lap} = \left\| \nabla^2 I_{out} - \nabla^2 I_{gt} \right\|_1 \quad (20)$$

By computing the L1 distance between the second-order derivatives of the output and target (approximated channel-wise using a 2D convolution with the Laplacian kernel and groups = 3), this loss explicitly encourages the preservation and reconstruction of sharp edges and fine details.

$$K_{Lap} = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

- **Parameters-** For training SETAU-Net, we employ the AdamW optimizer for better model generalization. We start AdamW with a learning rate of  $5 \times 10^{-4}$  to facilitate rapid convergence in the early phases of training. AdamW's decoupling mechanism penalizes excessive network weights with a weight decay coefficient of 0.01 to reduce overfitting. The 'ReduceLROnPlateau' scheduler was used to adjust the learning rate during training. It monitors the average total training loss and reacts dynamically to learning progress. If the loss does not improve after  $p=2$  consecutive epochs, the learning rate is reduced by  $\gamma=0.1$ . A minimum threshold of  $1 \times 10^{-7}$  was set to prevent stagnation in learning rate. The training took 35 epochs and a practical batch size of 12 was used to get a steady gradient estimate. The input photos are processed at a standard 256x256 pixel resolution, which balances detail retention and processing efficiency in image restoration jobs. The SimAM attention [54] modules uses a standard  $\lambda$  value of  $1 \times 10^{-4}$ , while the Transformer bridge uses a standard setup of 4 attention heads to evenly divide the bottleneck dimension (320) by the number of heads.
- **Datasets and data augmentation-** We utilize three benchmark datasets common in underwater image enhancement research: EUVP [28], LSUI [27], and UFO-120 [29] for supervised training and assessment of SETAU-Net. The training phase utilizes a mix of EUVP [28], LSUI [27], and UFO-120 [29]. Our training uses the "Underwater Scenes" subset of the EUVP [28] dataset, which contains 2,185 pairs of poor and good perceptual quality images from seven camera types in diverse oceanic locations and visibility conditions to ensure representativeness of real-world robotic deployments. For scene diversity and high-quality references, the LSUI dataset [27] collected real-world underwater images and generated high-quality reference images through automated enhancement, objective filtering, and multiple rounds of human perceptual rating and refinement. It provides 3,423 pairs from compilation to our training split. Our training

split includes 1,500 training pairs from the UFO-120 dataset [29], which uses domain transfer techniques to imitate damaged images and is utilized for UIE, super-resolution, and salient object recognition. PyTorch's 'ConcatDataset' [18] combines these three training sets to create a 7,108-image-pair training pool that exposes the model to more degradation patterns. EUVP [28] (515 pairs), LSUI [27] (400 pairs), and UFO-120 [29] (120 pairs) test partitions are used to evaluate performance of the model. All training and testing datasets use a standard input dimension of  $256 \times 256$  pixels to preserve detail while optimizing computing performance. During training, we used a data augmentation pipeline to increase SETAU-Net's resilience and generalization capabilities. In order to maintain correspondence, the same geometric and photometric transformations were applied to each input-target image pair in a probabilistic manner using the Albumentations library. The augmentation sequence had moderate color jittering (a 30% chance of changing brightness, contrast, saturation, and hue), infrequent random rotations within  $\pm 15$  degrees, and horizontal and vertical flips (each with a 50% chance). In order to assist the model develop invariance to frequent underwater photography situations including perspective shifts and color distortions, this technique incorporates variability in orientation and color. In order to guarantee a balanced mixture of original and augmented data in every training epoch, all augmentations were performed with an overall probability of 0.5.

- **Implementation-** SETAU-Net is implemented and performed with the PyTorch deep learning framework [18]. Training was performed on a machine including an NVIDIA GeForce RTX 4060 GPU (8GB GDDR6), complemented by a 13th Gen Intel Core i7-13700HX 2.10 GHz CPU and 16 GB of RAM. To improve training efficacy, mixed precision training is employed using PyTorch's 'torch.amp' and 'GradScaler' functionalities [18]. This technique improves computational efficiency and reduces GPU memory consumption by performing some operations in lower precision (FP16) while maintaining numerical stability for others in FP32. The model architecture incorporates Batch Normalization (nn.BatchNorm2d) inside its convolutional blocks and Instance Normalization (nn.InstanceNorm2d) in the Transformer bridge layers, therefore guaranteeing stable training dynamics and appropriate feature normalization for diverse module types. Skip connections, crucial to the U-Net design for preserving spatial information, are implemented using feature concatenation (torch.cat) in the decoder blocks, amalgamating feature maps from corresponding encoder stages. The training configuration employs a consolidated data pipeline formed by integrating the EUVP [28], UFO-120 [29] and LSUI [27] training datasets with PyTorch's 'ConcatDataset' [18]. Model checkpoints are regularly stored throughout training (e.g., after each epoch), preserving the model's state dictionary, optimizer and scheduler states, the current epoch number, and recent training loss. This facilitates the resumption of the training process if stopped, and the final model weights are preserved following the conclusion of the training epochs. All training artifacts, including saved checkpoints and final model weights, are preserved in a specified directory to ensure repeatability.

### 3.2.3 Model Operation

1. The SETAU-Net model analyzes an input RGB underwater image  $X \in \mathbb{R}^{3 \times H \times W}$  using a sequence of specialized modules tailored to tackle the distinct issues of underwater image enhancement. The procedure starts with the 'CNNSobelEdgeModule', which specifically collects edge information vital for recovering structural features frequently obscured in underwater environments. This module calculates horizontal and vertical gradients with fixed Sobel kernels  $K_x$  and  $K_y$  through depthwise convolution. Each encoder block first applies a  $3 \times 3$  depthwise convolution (optionally with stride 2 for downsampling),

followed by a  $1 \times 1$  pointwise convolution, batch normalization (BatchNorm2d), and GELU activation. The depthwise convolution performs spatial filtering autonomously for each input channel, while the pointwise convolution projects features to the desired output channel dimension. The stride is set to 2 in the second, third, and fourth encoder blocks to progressively downsample the spatial dimensions, providing multi-scale contextual representations.

2. The use of depthwise convolution with groups = 3 ensures that the Sobel filtering is applied independently to each color channel, preserving channel-specific edge information. This initial module injects crucial structural priors while avoiding computational overhead. The kernels  $K_x$  and  $K_y$  are fixed (non-trainable) and initialized as:

$$k_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad k_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

3. The original image  $X$ ,  $G_x$ , and  $G_y$  are concatenated along the channel dimension to create a 9-channel tensor. The result is then passed through two consecutive convolutional blocks (each block: Conv  $\rightarrow$  BN  $\rightarrow$  ReLU), producing an edge-aware feature map  $E_0 \in \mathbb{R}^{C_0 \times H \times W}$ , where  $C_0 = 40$ .
4. The encoder path consists of four sequential encoder blocks, each designed to extract increasingly abstract features while reducing spatial resolution and increasing channel depth. Each encoder block starts with a  $3 \times 3$  depthwise convolution (which can use a stride of 2 to reduce size), followed by a  $1 \times 1$  pointwise convolution, batch normalization, and GELU activation. The output is then enhanced by the SimAM [54] attention mechanism, which calculates channel-wise attention weights by a parameter-free energy function, where  $x$  represents the feature map,  $\mu$  denotes its mean,  $\sigma^2$  signifies its variance, and  $\lambda$  is a negligible constant. The attention weights are obtained via a sigmoid activation and multiplied elementwise with the features. A residual connection adds the (possibly projected) input to the SimAM-refined output, ensuring stable gradient flow. The encoder produces feature maps  $E_1, E_2, E_3, E_4$  with spatial sizes halved at each stage and channel sizes  $C_0, 2C_0, 4C_0, 8C_0$  (i.e., 40, 80, 160, 320). The attention weights are obtained via a sigmoid activation and multiplied elementwise with the features, adaptively scaling feature responses based on their importance.

$$E = \frac{(x-\mu)^2}{4(\sigma^2+\lambda)} + 0.5 \quad (21)$$

5. At the bottleneck, the most compressed feature map  $E_4$  is processed by the transformer bridge. This module consists of two sequential blocks, each applying instance normalization, followed by a self-attention mechanism and a residual connection. The ‘SelfAttention’ module computes multi-head self-attention as follows: input features are projected to queries  $Q$ , keys  $K$ , and values  $V$  using  $1 \times 1$  convolutions, then reshaped for  $h$  heads (here  $h = 4$ ).
6. The attention output is modulated by a learned interaction term and combined with a depthwise convolutional path. After both transformer blocks, multi-scale context is aggregated by global and medium adaptive average pooling, followed by  $1 \times 1$  convolutions, upsampling, and fusion with the main path. The following procedure generates a contextually enhanced bottleneck feature map  $B \in \mathbb{R}^{320 \times H/8 \times W/8}$ .

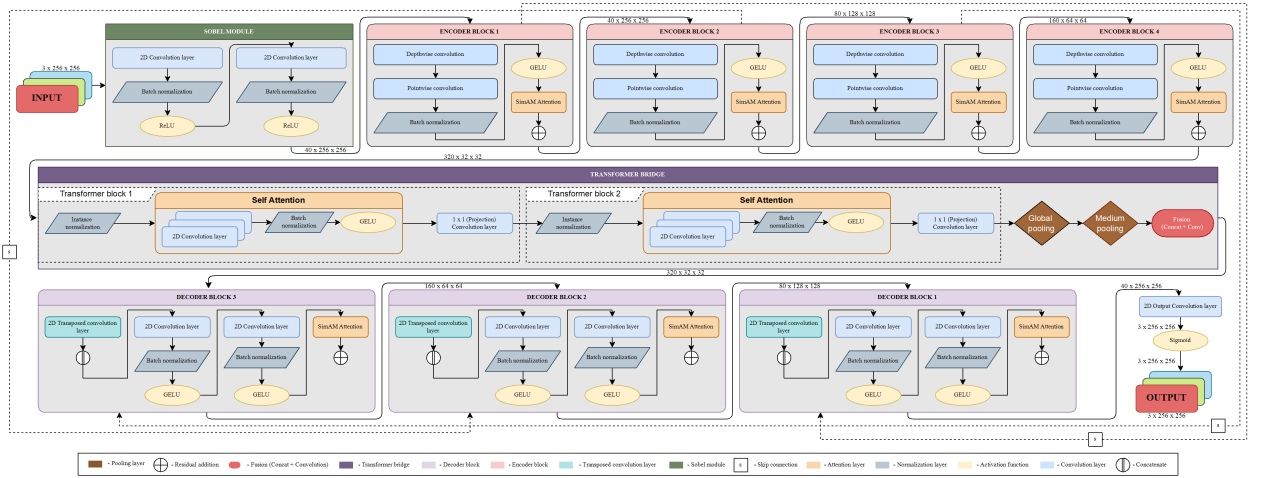
7. The decoder path parallels the encoder, recreating the image at ever greater resolutions. Each decoder block initially upsamples its input using a transposed convolution before concatenating it with the associated encoder feature map (skip connection). The integrated features are combined using a  $1 \times 1$  convolution, succeeded by two  $3 \times 3$  convolutions (each incorporating BatchNorm and GELU), SimAM attention [54], a residual connection, and a concluding GELU activation. This procedure is done over three phases, yielding feature maps  $D_3, D_2, D_1$  with channel dimensions  $4C_0, 2C_0, C_0$  (160, 80, 40) while reinstating the original spatial resolution. If the spatial dimensions do not match due to rounding issues during upsampling, padding is applied using ‘F.pad’.
8. Finally, the output layer employs a  $1 \times 1$  convolution to transform the  $C_0 = 40$  channels into 3 output channels (RGB), succeeded by a sigmoid activation.

$$I_{\text{out}} = \sigma(\text{Conv}_{1 \times 1}(D_1)), \quad \sigma(x) = \frac{1}{1+e^{-x}} \quad (22)$$

The above procedure guarantees that the improved image  $I_{\text{out}} \in \mathbb{R}^{3 \times H \times W}$  possesses pixel values throughout the interval  $[0,1]$ . The complete architecture is entirely differentiable and trained end-to-end, allowing the model to acquire both low-level and high-level characteristics essential for effective underwater image enhancement.

**Table 3.2.** SETAU-Net Architecture Configuration

Layer	Output channels	Kernel	Stride	Activation	Parameters	Description
Edge	40	3x3	1	ReLU	18,000	Edge enhancement using Sobel filters
Encoder 1	40	3x3	1	GELU	1,920	Initial feature extraction
Encoder 2	80	3x3	2	GELU	7,360	Downsampling and feature expansion
Encoder 3	160	3x3	2	GELU	28,800	Further downsampling & feature expansion
Encoder 4	320	3x3	2	GELU	104,960	Final downsampling and feature expansion
Bridge	320	-	-	-	1,959,040	Global context modeling with Transformer
Decoder 3	160	3x3	2	GELU	947,712	Upsampling and feature reconstruction
Decoder 2	80	3x3	2	GELU	237,120	Further upsampling & feature reconstruction
Decoder 1	40	3x3	1	GELU	59,200	Final upsampling and feature reconstruction
Output Layer	3	1x1	1	Sigmoid	123	Projection to RGB output



**Fig. 3.2.** SETAU-Net Architecture Diagram



## CHAPTER 4

### EXPERIMENTAL RESULTS AND ANALYSIS

This section presents the experimental setup and evaluation protocols used to assess the effectiveness of the proposed deep learning frameworks for image enhancement and restoration under adverse visual conditions. The experiments were conducted using publicly available benchmark datasets for both low-light and underwater scenarios, ensuring the models were rigorously tested across diverse and challenging conditions. Details of the datasets, data preprocessing steps, and the architectures of LEARN and SETAU-Net are provided to establish the experimental context. The performance of the proposed models was measured using widely adopted quantitative metrics such as PSNR, SSIM, and LPIPS, as well as qualitative visual comparisons. Furthermore, a comprehensive comparative analysis is conducted against existing state-of-the-art methods to highlight the strengths and improvements achieved by our approach. The results are discussed in detail, providing insights into the models' enhancement quality, computational efficiency, and practical applicability in real-world settings.

#### 4.1 Experimental Setup

This section outlines our experimental configuration. We employ a rigorous assessment process, enabling comprehensive quantitative and qualitative comparisons with leading methodologies.

LEARN and SETAU-Net are evaluated using PSNR (Peak Signal to Noise Ratio) to measure pixel-level accuracy between enhanced and ground truth images, SSIM (Structural Similarity Index) is used to assess structural similarity of enhanced and ground truth images. From -1 to 1, bigger SSIM values indicate more similarity. A window size of 7 is employed, with default settings applied to the other parameters for SETAU-Net while for LEARN the configuration are window size=11,  $C_1=0.01^2$  and  $C_2=0.03^2$ . The Structural Similarity Index (SSIM) with a range of [-1, 1] compares images using structural characteristics, contrast, and brightness. It is important because, especially when measuring structural distortions, it better matches human sense of image quality than MSE or PSNR. A greater SSIM score, approaching 1, indicates structural similarity between the enhanced image and ground truth. Superior because it means the improved image retains or accurately rebuilds the reference image's structural features and perceived quality qualities. Both metrics are calculated using 'skimage.metrics' module from scikit-image library. All test images are processed at 256×256 resolution using LANCZOS resampling from the PIL (Python Imaging Library) for consistent image loading and preprocessing across datasets.

$$\text{PSNR} = 10 \cdot \log_{10} \left( \frac{\text{MAX}_I^2}{\text{MSE}} \right) = 20 \cdot \log_{10} \left( \frac{\text{MAX}_I}{\sqrt{\text{MSE}}} \right) \quad (23)$$

Where;

$$\text{MSE} = \frac{1}{m \cdot n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2 \quad (24)$$

and,  $\text{MAX}_I$  = Maximum possible pixel value of the image.

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (25)$$

Where;

$\mu_x$  is the pixel sample mean of image  $x$ ,

$\mu_y$  is the pixel sample mean of image  $y$ ,

$\sigma_x^2$  is the sample variance of image  $x$ ,

$\sigma_y^2$  is the sample variance of image  $y$ ,

$\sigma_{xy}$  is the sample covariance of images  $x$  and  $y$ ,

$C_1 = (k_1 L)^2$ ,  $C_2 = (k_2 L)^2$  are constants to stabilize the division,

$L$  is the dynamic range of pixel values ( $2^{\text{bits per pixel}} - 1$ ),

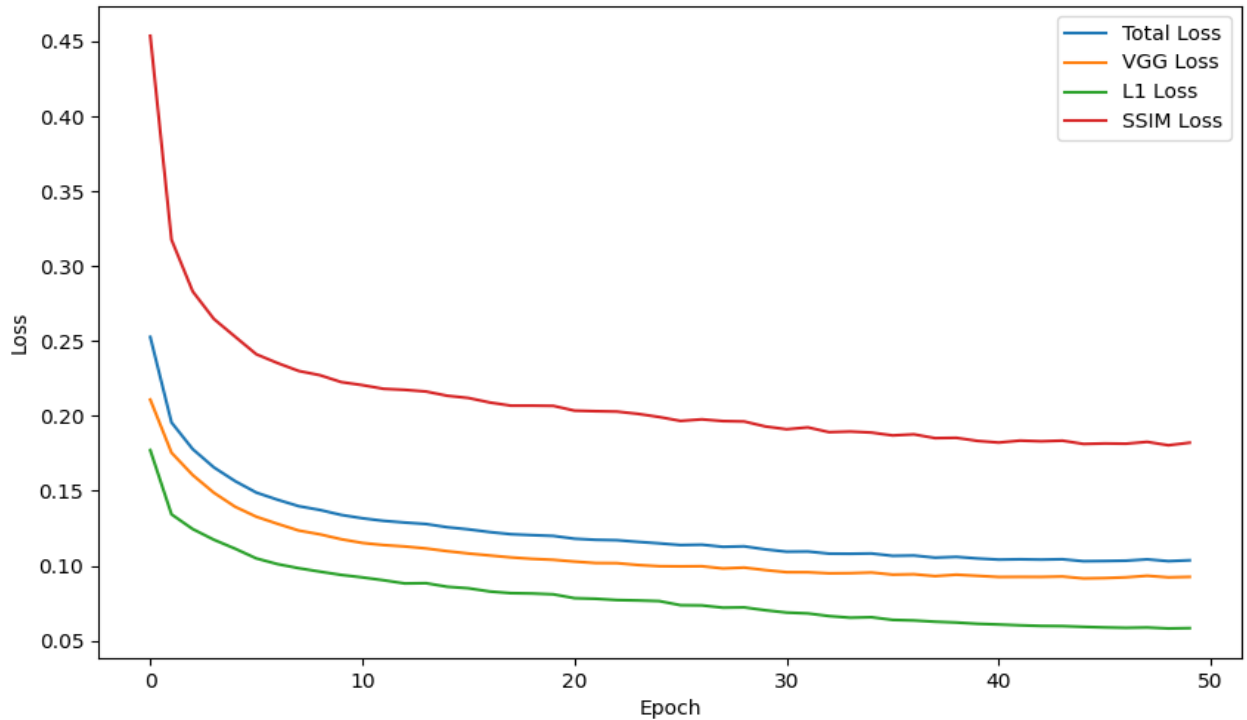
$k_1 = 0.01$  and  $k_2 = 0.03$  (default values).

Additionally, SETAU-Net is also evaluated using LPIPS (Learning Perceptual Image Patch Similarity). LPIPS uses features from deep convolutional neural networks (CNNs) pre-trained on large image datasets (e.g., AlexNet or VGG) to quantify the perceptual difference between two image patches,  $I$  and  $K$ . Instead of comparing pixel values, LPIPS determines the network's activation distance between two patches. Let  $\phi$  be the pre-trained deep network, and let  $\phi_l(I) \in \mathbb{R}^{H_l \times W_l \times C_l}$  represent the feature map (activations) obtained from layer  $l$  for the input image  $I$ . The activations are normalized on a channel-wise basis (represented as  $\widehat{\phi}_l$ ). The LPIPS distance  $d(I, K)$  is calculated as a summation over many layers.  $L$ :

$$d(I, K) = \sum_{l \in L} \frac{1}{H_l W_l} \sum_{h=1}^{H_l} \sum_{w=1}^{W_l} \left\| w_l \odot (\widehat{\phi}_l(I)_{h,w} - \widehat{\phi}_l(K)_{h,w}) \right\|_2^2 \quad (26)$$

In this formulation, let  $L$  be the collection of layers utilized from the network  $\phi$ . The variables  $H_l$  and  $W_l$  represent the spatial dimensions of the feature map at layer  $l$ .  $\widehat{\phi}_l(I)_{h,w}$  denotes the normalized activation vector (across channels  $C_l$ ) at spatial coordinates  $(h, w)$  in layer  $l$  for image  $I$ .  $w_l$  represents the channel-wise scaling factors (weights) for layer  $l$ . The weights are optimized using an independent dataset to enhance the metric's correlation with human perceptual evaluations. Let  $\odot$  represent element-wise multiplication. The notation  $\| \cdot \|_2^2$  represents the squared L2 distance, which is the summation of squared differences across channels. The spatial dimensions  $H_l$ ,  $W_l$  standardize the total, resulting in an average distance per spatial location inside the feature map. A low LPIPS score  $d(I, K)$  signifies enhanced perceptual similarity between images  $I$  and  $K$ . We employ the conventional AlexNet backbone for our assessment, with input images scaled to the range  $[-1, 1]$ .

LEARN and SETAU-Net undergo execution in PyTorch [18] and are tested on an NVIDIA GeForce RTX 4060 GPU (8GB GDDR6) alongside a 13th Gen Intel Core i7-13700HX processor running at 2.10 GHz with 16GB of RAM. The implementation leverages PyTorch's [18] mixed precision capabilities to accelerate computation while maintaining numerical stability.



**Fig. 4.1.** Training Loss Convergence of LEARN

## 4.2 Datasets

For LEARN, testing is performed on LOLv1 [6] (15 image pairs), LOLv2-Real [9] (100 pairs) and LOLv2-Synthetic [9] (100 pairs). We assess SETAU-Net using three publicly accessible underwater image enhancement datasets, each presenting distinct traits and problems. EUVP [28] (Enhancing Underwater Visual Perception) consists of paired underwater images exhibiting different levels of deterioration with their matching high-quality reference photographs. The "Underwater Scenes" subset 515 image pairs, obtained under various underwater circumstances and with different camera systems. LSUI [27] is a comprehensive dataset particularly created for the enhancement of underwater images. In accordance with [27, 28, 29], we present findings on a subset of 400 test pairings (LSUI-L400 [27]). UFO-120 [29] comprises 240 paired underwater photos exhibiting complex degradations, produced by domain transfer methodologies. This dataset functions as a significant benchmark for evaluating the generalization capabilities of enhancement techniques in unobserved underwater situations.

**Table 4.1.** Dataset image split pairs used for testing

Task	Dataset	Testing pairs
Underwater Image Enhancement	EUVP (Underwater scenes) [28]	515
	LSUI [27]	400
	UFO-120 [29]	240
Low Light Image Enhancement	LOL-v1 [6]	15
	LOL-v2 Real [9]	100
	LOL-v2 Synthetic [9]	100

### 4.3 Baselines

We evaluate SETAU-Net against traditional and state-of-the-art underwater image enhancement techniques, namely, RGHS [56], UDCP [57], UIBLA [41], UGAN [35], CLUIE-Net [59], TWIN [61], U-shaped Transformer [27], Spectroformer [58] and CE-VAE [60]. The datasets used for evaluation and comparison include LSUI-L400 [27] (400 paired test images), EUVP [28] (515 paired test images) and UFO-120 [29] (240 paired test images). The evaluation metrics used for comparison are PSNR, SSIM and LPIPS.

LEARN is evaluated against both traditional and state-of-the-art low-light image enhancement techniques, namely, RAUS [19], RetinexNet [6], KinD [11], EnlightenGAN [7], SID [10], MIRNet [12], and 3DLUT [20]. The datasets used for evaluation and comparison include LOLv1 [6] (15 paired test images), LOLv2-Real [9] (100 paired test images), and LOLv2-Synthetic [9] (100 paired test images). The evaluation metrics used for comparison are PSNR and SSIM.

### 4.4 Results and Analysis

This section presents the experimental results and analysis of the proposed deep learning frameworks for image enhancement under adverse visual conditions. The performance of LEARN and SETAU-Net is evaluated on multiple benchmark datasets using both quantitative metrics-such as PSNR, SSIM, and LPIPS-and qualitative visual comparisons. The results are systematically compared with traditional and state-of-the-art baseline methods to demonstrate the effectiveness, robustness, and efficiency of the proposed approaches. Detailed discussions highlight the strengths and limitations observed across various scenarios, offering insights into the practical applicability and generalization capabilities of the models in real-world low-light and underwater environments.

SETAU-Net exhibits robust performance on many underwater enhancement benchmarks while ensuring considerable computational economy, necessitating around 31.1 GFLOPs and 3.36 million parameters. The quantitative data are encapsulated in Table 4.2.

**Table 4.2.** Real-time Performance Evaluation of SETAU-Net

<b>Dataset</b>	<b>FPS</b>	<b>Processing time</b>
LSUI [27]	$118.93 \pm 10$	~8.41 ms per image
EUVP [28]	$115.21 \pm 10$	~8.54 ms per image
UFO-120 [29]	$118.03 \pm 10$	~8.47 ms per image
<i>Average</i>	$117.39 \pm 10$	~8.47 ms per image

In the LSUI-L400 [27] dataset, SETAU-Net exhibits superior performance, attaining a PSNR of 28.96 dB, an SSIM of 0.92, and an LPIPS of 0.07, with an image processing rate of  $118.93 \pm 10$  FPS (approximately 8.41 ms per image). This surpasses CE-VAE [60] (24.49 dB PSNR, 0.84 SSIM, 0.26 LPIPS) by +4.47 dB PSNR, +0.08 SSIM, and -0.19 LPIPS, and also exceeds U-Shaped Transformer [27] (23.02 dB PSNR, 0.82 SSIM, 0.29 LPIPS) across all metrics. On the EUVP [28] dataset, SETAU-Net achieves a PSNR of 25.90 dB, an SSIM of 0.86, and an LPIPS of 0.14 at  $115.21 \pm 10$  FPS (approximately 8.54 ms per image). While CE-VAE [60] (27.75 dB PSNR, 0.89 SSIM, 0.20 LPIPS) and U-Shaped Transformer [27] (27.59 dB PSNR, 0.88 SSIM, 0.23 LPIPS) report higher PSNR and SSIM, SETAU-Net provides improved perceptual quality, achieving a lower LPIPS by -0.06 compared to CE-VAE [60] and -0.09

compared to U-Shaped Transformer [27]. In the UFO-120 [29] dataset, SETAU-Net attains a PSNR of 27.28 dB, SSIM of 0.87, and LPIPS of 0.12 at a rate of  $118.03 \pm 10$  FPS (approximately 8.47 ms per image), surpassing CE-VAE [60] (24.38 dB PSNR, 0.79 SSIM, 0.28 LPIPS) by +2.90 dB PSNR, +0.08 SSIM, and -0.16 LPIPS. SETAU-Net consistently demonstrates superior perceptual quality (LPIPS) and maintains an average processing speed of  $117.39 \pm 10$  FPS (8.47 ms per image) across all datasets. With only 3.36 million parameters and 31.1 GFLOPs per  $256 \times 256$  image, SETAU-Net balances enhancement quality and computational efficiency, making it an effective solution for real-world underwater imaging applications that require both quality and speed, and outperforming more resource-intensive transformer-based and multi-branch models such as UGAN [35].

**Table 4.3.** Comparative Analysis of SETAU-Net and Existing Underwater Image Enhancement (For each metric/dataset, the best result is highlighted in red, second best is highlighted in blue and third best is highlighted in purple).  $\uparrow$  Denotes that a higher value for a particular metric is better while  $\downarrow$  denotes that a lower value for a particular metric is better

Methods	LSUI-L400 [27]			EUVP [28]			UFO-120 [29]		
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
RGHS [56]	18.44	0.80	0.31	18.05	0.78	0.31	17.48	0.71	0.37
UDCP [31]	13.24	0.56	0.39	14.52	0.59	0.35	14.50	0.55	0.42
UIBLA [57]	17.75	0.72	0.36	18.95	0.74	0.33	17.04	0.64	0.40
UGAN [35]	19.40	0.77	0.37	20.98	0.83	0.31	19.92	0.73	0.38
CLUIE-Net [59]	18.71	0.78	0.33	18.90	0.78	0.30	18.43	0.72	0.36
TWIN [61]	19.84	0.79	0.33	18.91	0.79	0.32	18.21	0.72	0.37
UST [27]	23.02	0.82	0.29	27.59	0.88	0.23	22.82	0.77	0.33
Spectroformer [58]	20.09	0.79	0.32	18.70	0.79	0.32	18.03	0.71	0.37
CE-VAE [60]	24.49	0.84	0.26	27.75	0.89	0.20	24.38	0.79	0.28
<b>SETAU-Net</b>	<b>28.96</b>	<b>0.92</b>	<b>0.07</b>	<b>25.90</b>	<b>0.86</b>	<b>0.14</b>	<b>27.28</b>	<b>0.87</b>	<b>0.12</b>

LEARN demonstrates strong performance across multiple low-light enhancement benchmarks with computational requirements of 46.67 GFLOPS and 7.62M parameters. On the LOLv1 [6] dataset, it achieves a PSNR of 22.86 dB and SSIM of 0.8374 with a processing speed of  $104.56 \pm 10$  FPS ( $0.0096 \pm 0.002$  seconds per image), outperforming methods like KinD [12] (20.86 dB PSNR, 0.79 SSIM) and RetinexNet [6] (16.77 dB PSNR, 0.56 SSIM) in terms of enhancement quality while maintaining comparable computational efficiency to KinD [12] (34.99 GFLOPS, 8.02M parameters). For the LOLv2-Real [9] dataset, LEARN reaches 24.62 dB PSNR and 0.8746 SSIM with  $97.25 \pm 10$  FPS ( $0.0131 \pm 0.0032$  seconds per image), significantly improving upon RAUS [19] (18.37 dB PSNR, 0.723 SSIM) and EnlightenGAN [7] (18.23 dB PSNR, 0.617 SSIM). On LOLv2-Synthetic [9], LEARN attains 22.54 dB PSNR and 0.8720 SSIM with a processing speed of  $82.33 \pm 10$  FPS ( $0.0110 \pm 0.0020$  seconds per image), showcasing better enhancement quality than methods like RetinexNet [6] (17.13 dB PSNR, 0.798 SSIM) and EnlightenGAN [7] (16.57 dB PSNR, 0.734 SSIM). Overall, LEARN achieves an average PSNR of 23.34 dB and SSIM of 0.8613 across all datasets, with an average processing time of  $0.0111 \pm 0.0005$  seconds per image ( $90.82 \pm 5$  FPS). These metrics demonstrate LEARN's ability to balance enhancement quality with real-time processing

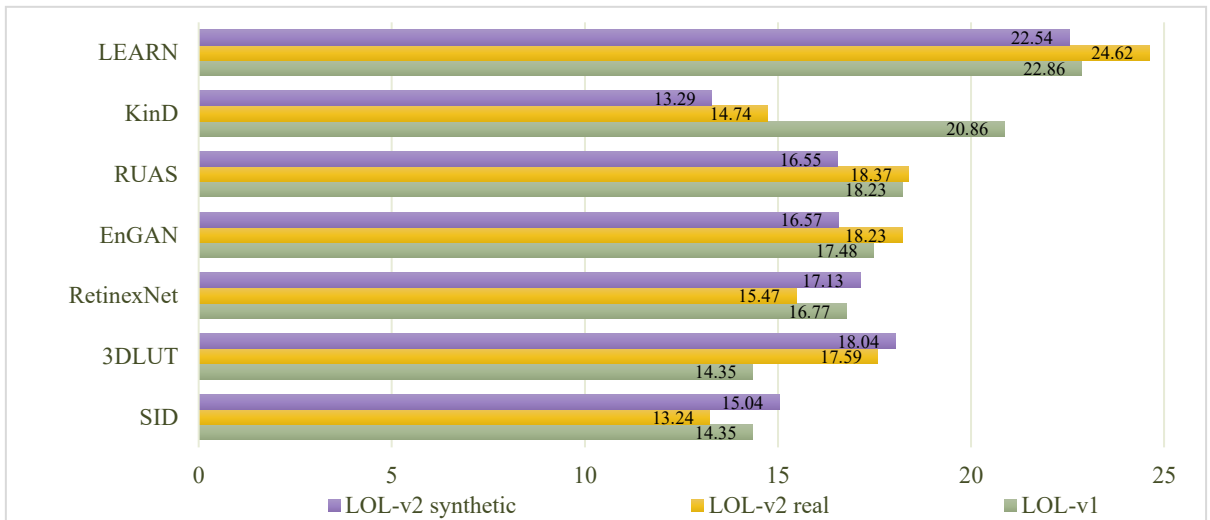
capabilities while maintaining a reasonable computational footprint, making it suitable for practical applications requiring both efficiency and high-quality results.

**Table 4.4.** Real-time Performance Evaluation of LEARN Model (in seconds)

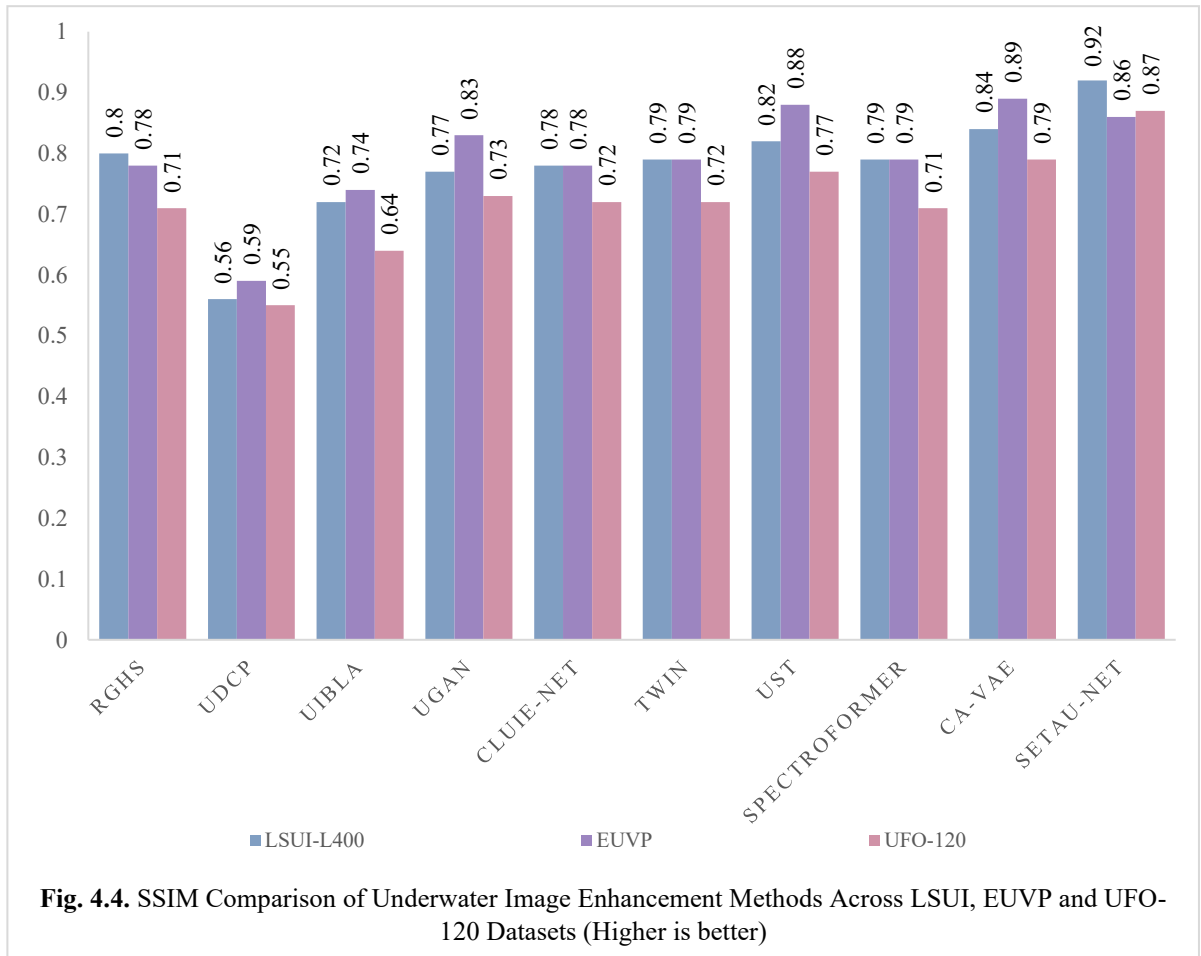
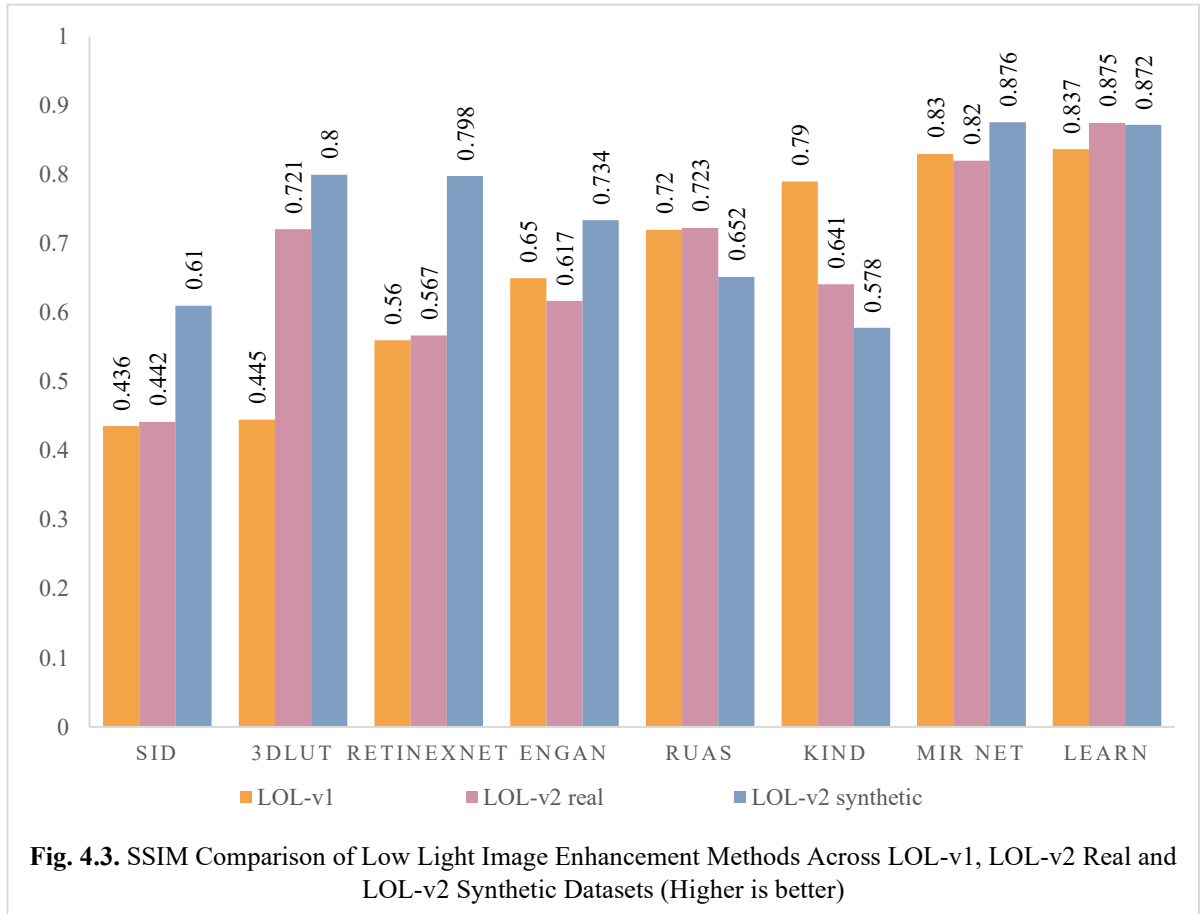
Dataset	FPS	Processing time
LOL-v1 [6]	$104.56 \pm 10$	$0.0096 \pm 0.002$
LOL-v2 Real [9]	$97.25 \pm 10$	$0.0131 \pm 0.0032$
LOL-v2 Synth [9]	$82.33 \pm 10$	$0.0110 \pm 0.0020$
<i>Average</i>	$90.82 \pm 5$	$0.0111 \pm 0.0005$

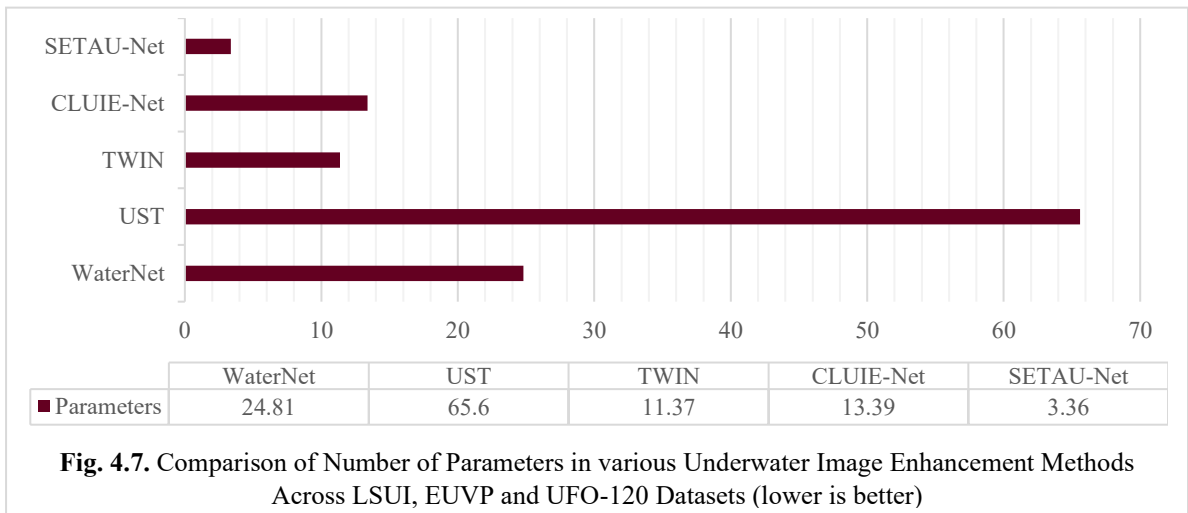
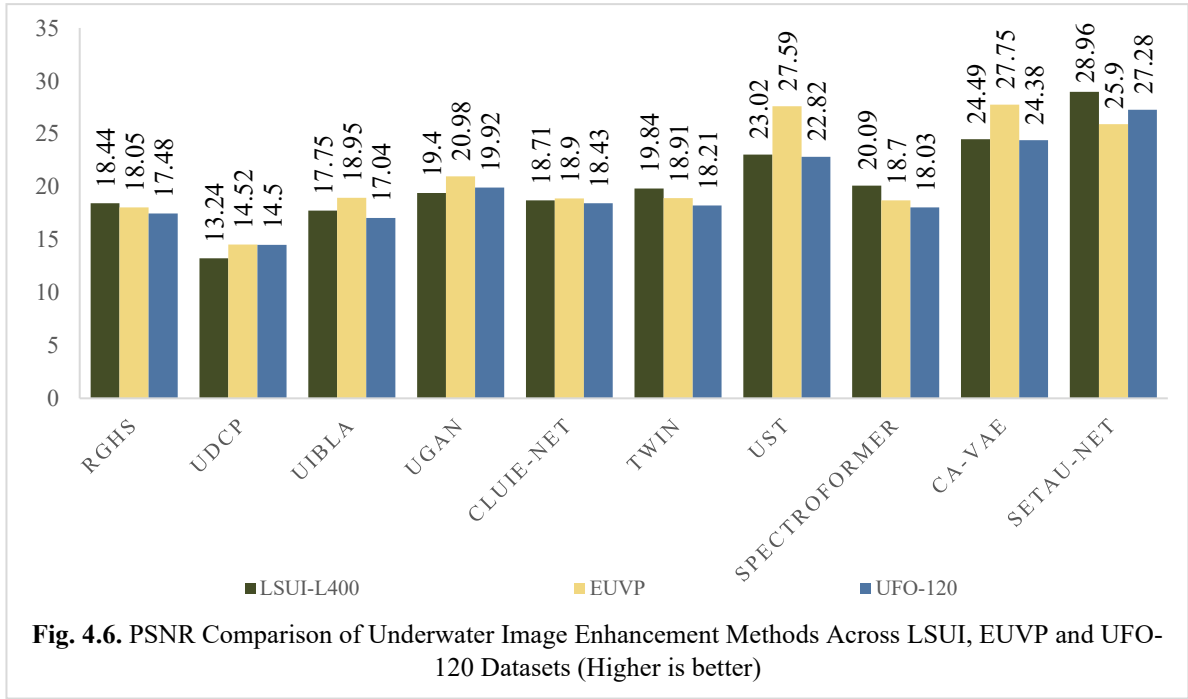
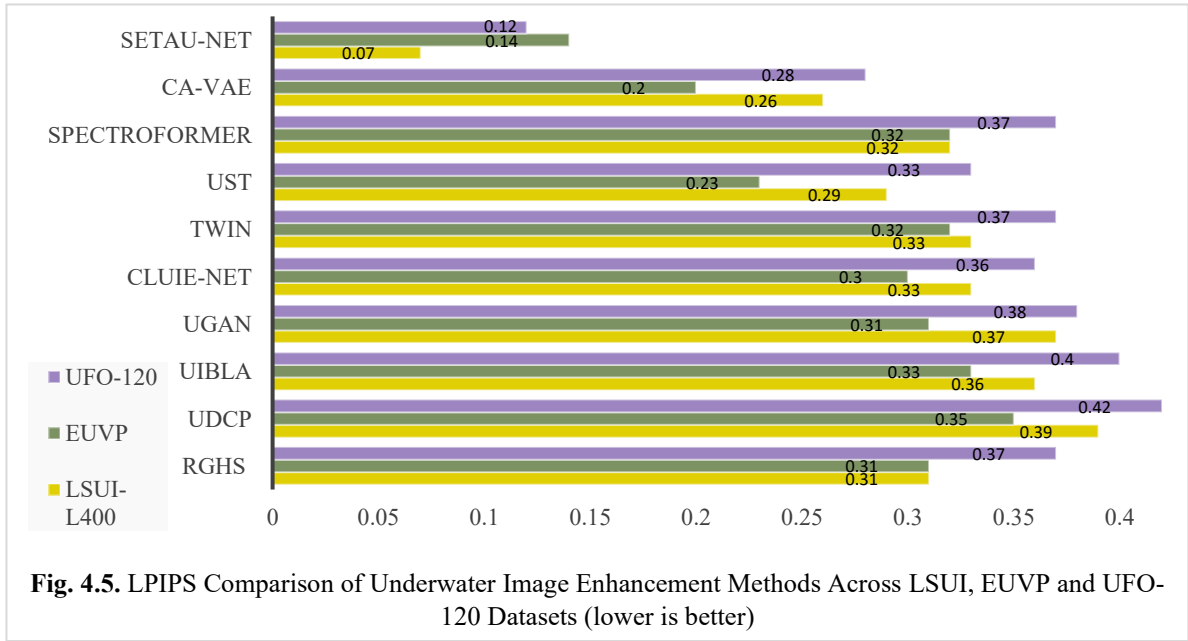
**Table 4.5.** Comparative Analysis of SETAU-Net and Existing Underwater Image Enhancement (For each metric/dataset, the best result is highlighted in **red** and second best is highlighted in **blue**).  $\uparrow$  Denotes that a higher value for a particular metric is better while  $\downarrow$  denotes that a lower value for a particular metric is better

Methods	Model Complexity		LOL-v1 [6]		LOL-v2 Real [9]		LOL-v2 Synth.[9]	
	GFLOPS $\downarrow$	Parameters $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	PSNR $\uparrow$	SSIM $\uparrow$	PSNR $\uparrow$	SSIM $\uparrow$
SID [10]	13.73	7.76	14.35	0.436	13.24	0.442	15.04	0.61
3DLUT [20]	0.075	0.59	14.35	0.445	17.59	0.721	18.04	0.8
RetinexNet [6]	587.47	0.84	16.77	0.56	15.47	0.567	17.13	0.798
EGAN [7]	61.01	114.35	17.48	0.65	18.23	0.617	16.57	0.734
RAUS [19]	0.83	0.003	18.23	0.72	18.37	0.723	16.55	0.652
KinD [12]	34.99	8.02	20.86	0.79	14.74	0.641	13.29	0.578
MIRNet [11]	785	31.76	24.14	0.830	20.02	0.820	21.94	0.876
LEARN	46.67	7.62	22.86	0.837	24.62	0.875	22.54	0.872












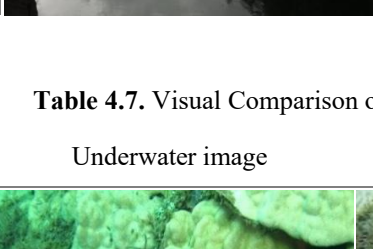
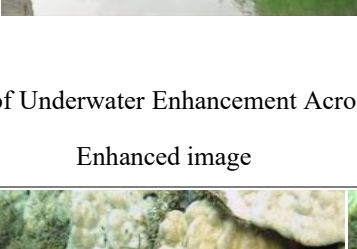
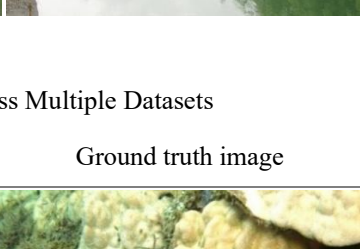
**Fig. 4.2.** PSNR Comparison of Low Light Image Enhancement Methods Across LOL-v1, LOL-v2 Real and LOL-v2 Synthetic Datasets (Higher is better)



















**Table 4.6.** Visual Comparison of Low Light Image Enhancement Across Multiple Datasets

Dataset	Low light image	Enhanced image	Ground truth image
LOL-v1 [6]			
			
LOL-v2 Real [9]			
LOL-v2 Synthetic [9]			

**Table 4.7.** Visual Comparison of Underwater Enhancement Across Multiple Datasets

Dataset	Underwater image	Enhanced image	Ground truth image
LSUI- L400 [27]			
			
EUVF [28]			
UFO- 120 [29]			

# CHAPTER 5

## CONCLUSION, FUTURE WORK AND SOCIAL IMPACT

### 5.1 Conclusion

In order to tackle the distinct difficulties presented by unfavourable visual situations, we have presented and thoroughly assessed two sophisticated deep learning frameworks in this research study: SETAU-Net for underwater image restoration and LEARN for low-light image enhancement. Both models achieve state-of-the-art performance on several public benchmarks while preserving lightweight architectures appropriate for real-time deployment, exhibiting a strong balance between computational efficiency and enhanced quality. LEARN is extremely relevant for a variety of applications, from autonomous driving and medical imaging to smartphone photography and surveillance, due to its capacity to provide high-fidelity, instantaneous picture improvement on devices with limited resources. Similarly, underwater photography, marine research, environmental monitoring, and autonomous robotics can all benefit from SETAU-Net's exceptional restoration accuracy and efficiency. The comprehensive results underscore the versatility, scalability, and societal relevance of these frameworks, highlighting their potential to advance technological capabilities and address real-world needs in a variety of critical applications.

LEARN offers a practical, effective way to maximize computing resources and improve low-light photos. The encoder-decoder architecture incorporates skip connections, residual blocks with larger kernels to make the receptive field bigger, and Convolutional Block Attention Modules (CBAM) to highlight significant channel- and space-specific details. A Laplacian enhancement module enhances detail sharpness and edge definition. The model has a consistent computational need of 46.67 GFLOPS with 7.62 million parameters across many low-light enhancement datasets, attaining a mean PSNR of 23.34 dB and an SSIM of 0.8613. LEARN demonstrates significant potential for providing high-fidelity, real-time image enhancements, achieving an average processing rate of  $90.82 \pm 5$  FPS ( $0.0111 \pm 0.0005$  seconds per image), so confirming its capability to give high-quality solutions instantaneously. The findings suggest that LEARN is applicable in resource-constrained real-world situations, as it sustains a constant balance between expedited processing and enhanced quality. LEARN could be helpful in various ways under suboptimal illumination. Without more hardware, it can improve low-light photography in smartphone camera systems. Improved view of low-light collected film could improve the security monitoring features of surveillance systems. This technique increases vision systems' ability to distinguish objects and hazards in low light, potentially improving autonomous driving safety. LEARN's clarity improvement characteristics in medical imaging will benefit low-light endoscopic and microscopic images. Design adjustment can make the model domain-specific, maximize hardware acceleration-based mobile device integration, or address motion blur or noise with alternative methods. These developments would make it more adaptable to many kinds of specialized work, therefore transforming it into a versatile solution for numerous technological and medical applications.

SETAU-Net presents a robust and efficient approach to underwater image enhancement, successfully achieving a balance between computational efficiency and restoration quality. Constructed with an encoder-decoder framework featuring skip connections, SimAM attention, a CNN Sobel edge enhancement module for clear edge guidance, and a transformer bridge for comprehensive global context, SETAU-Net reaches leading performance on multiple public underwater benchmarks. The quantitative analysis unequivocally demonstrates that SETAU-Net surpasses both traditional and deep learning methodologies. In the LSUI-L400 [27] dataset,

SETAU-Net exhibits a PSNR of 28.96 dB, an SSIM of 0.92, and a notably low LPIPS of 0.07. In the context of EUVP [28], SETAU-Net achieves a PSNR of 25.90 dB, an SSIM of 0.86, and an LPIPS of 0.14, demonstrating superior perceptual quality compared to all other methods evaluated. In the context of the demanding UFO-120 [29] dataset, SETAU-Net demonstrates a performance of 27.28 dB PSNR, 0.87 SSIM, and 0.12 LPIPS, consistently exceeding all baseline models. The model demonstrates high efficiency, utilizing only 31.1 GFLOPs and comprising 3.36 million parameters. The findings indicate an average processing speed of  $117.39 \pm 10$  FPS ( $\sim 8.47$  ms per image), confirming the model's appropriateness for real-time applications, practical deployment, and resource-limited environments. Practical use in a variety of underwater imaging applications is made possible by the lightweight architecture, which enables smooth implementation on embedded and mobile systems. The features include improved clarity for consumer underwater photography, assistance for marine research, ease of environmental monitoring, and improvements in vision systems for autonomous cars and underwater robotics. Applications of SETAU-Net's real-time, high-fidelity underwater photography advancements are anticipated to have an influence on a number of domains, including autonomous navigation, marine exploration, environmental preservation, and underwater surveillance. Enhancing visibility and detail in difficult-to-reach places increases user experience, encourages scientific research, and improves safety. By combining speed, quality, and efficiency, SETAU-Net presents itself as a viable substitute for the development of underwater photography technology in both commercial and recreational settings.

Overall, the results affirm that both LEARN and SETAU-Net offer practical, high-performance solutions for challenging image enhancement and restoration tasks. Their efficiency, adaptability, and strong restoration quality position them as valuable tools for a wide range of real-world applications, paving the way for further advancements in intelligent imaging technologies.

## 5.2 Future work

There are still a number of intriguing avenues for further study, despite the suggested frameworks' excellent performance and usefulness. Important next developments include significantly lowering processing needs, integrating multi-scale or domain-specific modules, and improving adaptation to increasingly varied and harsh visual situations. The impact and versatility of these approaches will be further expanded by looking into unsupervised or self-supervised learning to lessen reliance on paired datasets, expanding the models to handle additional degradations like motion blur or severe noise, and optimizing deployment for edge and mobile devices. Further developments along these lines will contribute to the continued stability, effectiveness, and accessibility of deep learning-based picture augmentation in a greater variety of real-world applications.

**LEARN-** Leveraging LEARN's established ability to integrate computational efficiency with superior quality presents multiple opportunities for application expansion. The  $256 \times 256$  input resolution hinders the retention of intricate features in high-resolution images, and the static Laplacian kernel configuration restricts adaptability to diverse edge profiles, notwithstanding LEARN's robust performance in low-light enhancement and real-time processing ( $\sim 90.82$  FPS).

To tackle these challenges, a resolution-agnostic design utilizing hybrid CNN-Transformer blocks or progressive resizing during training could effectively support elevated resolutions. Furthermore, the dependence on paired training data (LOLv1/v2 [6, 9], SID [10]) limits its applicability in unpaired real-world contexts. Future research may investigate CycleGAN-style training to facilitate enhancement without the necessity of paired low- and normal-light data. Unsupervised and semi-supervised techniques, such as self-supervised denoising and domain adaptation, can mitigate the issue of static noise distributions in training data, which constrains

noise adaptability. Integrating neural architecture search (NAS) with pruning or 8-bit quantization can diminish the computational load (46.67 GFLOPS) by 40-60%, thereby enhancing the feasibility of mobile implementation and increasing efficiency in edge deployment. To enhance efficiency in edge deployment, neural architecture search (NAS) integrated with pruning or 8-bit quantization can diminish the computational burden (46.67 GFLOPS) by 40-60%, thereby enabling mobile implementation. The static Learnable edge detection modules or adaptive frequency decomposition can improve multi-scale edge handling over Laplacian kernel design, which struggles with various edge shapes. Perceptual color consistency losses or advanced color correction modules can reduce excessive low-light color distortions in enhancement models. Addressing these limitations could improve LEARN's adaptability, computing demands, and performance in real-world applications like nighttime surveillance and mobile photography.

**SETAU-Net-** This framework demonstrates computational efficiency alongside high enhancement quality, while improving performance on various underwater datasets, coupled with notable processing speed. However, particular limitations must be addressed to improve its effectiveness in practical underwater applications. SETAU-Net's 256 x 256 input resolution hinders its ability to preserve fine-scale details in high-resolution underwater photography, crucial for scientific documentation and marine biology applications. The fixed-scale Sobel module detects edges well, but it may struggle to preserve multi-scale features under varying underwater circumstances and object sizes. By training at numerous resolutions or extracting features independent of scale, a solution can adapt to different situations. Coastal, oceanic, and turbid waters have variable light absorption and scattering qualities, but SETAU-Net does not adjust to them. Some enhanced images have color casts, according to comparisons. Even if these casts are relatively low, they affect generalizability among aquatic conditions. A water-type classification branch or color correction module could improve underwater image enhancement techniques. 3.36 million parameters and 31.1 GFLOPs make up SETAU-Net's lightweight design. On resource-limited platforms like AUVs and portable underwater photography systems, optimization may boost its utility. The knowledge distillation procedure from SETAU-Net to smaller student networks improves quality while reducing computing loads. With low performance deterioration, network pruning, filter factorization, and 8-bit quantization can potentially Reduce computational load by 30-50%. The SimAM attention parameters, transformer bridge design, and composite loss weights of SETAU-Net could be optimized through systematic parameter tuning to improve its effectiveness using techniques like Bayesian optimisation and population-based training to automatically find optimal hyperparameters. Adaptive loss weighting can balance the composite loss function during training. Designing parameter search spaces for architectural needs helps solve underwater augmentation problems. By solving these restrictions, future SETAU-Net versions could improve underwater photography flexibility, computational efficiency, and quality for marine biology, underwater archaeology, and ocean engineering where effective color correction and image clarity is essential.

### 5.3 Social Impact

Advancements in real-time image enhancement technologies have far-reaching implications for both everyday life and specialized fields. By addressing critical challenges in low-light and underwater environments, modern deep learning models like LEARN and SETAU-Net offer tangible benefits across diverse societal and industrial domains.

LEARN combines practicality with significant society benefit in a variety of scenarios by enabling real-time improvement of low-light images. The architecture is intended to be lightweight and efficient in computing to handle real-time applications like security monitoring, autonomous driving, and live video processing. Because of this, it is ideal for low-power devices

such as smartphones, security cameras, and embedded platforms. Fields that potentially benefit from LEARN's enhanced visibility and edge detail preservation under challenging lighting conditions include forensics, medical imaging, and traffic safety, among others. LEARN's ability to generalize across makes it suitable for both well-lit indoor environments and utterly dark outdoor settings. LEARN's equilibrium of real-time functionalities and social significance demonstrates its accessibility for practical implementation by improving safety, diagnostics, and user experiences in low-light conditions.

On the other hand, in today's world, which is becoming more digital and interconnected by the day, high-quality underwater imaging is crucial for a variety of technological and societal applications. This is why SETAU-Net was developed. The usefulness of conventional imaging and analysis techniques can be significantly reduced in underwater environments because of the special visual challenges posed by light absorption, scattering, and color distortion. Because of its resource-efficient, real-time design, SETAU-Net is highly feasible for use in fields where speedy, dependable image enhancement is essential for mission success, such as autonomous underwater vehicles, marine robotics, environmental monitoring, and scientific exploration. In the context of maritime research, SETAU-Net allows for clearer observation and documenting of underwater habitats, which aids biodiversity evaluation, habitat conservation, and early detection of environmental changes or risks, amongst many other usages. Its applications include industrial inspection, underwater archaeology, and recreational diving, where improved optical clarity can boost safety, operational efficiency, and user experience. By utilizing advanced deep learning techniques to produce robust, high-fidelity image restoration, SETAU-Net advances digital imaging technology while also supporting the greater societal objective of sustainable ocean development. In an era where the health of marine habitats is strongly related to global well-being, SETAU-Net provides researchers, policymakers, and industry experts with superior visual data, ultimately enabling more informed decision-making and beneficial social effect.

In summary, the practical deployment of LEARN and SETAU-Net stands to deliver meaningful benefits across multiple domains, improving safety, efficiency, and decision-making. These innovations exemplify how cutting-edge deep learning can drive positive social impact through accessible, high-quality imaging solutions.

## References

- [1] Abdullah-Al-Wadud, M., Kabir, M. H., Dewan, M. A. A., & Chae, O. (2007). A dynamic histogram equalization for image contrast enhancement. *IEEE transactions on consumer electronics*, 53(2), 593-600.
- [2] Celik, T., & Tjahjadi, T. (2011). Contextual and variational contrast enhancement. *IEEE Transactions on Image Processing*, 20(12), 3431-3441.
- [3] Cheng, H. D., & Shi, X. J. (2004). A simple and effective histogram equalization approach to image enhancement. *Digital signal processing*, 14(2), 158-170.
- [4] Pisano, E. D., Zong, S., Hemminger, B. M., DeLuca, M., Johnston, R. E., Muller, K., ... & Pizer, S. M. (1998). Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms. *Journal of Digital imaging*, 11, 193-200.
- [5] Land, E. H. (1977). The retinex theory of color vision. *Scientific american*, 237(6), 108-129.
- [6] Wei, C., Wang, W., Yang, W., & Liu, J. (2018). Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*.
- [7] Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., ... & Wang, Z. (2021). Enlighten: Deep light enhancement without paired supervision. *IEEE transactions on image processing*, 30, 2340-2349.
- [8] Guo, C., Li, C., Guo, J., Loy, C. C., Hou, J., Kwong, S., & Cong, R. (2020). Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 1780-1789).
- [9] Yang, W., Wang, W., Huang, H., Wang, S., & Liu, J. (2021). Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE Transactions on Image Processing*, 30, 2072-2086.
- [10] Chen, C., Chen, Q., Do, M. N., & Koltun, V. (2019). Seeing motion in the dark. In *Proceedings of the IEEE/CVF International conference on computer vision* (pp. 3185-3194).
- [11] Zamir, S. W., Arora, A., Khan, S., Hayat, M., Khan, F. S., Yang, M. H., & Shao, L. (2020). Learning enriched features for real image restoration and enhancement. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16* (pp. 492-511). Springer International Publishing.
- [12] Zhang, Y., Zhang, J., & Guo, X. (2019, October). Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM international conference on multimedia* (pp. 1632-1640).
- [13] Guo, X., Li, Y., & Ling, H. (2016). LIME: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing*, 26(2), 982-993.
- [14] Li, M., Liu, J., Yang, W., Sun, X., & Guo, Z. (2018). Structure-revealing low-light image enhancement via robust retinex model. *IEEE transactions on image processing*, 27(6), 2828-2841.



- [15] Cai, Y., Bian, H., Lin, J., Wang, H., Timofte, R., & Zhang, Y. (2023). Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 12504-12513).
- [16] Lim, S., & Kim, W. (2020). DSLR: Deep stacked Laplacian restorer for low-light image enhancement. *IEEE Transactions on Multimedia*, 23, 4272-4284.
- [17] Luo, Z., Tang, J., Zhou, K., Huang, Z., Zhang, J., & Hou, Y. (2024, October). Unsupervised Low Light Image Enhancement via SNR-Aware Swin Transformer. In *2024 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 4157-4162). IEEE.
- [18] Paszke, A. (2019). Pytorch: An imperative style, high-performance deep learning library. *arXiv preprint arXiv:1912.01703*.
- [19] Liu, R., Ma, L., Zhang, J., Fan, X., & Luo, Z. (2021). Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10561-10570).
- [20] Zeng, H., Cai, J., Li, L., Cao, Z., & Zhang, L. (2020). Learning image-adaptive 3d lookup tables for high performance photo enhancement in real-time. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(4), 2058-2073.
- [21] He, T., Song, T., Liu, Y., Yang, F., Chen, R., & Li, Z. (2025, April). Collaborative Dual-Branch Spatial-Frequency Enhancement Network for Low-Light Images. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1-5). IEEE.
- [22] Sun, M., Wang, X., & Yi, R. (2025, April). MLSwinTNet: A Multi-Level Feature Interaction Network for Low-Light Image Enhancement. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1-5). IEEE.
- [23] Masood, M. K. K., Nava, E., & Otero, P. (2025). A Novel Intensity-Corrected Blue Channel Compensation and Edge-Preserving Contrast Enhancement Using Laplace Filter and Sigmoid Function for Sand-Dust Image Enhancement. *IEEE Access*.
- [24] Chatterjee, S., Koul, S., Kaul, I., & Singh, K. (2023, December). ADC-Net: Attention-based Dense Convolutional Network for Underwater Image Enhancement. In *2023 5th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)* (pp. 790-795). IEEE.
- [25] Zhu, J., Chen, T., Xue, P., & Zhao, X. (2023, December). An Algorithm for Low-Light Image Enhancement Integrating Residual Dense Block and Attention Mechanis. In *2023 5th International Academic Exchange Conference on Science and Technology Innovation (IAECST)* (pp. 742-748). IEEE.
- [26] Yan, Q., Hu, T., Wu, P., Dai, D., Gu, S., Dong, W., & Zhang, Y. (2025). Efficient Image Enhancement with A Diffusion-Based Frequency Prior. *IEEE Transactions on Circuits and Systems for Video Technology*.
- [27] Peng, L., Zhu, C., & Bian, L. (2023). U-shape transformer for underwater image enhancement. *IEEE Transactions on Image Processing*, 32, 3066-3079.
- [28] Islam, Md Jahidul, Youya Xia, and Junaed Sattar. "Fast underwater image enhancement for improved visual perception." *IEEE Robotics and Automation Letters* 5.2 (2020): 3227-3234.

- [29] Islam, M. J., Luo, P., & Sattar, J. (2020). Simultaneous enhancement and super-resolution of underwater imagery for improved visual perception. *arXiv preprint arXiv:2002.01155*.
- [30] Ghani, A. S. A., & Isa, N. A. M. (2015). Underwater image quality enhancement through integrated color model with Rayleigh distribution. *Applied soft computing*, 27, 219-230.
- [31] Li, C. Y., Guo, J. C., Cong, R. M., Pang, Y. W., & Wang, B. (2016). Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior. *IEEE Transactions on Image Processing*, 25(12), 5664-5677.
- [32] Han, P., Liu, F., Yang, K., Ma, J., Li, J., & Shao, X. (2017). Active underwater descattering and image recovery. *Applied Optics*, 56(23), 6631-6638.
- [33] Berman, D., Levy, D., Avidan, S., & Treibitz, T. (2020). Underwater single image color restoration using haze-lines and a new quantitative dataset. *IEEE transactions on pattern analysis and machine intelligence*, 43(8), 2822-2837.
- [34] Neumann, L., Garcia, R., Jánosik, J., & Gracias, N. (2018). Fast underwater color correction using integral images. *Instrumentation Viewpoint*, (20), 53-54.
- [35] Fabbri, C., Islam, M. J., & Sattar, J. (2018, May). Enhancing underwater imagery using generative adversarial networks. In *2018 IEEE international conference on robotics and automation (ICRA)* (pp. 7159-7165). IEEE.
- [36] Guo, Y., Li, H., & Zhuang, P. (2019). Underwater image enhancement using a multiscale dense generative adversarial network. *IEEE Journal of Oceanic Engineering*, 45(3), 862-870.
- [37] Hu, K., Zhang, Y., Weng, C., Wang, P., Deng, Z., & Liu, Y. (2021). An underwater image enhancement algorithm based on generative adversarial network and natural image quality evaluation index. *Journal of Marine Science and Engineering*, 9(7), 691.
- [38] Park, J., Han, D. K., & Ko, H. (2019). Adaptive weighted multi-discriminator CycleGAN for underwater image enhancement. *Journal of Marine Science and Engineering*, 7(7), 200.
- [39] Zhang, H., Sun, L., Wu, L., & Gu, K. (2021). DuGAN: An effective framework for underwater image enhancement. *IET Image Processing*, 15(9), 2010-2019.
- [40] Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision* (pp. 2223-2232).
- [41] Ancuti, C. O., Ancuti, C., De Vleeschouwer, C., & Bekaert, P. (2017). Color balance and fusion for underwater image enhancement. *IEEE Transactions on image processing*, 27(1), 379-393.
- [42] Bazeille, S., Quidu, I., Jaulin, L., & Malkasse, J. P. (2006). Automatic underwater image pre-processing. In *CMM'06* (p. xx).
- [43] Lu, H., Li, Y., & Serikawa, S. (2013, September). Underwater image enhancement using guided trigonometric bilateral filter and fast automatic color correction. In *2013 IEEE international conference on image processing* (pp. 3412-3416). IEEE.



- [44] Li, C., Quo, J., Pang, Y., Chen, S., & Wang, J. (2016, March). Single underwater image restoration by blue-green channels dehazing and red channel correction. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1731-1735). IEEE.
- [45] Park, D., Han, D. K., & Ko, H. (2017). Enhancing underwater color images via optical imaging model and non-local means denoising. *IEICE TRANSACTIONS on Information and Systems*, 100(7), 1475-1483.
- [46] Zhang, W., Wang, Y., & Li, C. (2022). Underwater image enhancement by attenuated color channel correction and detail preserved contrast enhancement. *IEEE Journal of Oceanic Engineering*, 47(3), 718-735.
- [47] Peng, Y. T., & Cosman, P. C. (2017). Underwater image restoration based on image blurriness and light absorption. *IEEE transactions on image processing*, 26(4), 1579-1594.
- [48] Qiao, N., Dong, L., & Sun, C. (2022). Adaptive deep learning network with multi-scale and multi-dimensional features for underwater image enhancement. *IEEE Transactions on Broadcasting*, 69(2), 482-494.
- [49] Li, C., Guo, C., Ren, W., Cong, R., Hou, J., Kwong, S., & Tao, D. (2019). An underwater image enhancement benchmark dataset and beyond. *IEEE transactions on image processing*, 29, 4376-4389.
- [50] Xing, Z., Cai, M., & Li, J. (2022, October). Improved shallow-uwnet for underwater image enhancement. In *2022 IEEE International Conference on Unmanned Systems (ICUS)* (pp. 1191-1196). IEEE.
- [51] Tang, Y., Iwaguchi, T., Kawasaki, H., Sagawa, R., & Furukawa, R. (2022). AutoEnhancer: Transformer on U-Net architecture search for underwater image enhancement. In *Proceedings of the Asian conference on computer vision* (pp. 1403-1420).
- [52] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- [53] Shen, Z., Xu, H., Luo, T., Song, Y., & He, Z. (2023). UDAformer: Underwater image enhancement based on dual attention transformer. *Computers & Graphics*, 111, 77-88.
- [54] Yang, L., Zhang, R. Y., Li, L., & Xie, X. (2021, July). Simam: A simple, parameter-free attention module for convolutional neural networks. In *International conference on machine learning* (pp. 11863-11874). PMLR.
- [55] Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12), 2481-2495.
- [56] Huang, D., Wang, Y., Song, W., Sequeira, J., & Mavromatis, S. (2018). Shallow-water image enhancement using relative global histogram stretching based on adaptive parameter acquisition. In *MultiMedia Modeling: 24th International Conference, MMM 2018, Bangkok, Thailand, February 5-7, 2018, Proceedings, Part I 24* (pp. 453-465). Springer International Publishing.
- [57] Drews, P. L., Nascimento, E. R., Botelho, S. S., & Campos, M. F. M. (2016). Underwater depth estimation and image restoration based on single images. *IEEE computer graphics and applications*, 36(2), 24-35.

- [58] Khan, R., Mishra, P., Mehta, N., Phutke, S. S., Vipparthi, S. K., Nandi, S., & Murala, S. (2024). Spectroformer: Multi-domain query cascaded transformer network for underwater image enhancement. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision* (pp. 1454-1463).
- [59] Li, K., Wu, L., Qi, Q., Liu, W., Gao, X., Zhou, L., & Song, D. (2022). Beyond single reference for training: Underwater image enhancement via comparative learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(6), 2561-2576.
- [60] Pucci, R., & Martinel, N. (2024). Capsule enhanced variational autoencoder for underwater image reconstruction. *arXiv preprint arXiv:2406.01294*.
- [61] Liu, R., Jiang, Z., Yang, S., & Fan, X. (2022). Twin adversarial contrastive learning for underwater image enhancement and beyond. *IEEE Transactions on Image Processing*, 31, 4922-4936.
- [62] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18* (pp. 234-241). Springer international publishing.
- [63] Li, C., Guo, C., Ren, W., Cong, R., Hou, J., Kwong, S., & Tao, D. (2019). An underwater image enhancement benchmark dataset and beyond. *IEEE transactions on image processing*, 29, 4376-4389.

# Accept Letters and Registration Payment Receipts

## Acceptance Letter



Paper ID: ICDAM - 549  
Date: 08-04-2025

**6<sup>th</sup> INTERNATIONAL CONFERENCE ON DATA ANALYTICS & MANAGEMENT-ICDAM 2025**

**Letter of Acceptance**

**Paper Title:** LEARN: Laplacian Enhanced Attention and Residual Network for Low Light Image Enhancement

**Author (s):** Jyotirmaya Tembhurne, Rahul Katarya


**Congratulations!**

Based on the recommendations of the Technical Program Committee of (ICDAM-2025) we are pleased to inform you that your manuscript has been **Accepted as a regular paper** and will be processed for Publication in the Springer Series "Lecture Notes in Networks and Systems" [ISSN: 2367-3389; 2367-3370] (Scopus Indexed). The paper Shall appear in **ICDAM-2025** in Lecture Notes in Networks and Systems. We would like to invite you to register and participate in the conference. We will encourage more quality submissions from you and your colleagues in the future.

  
**TPC Chair /  
Conference Chair / Editor  
ICDAM-2025-SPRINGER**

General Series Name: Lecture Notes in Networks and Systems  
Series ISSN: 2367-3389; 2367-3370  
<https://www.springer.com/series/15179>

## Payment Receipt



To ICICC


**₹15,355**

ICDAM Registration payment  
online mode ID 549

**Pay again**

Completed

4 Apr 2025, 3:51 pm




Union Bank of India  
2693

UPI transaction ID  
546023951539

To: ICICC  
eze0002677@cub

From: JYOTIRMAYA TEMBHURNE  
(Union Bank of India)  
Google Pay · jayt1712@okaxis

Google transaction ID  
CICAgliinZodbA

POWERED BY  


**Pay**



Paper ID: ICDAM - 1297  
Date: 16-05-2025

**6<sup>th</sup> INTERNATIONAL CONFERENCE ON DATA ANALYTICS & MANAGEMENT-ICDAM 2025**

**Letter of Acceptance**

**Paper Title:** SETAU-NET: Sobel Enhancement with Transformer Attention U-Net for Underwater Image Enhancement


**Author (s):** Jyotirmaya Tembhurne (Delhi Technological University)\*; Rahul Katarya (Delhi Technological University)

**Congratulations!**

Based on the recommendations of the Technical Program Committee of (ICDAM-2025) we are pleased to inform you that your manuscript has been **Accepted as a regular paper** and will be processed for Publication in the Springer Series "Lecture Notes in Networks and Systems" [ISSN: 2367-3389; 2367-3370] (Scopus Indexed). The paper Shall appear in **ICDAM-2025** in Lecture Notes in Networks and Systems. We would like to invite you to register and participate in the conference. We will encourage more quality submissions from you and your colleagues in the future.

  
**TPC Chair /  
Conference Chair / Editor  
ICDAM-2025-SPRINGER**

General Series Name: Lecture Notes in Networks and Systems  
Series ISSN: 2367-3389; 2367-3370  
<https://www.springer.com/series/15179>



To ICICC


**₹15,355**

ICDAM Registration Payment  
Online mode ID 1297

**Pay again**

Completed

16 May 2025, 12:17 pm




Union Bank of India  
2693

UPI transaction ID  
550274397543

To: ICICC  
eze0002677@cub

From: JYOTIRMAYA TEMBHURNE  
(Union Bank of India)  
Google Pay · jayt1712@okaxis

Google transaction ID  
CICAgKjgvNnkDA

POWERED BY  


**Pay**

# **DELHI TECHNOLOGICAL UNIVERSITY**

(Formerly Delhi College of Engineering)  
Shahbad Daultpur, Main Bawana Road, Delhi-42

## **PLAGIARISM VERIFICATION**

Title of the Thesis **Enhancement and Restoration of Images Under Adverse Visual Conditions Using Deep Learning Techniques**

Total Pages **47** Name of Scholar **Jyotirmaya Tembhurne**

Supervisor **Rahul Katarya**

Department **Computer Science and Engineering**

This is to report that the above thesis was scanned for similarity detection. Process and outcome are given below:

Software used: **Turnitin** Similarity Index: **7%** Total word count: **18,212**

Date:

**Candidate's Signature**

**Signature of Supervisor**

# THESIS JT.pdf



Delhi Technological University

## Document Details

### Submission ID

trn:oid:::27535:97357889

### Submission Date

May 23, 2025, 1:43 PM GMT+5:30

### Download Date

May 23, 2025, 1:46 PM GMT+5:30

### File Name

THESIS JT.pdf

### File Size

2.0 MB

47 Pages

18,212 Words

109,503 Characters





# 7% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.




## Filtered from the Report

- Bibliography
- Quoted Text
- Cited Text
- Small Matches (less than 10 words)

## Match Groups



-  **94 Not Cited or Quoted 7%**  
Matches with neither in-text citation nor quotation marks
-  **0 Missing Quotations 0%**  
Matches that are still very similar to source material
-  **0 Missing Citation 0%**  
Matches that have quotation marks, but no in-text citation
-  **0 Cited and Quoted 0%**  
Matches with in-text citation present, but no quotation marks

## Top Sources

- 5%  Internet sources
- 4%  Publications
- 4%  Submitted works (Student Papers)

## Integrity Flags

### 2 Integrity Flags for Review

-  **Replaced Characters**  
53 suspect characters on 17 pages  
Letters are swapped with similar characters from another alphabet.
-  **Hidden Text**  
6 suspect characters on 4 pages  
Text is altered to blend into the white background of the document.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

## Match Groups

- 94 Not Cited or Quoted 7%**  
Matches with neither in-text citation nor quotation marks
- 0 Missing Quotations 0%**  
Matches that are still very similar to source material
- 0 Missing Citation 0%**  
Matches that have quotation marks, but no in-text citation
- 0 Cited and Quoted 0%**  
Matches with in-text citation present, but no quotation marks

## Top Sources

- 5% Internet sources
- 4% Publications
- 4% Submitted works (Student Papers)

## Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	arxiv.org	1%
2	Internet	repository.naturalis.nl	<1%
3	Internet	assets.researchsquare.com	<1%
4	Submitted works	University of Sydney on 2023-11-04	<1%
5	Internet	assets-eu.researchsquare.com	<1%
6	Internet	www.mdpi.com	<1%
7	Internet	openaccess.thecvf.com	<1%
8	Internet	ebin.pub	<1%
9	Publication	Wang Mao-sen, Niu Shao-zhang. "An enhancement algorithm for mobile docume...	<1%
10	Publication	"Computer Vision – ACCV 2024 Workshops", Springer Science and Business Media ...	<1%

11	Submitted works	Hong Kong Baptist University on 2023-04-25	<1%
12	Submitted works	Universitas Dian Nuswantoro on 2016-01-14	<1%
13	Submitted works	University of Bristol on 2023-08-03	<1%
14	Publication	Shangwang Liu, Feiyan Si, Yinghai Lin. "CSUnet: a dual attention and hybrid conv...	<1%
15	Publication	"Pattern Recognition", Springer Science and Business Media LLC, 2025	<1%
16	Submitted works	Technical University of Cluj-Napoca on 2024-07-10	<1%
17	Publication	Xin Zhang, Xia Wang. "MARN: Multi-scale attention retinex network for low-light i...	<1%
18	Submitted works	Indian Institute of Technology Roorkee on 2025-01-27	<1%
19	Submitted works	University of Northampton on 2025-05-18	<1%
20	Internet	pmc.ncbi.nlm.nih.gov	<1%
21	Submitted works	Queen Mary and Westfield College on 2022-11-20	<1%
22	Publication	Shuang Wang, Qianwen Lu, Boxing Peng, Yihe Nie, Qingchuan Tao. "DPEC: Dual-P...	<1%
23	Submitted works	Asian Institute of Technology on 2024-07-17	<1%
24	Submitted works	Cardiff University on 2024-10-07	<1%



25	Publication	Nadeem Sarwar, Asma Irshad, Qamar H. Naith, Kholod D.Alsufiani, Faris A. Almal...	<1%
26	Publication	Runmin Cong, Wenyu Yang, Wei Zhang, Chongyi Li, Chun-Le Guo, Qingming Huan...	<1%
27	Submitted works	University of Lancaster on 2024-05-05	<1%
28	Submitted works	University of Newcastle upon Tyne on 2025-05-06	<1%
29	Submitted works	Adama Science and Technology University on 2023-06-13	<1%
30	Publication	Binghao Huang, Huimin Meng, Lianchao Huang, Chunsi Zhao, Nianmin Yao. "PFL...	<1%
31	Publication	Chunlei Wu, Fengjiang Wu, Jie Wu, Leiquan Wang, Qinfu Xu. "Gradient-guided low...	<1%
32	Publication	Eilif Hjelseth, Sujesh F. Sujana, Raimar J. Scherer. "ECPPM 2022 - eWork and eBusin...	<1%
33	Publication	Lichuan Wang, Shuchun Wang. "A survey of Image Compression Algorithms base...	<1%
34	Submitted works	Liverpool John Moores University on 2024-03-19	<1%
35	Submitted works	University of Hertfordshire on 2024-12-01	<1%
36	Submitted works	University of San Diego on 2024-12-10	<1%
37	Submitted works	University of Sunderland on 2024-05-21	<1%
38	Submitted works	University of Technology, Sydney on 2024-05-24	<1%

39	Submitted works	University of Westminster on 2024-11-17	<1%
40	Internet	drsr.daiict.ac.in	<1%
41	Internet	link.springer.com	<1%
42	Internet	thesai.org	<1%
43	Internet	www.marketresearch.com	<1%
44	Publication	"Data Science and Applications", Springer Science and Business Media LLC, 2024	<1%
45	Publication	"Digital Multimedia Communications", Springer Science and Business Media LLC, ...	<1%
46	Publication	"Machine Learning in Medical Imaging", Springer Science and Business Media LL...	<1%
47	Submitted works	Association of Educators on 2023-10-30	<1%
48	Publication	C.H. Chen. "Signal and Image Processing for Remote Sensing", CRC Press, 2024	<1%
49	Submitted works	City University on 2023-10-18	<1%
50	Publication	Debasis Chaudhuri, Jan Harm C Pretorius, Debashis Das, Sauvik Bal. "Internationa...	<1%
51	Publication	Huan Hu, Fengwen Liu, Nan Su, Wenqiang Hu. "ECF-Net: lumber defect segmenta...	<1%
52	Publication	Jiachen Dang, Yong Zhong, Xiaolin Qin. "PPformer: Using pixel-wise and patch-wis...	<1%

53	Submitted works	La Trobe University on 2023-11-15	<1%
54	Publication	Mingjie Wang, Keke Zhang, Hongan Wei, Weiling Chen, Tiesong Zhao. "Underwat...	<1%
55	Submitted works	Obudai Egyetem on 2025-05-16	<1%
56	Submitted works	The University of Tokyo on 2024-08-30	<1%
57	Submitted works	Tilburg University on 2025-05-18	<1%
58	Submitted works	University of Edinburgh on 2025-04-21	<1%
59	Publication	V. Sharmila, S. Kannadhasan, A. Rajiv Kannan, P. Sivakumar, V. Vennila. "Challeng...	<1%
60	Publication	Yılmaz, İlyas Eren. "Systematic Evaluation of the Effects of Low-resolution and Mo...	<1%
61	Internet	dokumen.pub	<1%